Research Article

Statistics in Medicine

Received 25 November 2014, Accepted 1 February 2016 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6920

Estimating optimal treatment regimes via subgroup identification in randomized control trials and observational studies

Haoda Fu,[a*†] Jin Zhou[b] and Douglas E. Faries[a]

With new treatments and novel technology available, personalized medicine has become an important piece in the new era of medical product development. Traditional statistics methods for personalized medicine and subgroup identification primarily focus on single treatment or two arm randomized control trials. Motivated by the recent development of outcome weighted learning framework, we propose an alternative algorithm to search treatment assignments which has a connection with subgroup identification problems. Our method focuses on applications from clinical trials to generate easy to interpret results. This framework is able to handle two or more than two treatments from both randomized control trials and observational studies. We implement our algorithm in `C++` and connect it with `R`. Its performance is evaluated by simulations, and we apply our method to a dataset from a diabetes study. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: multiple treatments; observational studies; personalized medicine; randomized control trials; subgroup identification; value function

1. Introduction

Responses to treatments can vary widely because of the heterogeneity in most patient populations. One treatment that works for a majority of individuals may not work for a subset of patients with certain characteristics. Thus, significant improvements in treating patients could potentially result from treating individuals based on their specific characteristics rather than a one-size-fits-all approach. Recently, personalized medicine has attracted great attention in medical and biostatistics research. As quoted from an article in Pharmacogenomics: 'therapy with the right drug at the right dose in the right patient' is a great description of how personalized medicine will affect the future of treatment [1].

With new treatments and novel technology available, personalized medicine has become an important piece in the new era of medical product development. At the evolution of drug development, for most therapeutic areas, there are multiple treatments available for the same disease. For example, in treating Type II diabetes mellitus, there are different classes of oral antidiabetic agents (Metformin, Sulfonylurea, Thiazolidinedione, DPP-4, SGLT-2, and so on), and multiple classes of injectable treatments such as GLP-1 and Insulin [2]. In each class, there are multiple drugs, and each drug often has different doses. The various of choices of treatments with different doses for a single disease provide the possibility of personalized medicine. At the same time, the information revolution has allowed the collection of vast amounts of data. Healthcare claims databases and other types of electronic medical records now contain multiple years of medical information for millions of patients. With the advent of genomics and our deepening knowledge of translational medicine, we are able to better characterize a patient, and open great opportunities to develop personalized medicine.

There are notable examples of marketed compounds that make tailored therapeutics a reality. For instance, Tamoxifen used to be a drug commonly prescribed to women with ER+ breast cancer, but 65% of women initially taking it developed resistance. After some research, it was discovered that women

[a]*Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, U.S.A.*
[b]*Biostatistics Department, University of Arizona, Tucson, AZ 85721, U.S.A.*
[*]*Correspondence to: Haoda Fu, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, U.S.A.*
[†]*E-mail: fu_haoda@lilly.com*

with certain mutation in their CYP2D6 gene, a gene that encodes the metabolizing enzyme, were not able to efficiently break down Tamoxifen, making it an ineffective treatment for their cancer [3]. Since then, women are now genotyped for those specific mutations, so that immediately these women can have the most effective treatment therapy. Another example is that Trastuzumab (Herceptin®) is a monoclonal antibody drug that interferes with the HER2/neu receptor. Its main use is to treat certain breast cancers. This drug is only used if a patient's cancer is tested for overexpression of the HER2/neu receptor [4]. Scientists also discovered that metabolic factors may play a role in tailoring as with the recent understanding of how a genetic polymorphism in cytochrome P450 enzyme 2C19 affects the ability of patients to metabolize clopidogrel (Plavix®) to its active form, thereby detrimentally impacting platelet aggregation and clinical outcomes for this subgroup of patients [5].

Despite these successful examples, there remain tremendous challenges in developing solutions for personalized medicine. Besides the medical, operational, ethical, and regulatory issues, there is considerable statistical complexity as well. Much has been written about the dangers of subgroup analyses [6–8], and various authors have proposed guidelines for analysis and reporting on subgroups [9, 10]. In particular, the FDA has already started to take initiatives to integrate personalized medicine into their regulatory policies. They developed a report in October 2013 entitled, 'Paving the Way for Personalized Medicine: FDA's role in a New Era of Medical Product Development', in which they outlined steps they would have to take to integrate genetic and biomarker information for clinical use and drug development. They determined that they would have to develop specific regulatory science standards, research methods, reference material, and other tools in order to incorporate personalized medicine into their current regulatory practices.

There are many existing methods to evaluate heterogeneity. Traditionally, we often conduct the subgroup analysis in clinical trials. Statisticians test for differential treatment effects among subgroups of patients in a very straightforward way (i.e., by adding a treatment-by-subgroup interaction term to a model). Consistent recommendations are that a small number of subgroups be specified in advance, that interaction tests be used, that all subgroup analyses be reported (a priori and post hoc), that multiplicity adjustments be considered, and that interpretation of findings should be viewed cautiously even in the presence of significant multiplicity-adjusted p-values [9, 10]. However, despite our best scientific endeavors and preclinical models, most often, we do not know all the potential effects of a new molecule or biologic. Under such very common circumstances, the traditional approach described previously is not enough in searching for important factors or subgroups of patients.

Therefore, novel methodologies are developed for personalized medicine, which are generally in three categories. The first approach focuses on treatment by subgroup interaction detection. For example, [11] and [12] developed the recursive partitioning methods by building splitting rules based on covariate–treatment interactions. [13] also provided a nice review in this category. The second category of methods are two-step methods such as [14–17]. The first step is to estimate differential treatment effect of each individual patient measured by a score function then use these scores as responses to establish relationship with covariates as the second step. The third class of methods is based on value functions which evaluate the patients benefit. The optimal personalized treatment recommendation rule is based on maximizing a special value function [18–20].

New methodologies greatly extend our ability to explore solutions for personalized medicine, but there are also some limitations. The results from interaction trees are easy to interpret. However, although these algorithms maximize the covariate–treatment interactions at each level of their trees, final trees do not connect with any objective function. Thus, it is hard to define an optimal solution for patients. As consequences, people often struggle with whether to choose a large subgroup with a moderate treatment effect or pick a small subgroup with a significant treatment benefit. Those two-steps methods focus on studies with two treatments, and their methods are not easy to be applied to multiple treatment situations.

The value function approaches provided a nice framework for personalized medicine. Instead of directly searching for subgroups, the value function approaches estimate a treatment regime, which is a rule to assign a treatment, among a set of possible treatments, to a patient as a function of this patient observed characteristics. The goal is to identify the optimal treatment regime that, if followed by the entire population of patients, would lead to the best outcomes on average, which is measured by a value function. Under this framework, there are vigorous methodology developments recently. Very notably, Qian and Murphy developed methods based on l_1 penalized least squares [18]. Zhao *et al.* [19] and Zhang *et al.* [21] connected the value function optimization to weighted classification problems. Zhao *et al.* [19] used a hinge loss function and support vector machine (SVM) algorithms to search the optimal treatment

regime. Zhang *et al.* [20] developed a doubly robust algorithm from a causal inference point of view. Xu *et al.* [22] recently proposed a smooth approximation of the 0–1 loss function to improve the performance for outcome weighted learning algorithms. Although the current algorithms provide accurate results in theory, they are not easy to be applied in clinical trial settings where simple decision rules are preferred. A desired simple rule is well illustrated by the drug label for prasugrel (Effient®) relating to product safety warnings for patients with body weight < 60 kg and age $\geqslant 75$. These simple decision rules for personalized methods are important because they are easy to be implemented and tested in future clinical trials by incorporating them into the protocol through inclusion and exclusion criteria, and in practice, clinicians also prefer to a simple rule for their decision-making [23]. In addition, many methods primarily focus on learning personalized medicine solution based on randomized control trials (RCTs). However, there are growing needs to develop personalized medicine solution based on observational studies, electronic medical records, and insurance claim database [20, 24]. The use of data from observational studies adds additional challenge of addressing the confounding present in such data as part of the personalized medicine solution.

Under the framework of value function approach, in this paper, we develop an alternative algorithm focusing on applications from drug development. We may comprise certain theoretical properties with practical requirements. The prioritized needs are as follows:

(1) The results are easy to interpret and easy to be implemented in clinical settings.
(2) The method is able to handle data from RCTs and also from observational studies.
(3) The method is capable to deal with multiple treatments.
(4) The framework is general enough to handle binary, continuous, time-to-event endpoints.

Based on these requirements, we develop our algorithm which is different than other methods in three aspects: we directly unitize a weighted 0–1 loss function, search on a simple model space consisting of the union of sets on a grid of values, and obtain optimal solution by a comprehensive searching algorithm. We demonstrate the scale and speed of a comprehensive search algorithm by `C++`, and our codes are provided in the Appendix for a broad use.

The rest of this paper is organized as follows: we propose our method in Section 2. Within this section, we start to introduce the general framework of individualized treatment recommendation, and then demonstrate its connection with subgroup identification. We present methods for both two treatment and multiple treatments. How to improve the numerical stability is also discussed. In Section 3, we introduce our search algorithm and demonstrate numerical results through multiple simulation studies. Real data analysis is conducted in Section 4. We close this paper with a discussion in Section 5.

2. Method

In this section, we review the framework of individualized treatment recommendation [18, 19], connect it with subgroup identification problems for multiple treatment subgroup identification methods, and in the end numerical stability issue is discussed.

2.1. Individualized treatment recommendation (ITR)

We have a random sample of size N from a large population. For each unit i in the sample, where $i = 1, \ldots, N$, let T_i be the treatment assignment, Y_i be the response, and X_i be a vector of covariates. Let (Y, T, X) be the generic random variable of $\{(Y_i, T_i, X_i)\}$, $\mathcal{P}$ be the distribution of (Y, T, X), and E be the expectation with respect to $\mathcal{P}$. For any given ITR $r(\cdot)$, which is a rule defining a treatment recommendation for each individual in a population, we let $\mathcal{P}^r$ be the distribution of (Y, T, X) given that $T = r(X)$.

The research question for ITR or subgroup identification is only valid when multiple treatment options are available for the same subject. In other words, if only one treatment option is allowed or available for certain subjects, the optimal treatment is the only available treatment where the research question is trivial. Therefore, without loss of generality, our population space Ω_x is defined as $\Omega_x = \{x : P(T = t|X = x) \in (0, 1), \forall t \in \mathcal{T}\}$ where $\mathcal{T}$ is a finite collection of treatment options. Because $d\mathcal{P} = p(y|x, t)p(t|x)p(x)$ and $d\mathcal{P}^r = p(y|x, t)I_{t=r(x)}p(x)$, we have

$$\frac{d\mathcal{P}^r}{d\mathcal{P}} = \frac{I_{t=r(x)}}{p(t|x)}. \tag{2.1}$$

The expected value of treatment benefit with respect to r is

$$V(r) = E^r(Y) = \int Y d\mathcal{P}^r = \int Y \frac{d\mathcal{P}^r}{d\mathcal{P}} d\mathcal{P} = E\left\{\frac{I_{T=r(X)}}{p(T|X)} Y\right\}. \tag{2.2}$$

Our goal is to estimate r_o, such that,

$$r_o \quad \in \quad \arg\max_{r \in R} V(r), \tag{2.3}$$

where R is a collection of ways to assign treatments, and it is a space of functions $R = \{r(X)|X \in \Omega_x\}$.

We notice that the responses Y are generic. We can put continuous and binomial responses directly. Following the idea in [22], for time to event responses, we can let Y be the observed time, that is, $Y_i = \min(E_i, C_i)$ where E_i is the event time and C_i is the censoring time.

Figure 1 is an intuitive explanation of Equation (2.2). In this simple example, we have only one covariate X. For each $X = x$, there are two subjects with one on treatment and one on control. If $f_t(x)$ is the response function when subjects taking treatment, and $f_c(x)$ is the response function while subjects taking control. By definition (2.2), the optimal treatment recommendation is $r(x) = I\{f_t(x) > f_c(x)\}$ where $I\{\cdot\}$ is an indicator function, $I\{\cdot\} = 1$ is to recommend patients to take treatment, and $I\{\cdot\} = 0$ for control. Under this recommendation, each subject x receives the benefit as $f(x) = \max\{f_t(x), f_c(x)\}$. This example essentially illustrates the case for a randomized control trial with 1:1 randomization ratio. In reality, the randomization ratio may not be 1:1, and in observational study, the treatment assignment may depend on covariates. The denominator of equation (2.2) $P(T|X)$ takes these factors into account. In particular, when there are two treatments, $P(T|X)$ is the propensity score which has been widely used in causal inference [25]. Because $P(T|X)$ is in the denominator, Equation (2.2) is also connected with the inverse probability weighting methods [26, 27]. Therefore, subgroup identification and patient level predictions based on real-world data using this method will be adjusted for biases from measured confounders.

Figure 2 illustrates another example common to placebo-controlled randomized clinical trials. In panel (a), subjects on treatment group have better outcome than those on the control arm regardless of their covariates. Obviously, the optimal treatment recommendation for each subject is to take treatment. However, to develop a new drug, we often look for a nontrivial benefit over the traditional therapy or placebo. Therefore, we can shift the outcomes from the control group by adding a nontrivial benefit margin as shown in panel (b), so that the ITR method can help us to identify the subjects with a pre-defined nontrivial benefit. This is important for payers because the new treatments usually have higher cost than generic

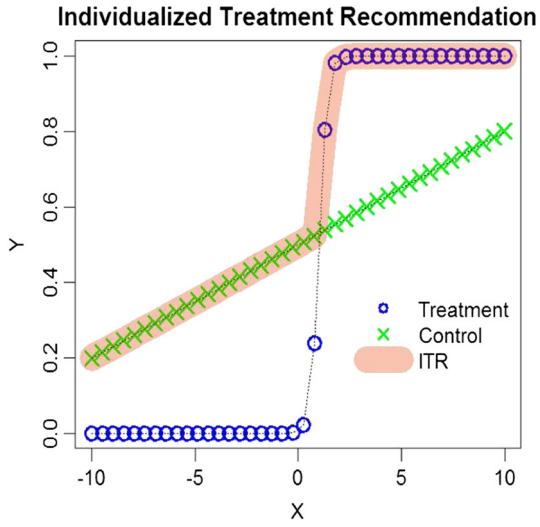


Figure 1. *Illustration of individualized treatment recommendation.* Each subject has one covariate from −10 to 10, and Y is their response. A higher value of Y means a better outcome. Each circle represents one subject receiving treatment, and each 'X' represents one subject receiving control. The shaded area is the optimal treatment recommendation. By following the recommendation, each patient will receive the maximal benefit, then the total benefit for the whole patient population is maximized.

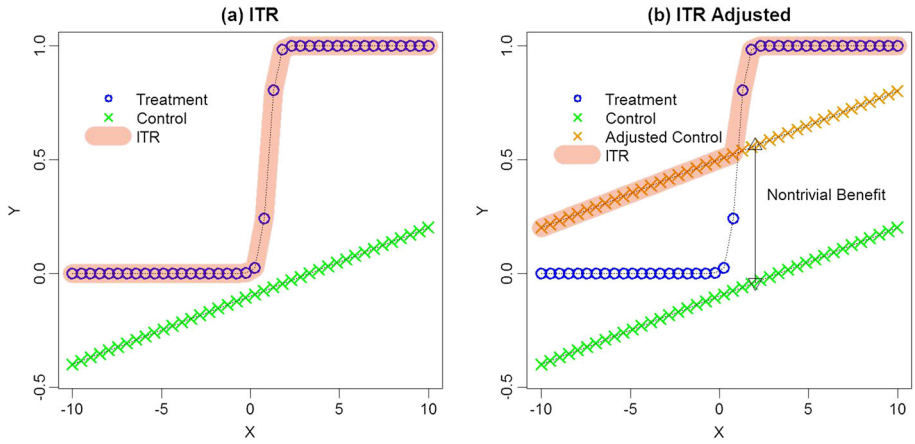


Figure 2. *Illustration of individualized treatment recommendation.* In panel (a), patients on treatment arm have better outcome than the control arm regardless of their covariates. Obviously, the optimal treatment recommendation for each subject is to take treatment. In panel (b), we illustrate that we can adjust the outcomes from control group by adding a nontrivial benefit margin then apply our ITR for the adjusted data.

drugs, and we look for the nontrivial benefit to justify the price. This is also helpful to design a future trial because, to increase the chance of success, it is important to choose an appropriate study population for which the new treatment is expected to have nontrivial overall benefits that compensate for its risks and/or costs. Figure 2 provides an intuitive example on how our ITR method could connect with subgroup identification methods.

Qian and Murphy [18] proposed Equation (2.2) and solve it by a two-step procedure. In the first step, a linear model is fitted to sufficiently approximate the conditional expectation, with the estimated rule derived via l_1 penalized least squares approach. Using a sparse l_1 penalty when modeling the conditional expectation, the method introduces parsimony and facilitates ease of interpretation. One problem with this approach is the mismatch between minimizing the prediction error and the goal of maximizing the value function [28]. A model which may not accurately predict the conditional expectation cannot guarantee a finding of the treatment strategy that yields the optimal value function. In addition, it suffers from overfitting problem when including all the interaction effects with high-dimensional covariate space [29]. [19] showed that solving Equation (2.3) is equivalent to optimize a classification problem with a weighted 0–1 loss function, then they use a hinge loss function to approximate the 0–1 loss function and develop an algorithm based on SVM techniques. They also proved that their solutions converge to the truth. However, in a finite sample, the optimizer from the hinge loss may not be the same as the solution from a 0–1 loss function. Furthermore, [16] pointed out that instead of focusing on finding the optimal or largest subgroup with enhanced treatment effects, it may be more desirable to determine a sub-optimal subgroup from simple and interpretable predictive variables. However, Zhao *et al.* SVM [19] based algorithm solves the optimization problem in a dual space which cannot select important variables directly, and results are difficult to be used to identify easy-to-interpret subgroups. Zhang *et al.* [20] derived the formula from a potential outcome framework with binary treatments. They provided a doubly robust method to estimate the solution. Similar as previous methods, their solution is in a form a linear combinations of covariates which may not be an easy-to-interpret solution.

These limitations motivated us to develop a simple yet direct approach. In this paper, we connect optimizing the value function (2.2) with subgroup identification and provide an algorithm directly searching all subgroups with rectangle shapes. Our goal is to generate easy to interpret subgroups, and the methods not only handle data from randomized control clinical trials but also observational studies.

2.2. *Subgroup identification*

We start with a simple case with binary treatment. Let T indicates whether the treatment of interest was received, with $T = 1$ indicating that subjects receive the active treatment, and $T = 0$ for the control. Using the potential outcome notation, let $Y(0)$ denote the outcome under control, and $Y(1)$ the outcome

under treatment. We observe T and Y, where $Y \equiv TY(1) + (1-T)Y(0)$. Because,

$$V(r) = E\left\{\frac{I_{T=r(X)}}{p(T|X)}Y\right\} = E\left[I_{r(X)=1}\left\{E(Y|T=1,X) - E(Y|T=0,X)\right\}\right] + E\left\{E(Y|T=0,X)\right\},$$

to maximize $V(r)$ with respect to r, we have the optimizer as

$$r_o(X) = \begin{cases} 1, & E(Y|T=1,X) > E(Y|T=0,X), \\ 0, & E(Y|T=1,X) \leqslant E(Y|T=0,X). \end{cases} \tag{2.4}$$

The interpretation of equation (2.4) is straightforward which simply assigns a treatment to patients who can benefit more from it. In other words, a patient with a covariate vector X is recommended to take treatment when $\{E(Y|T=1,X) - E(Y|T=0,X)\} > 0$. Equation (2.4) also connects our method with other personalized medicine methods. The contrast is a score function $\{E(Y|T=1,X) - E(Y|T=0,X)\}$ which is used by [14] and [16].

To have a more close look on Equation (2.4), in the context of personalized medicine, we often assume that responses Y are from the following model:

$$Y \quad = \quad \beta_0 + g(X) + Td(X) + \epsilon, \tag{2.5}$$

where β_0 is the overall mean, $g(X)$ is a function of prognostic markers, and $d(X)$ is a function of predictive markers. Both $g(X)$ and $d(X)$ are centered to 0. Based on Equation (2.4), it is easy to see that the optimal solution $r_o(X) = I\{d(X)>0\}$ which is a function of treatment by covariate interaction in Equation (2.5). Therefore, our method also target on the treatment by covariate interactions which connects the interaction tree approaches [11, 12].

In practice, we are not only interested in developing individualized treatment recommendation but also interested in identifying subgroup of patients which are suitable for a certain treatment. The ideal subgroups are defined by simple and easy to interpret rules. Therefore, if our purpose is to provide treatment recommendation for a group of patients, that is, $r(X) = 1, \forall X \in A$. We can relax the previous condition (2.4) to

$$r_o(X) = \begin{cases} 1, & E\left\{E(Y|T=1,X)|X \in A\right\} > E\left\{E(Y|T=0,X)|X \in A\right\}, \\ 0, & E\left\{E(Y|T=1,X)|X \in A\right\} \leqslant E\left\{E(Y|T=0,X)|X \in A\right\}, \end{cases} \tag{2.6}$$

where the inner expectation is with respect to Y and the outer with respect to X. The aforementioned equation is equivalent to

$$r_o(X) = \begin{cases} 1, & E\left\{\frac{Y\cdot T}{P(T=1|X)}|X \in A\right\} > E\left\{\frac{Y\cdot(1-T)}{P(T=0|X)}|X \in A\right\}, \\ 0, & E\left\{\frac{Y\cdot T}{P(T=1|X)}|X \in A\right\} \leqslant E\left\{\frac{Y\cdot(1-T)}{P(T=0|X)}|X \in A\right\}, \end{cases} \tag{2.7}$$

which is also the solution by restricting the space R in the form as $r(X) = 1$ when $X \in A$, and $r(X) = 0$ when $X \in A^c$.

Therefore, our subgroup identification method for binary treatment situation is described by the following proposition.

Proposition 1
Let

$$r_o^A \quad = \quad I\left[E\left\{\frac{Y\cdot T}{P(T=1|X)}\middle| X \in A\right\} - E\left\{\frac{Y\cdot(1-T)}{P(T=0|X)}\middle| X \in A\right\}\right], \tag{2.8}$$

where $I(x) = 1$ if $x > 0$, otherwise, $I(x) = 0$. Let A_o^1 and A_o^0 be partitions of the patient population Ω_x, that is, $A_o^1 \bigcup A_o^0 = \Omega_x$ and $A_o^1 \bigcap A_o^0 = \emptyset$. The subgroup A_o^1 of patients should be assigned to treatment, that is $T = 1$, is

$$A_o^1 \quad = \quad \arg\max_{A\in\mathcal{A}} E\left\{\frac{I_{T=r_o^A(X)}}{p(T|X)}Y\right\}, \tag{2.9}$$

where $\mathcal{A}$ is the collection of possible subgroups, and the rest of patients should be assigned to control, that is, $A_o^0 = \Omega_x - A_o^1$.

Equation (2.8) is evaluated by observed data where $E\left\{\frac{YT}{P(T=1|X)}|X \in A\right\}$ is estimated by

$$\sum_{x_i \in A} \frac{t_i y_i}{\hat{e}(x_i)} \Big/ \sum_{x_i \in A} \frac{t_i}{\hat{e}(x_i)},$$

and $E\left\{\frac{Y(1-T)}{P(T=0|X)}|X \in A\right\}$ is estimated by

$$\sum_{x_i \in A} \frac{(1-t_i) y_i}{1-\hat{e}(x_i)} \Big/ \sum_{x_i \in A} \frac{1-t_i}{1-\hat{e}(x_i)},$$

where $\hat{e}(x_i)$ is a propensity score which can be estimated by a logistic regression.

Proposition 1 provides a unified framework for individualized treatment recommendation and subgroup identification. The framework is very general. Here, responses Y can be binary, continuous, or time-to-event endpoints, and covariates X can cover a broad range of variables. This framework also allows the treatment assignment to depend on covariates, so that it can handle both RCTs as well as observational studies.

If we require nontrivial treatment benefit for a new treatment as illustrated in Figure 2, the modified response can be written as follows:

$$Y_i^m \equiv T_i Y_i(1) + (1 - T_i) m\{Y_i(0)\},$$

where function $m(\cdot)$ is used to define the additional benefit margin. For continuous endpoint, $m(\cdot)$ could be defined as $m\{Y_i(0)\} = Y_i(0) + \Delta$, where $\Delta > 0$.

Because treatment T in Equation (2.2) is general which allows multiple treatments, by the similar derivation of Proposition 1; it is easy to show that the individualized treatment recommendation for multiple treatments can be obtained as in the following proposition.

Proposition 2
When there are multiple treatments, the optimal individualized treatment recommendation is

$$t_o \quad = \quad \arg\max_{t_p \in \mathcal{T}} E(Y|T = t_p, X).$$

The optimal individualized treatment recommendation for a subgroup of patients A is

$$t_o^A \quad = \quad \arg\max_{t_p \in \mathcal{T}} E\left\{\frac{Y \cdot I_{T=t_p}}{P(T = t_p|X)}\middle| X \in A\right\},$$

Let $\mathcal{G} = A_1, \ldots, A_g$ is a partition of the patient population, that is, $\bigcup_{i=1}^g A_i = \Omega_x$ and $A_i \bigcap A_j = \emptyset, \forall i \neq j$. So $t_o^{A_i}$ is the optimal individualized treatment recommendation for the subgroup of A_i. The optimal subgroups $\mathcal{G}_o$ can be found by

$$\mathcal{G}_o \quad = \quad \arg\max_{\mathcal{G} \in \mathfrak{G}} E\left\{\frac{I_{T=t_o^{G_X}}}{p(T|X)} Y\right\}, \tag{2.10}$$

where $\mathfrak{G}$ is the collection of possible partitions, and G_X is the group where X belongs to.

2.3. Numerical stabilization

Because we do not know the distribution of $\mathcal{P}$, the expectation in Equation (2.2) cannot be directly evaluated. We search the optimal ITR based on observed data. Therefore, r_o is estimated by

$$\hat{r}_o \quad \in \quad \arg\max_{r \in R} \frac{1}{N} \sum \frac{Y_i I_{T_i = r(X_i)}}{p(T = T_i|X_i)}. \tag{2.11}$$

Searching $\widehat{r}_o$ based on random samples is a stochastic optimization problem. Through various of simulation studies, it was found that algorithms for directly optimizing Equation (2.11) based on the original Y_i may not stable.

To fix this problem, following the idea in [30], we introduce the following numerical stabilization algorithm. Instead of using the original Y as responses, we fit a model with Y versus X then put the residuals as response. To demonstrate this approach is valid, we show that

$$\arg\max_{r\in R} E\left\{\frac{I_{T=r(X)}}{p(T|X)}Y\right\} \quad = \quad \arg\max_{r\in R} E\left[\frac{I_{T=r(X)}}{p(T|X)}\{Y-m(X)\}\right],$$

where $m(X)$ is any function of X. This equivalence comes from

$$\begin{aligned} E\left[\frac{I_{T=r(X)}}{p(T|X)}\{Y-m(X)\}\right] &= E\left\{\frac{I_{T=r(X)}}{p(T|X)}Y\right\} - E\left[\frac{I_{T=r(X)}}{p\{T=r(X)|X\}}m(X)\right] \\ &= E\left\{\frac{I_{T=r(X)}}{p(T|X)}Y\right\} - E\left[\frac{p\{T=r(X)|X\}}{p\{T=r(X)|X\}}m(X)\right] \\ &= E\left\{\frac{I_{T=r(X)}}{p(T|X)}Y\right\} - E\{m(X)\}. \end{aligned} \tag{2.12}$$

A key observation here is that $m(X)$ can be any function of X without involving T. Therefore, we do not require a true model. Using the residuals as new responses in the context of subgroup, identification to improve numerical stability is similar to centering the covariates to fit a linear model. In linear model, we are interested in the coefficients of independent variables, and the intercepts can be removed. In subgroup identification, we are interested in the treatment by covariates interactions, and the covariates effects can be removed.

To see why the method works, we provide an illustration example with continuous response, and the principles can be applied to other response types as well. For example, if our data are generated from Equation (2.5), the accuracy varies significantly by simply changing the constant β_0. As shown in the succeeding text,

$$\begin{aligned} \mathbb{E}_N\left\{\frac{I_{T=r(X)}}{p(T|X)}Y\right\} &= \frac{1}{N}\sum\frac{I_{T=r(X_i)}Td(X_i)}{P\{T=r(X_i)\}} \\ &\quad + \mathcal{N}\left(\frac{1}{N}\sum h(X_i), \frac{1}{n^2}\sum\frac{1-P\{T=r(X_i)|X_i\}}{P\{T=r(X_i)|X_i\}}h^2(X_i)\right) + o_p(1), \end{aligned}$$

where $h(X) = \beta_0 + g(X)$, and $\mathbb{E}_N(\cdot)$ is to calculate sample average. When $h(X) \gg d(X)$, the random variable in the second term dominates the results which result in unstable optimizer. Suppose that we have a good estimate of $h(X)$ as $\widehat{h}(X)$, and it is easy to show that

$$\mathbb{E}_N\left[\frac{I_{T=r(X)}}{p(T|X)}\{Y-\widehat{h}(X)\}\right] = \frac{1}{n}\sum\frac{I_{T=r(X_i)}Td(X_i)}{P\{T=r(X_i)\}} + o_p(1).$$

This eliminates the impact of $h(X)$ and stabilizes the solution. The difference between estimating r from Equation (2.2) versus from Equation (2.11) is in Equation (2.12). Because the $\widehat{r}$ in Equation (2.11) involves variability from $I_{T=r(X)}$.

In the end, we pointed out that, as discussed before, we do not require that $m(X)$ is an estimate of a true model $h(X)$. Based on our simulation and empirical experience, a simple linear regression can improve the numerical stability of our searching algorithm. However, what the optimal $m(X)$ to minimize the numerical instability is out of scope of this paper, and it requires some additional research.

3. Implementation and numerical evaluation

In this section, we describe our algorithm and implementation for two treatments situation, then we use five simulation studies to evaluate its performance.

3.1. Implementation

As we discuss in the introduction, our paper focuses on meeting the needs for drug development. We are interested in subgroups with open rectangle shape (e.g., Age $\leqslant 75$ and BMI > 18). Although more complicated subgroups could potentially improve the accuracy, the results could become difficult to implement in real-world clinical settings. The subgroups are defined by certain number of variables. The number of variables to define a subgroup is called the *depth*. In practice, arguably, we prefer a depth is less or equal to three which is aligned with the recommendation from clinical practice discussed by other authors [12]. Therefore, we propose a comprehensive search algorithm. It is implemented in `C++`, and we connect it with `R` [31] through `Rcpp` package [32].

Our searching algorithm for patients who should be on treatment can be described as the following steps:

(1) We have the observed data as (Y, T, X) then use the observed data and a logistic regression to estimate propensity scores $e = P(T = 1|X)$. If the data are from RCTs, we directly use the true randomization ratio as input data for the next step.
(2) We fit a linear model with Y versus X and denote the residuals as $\widetilde{Y}$. The data input to the algorithm are $(\widetilde{Y}, T, e, X)$. We have two remarks here. First, we can fit other models to Y to improve the accuracy. Second, by the derivation in Section 2.3, we can fit linear model to binary data or time-to-event outcomes.
(3) For a given number of covariates $X_1, \cdots, X_p$, we select three covariates as X_{k1}, X_{k2}, X_{k3}, where $k = 1, \ldots, C(3, p)$, and $C(3, p)$ is the number of ways to choose three elements out of p.
(4) For each selected covariate, we choose a split value c_{k1}, c_{k2}, c_{k3}.
(5) For each split value, we select a direction to define the subgroup. It can be $\leqslant$ or $>$. Then, we can finish defining one possible subgroup, for example, $A_{k,j} = (X_{k1} \leqslant c_{k1} \cap X_{k2} \leqslant c_{k2} \cap X_{k3} > c_{k3})$, where $j = 1, \ldots, 8$. Because we choose depth is equal to three, for a fixed three variables and three split values, we have eight ways to define subgroups based on different directions.
(6) After we have a subgroup in step three, we assign treatment to this subgroup $A_{k,j}$ and control otherwise. Then, we evaluate our value function,

$$V_{k,j} = \frac{1}{N} \sum \frac{\widetilde{Y}_i I_{T_i = I_{X_i \in A_{k,j}}}}{T_i e_i + (1 - T_i)(1 - e_i)}.$$

(7) By selecting different covariates, split values, and directions, we can evaluate all the value functions $V_{k,j}$ and provide a subgroup associated maximal value of $V_{k,j}$.

In step four, for continuous variables, we can choose cut points as deciles. Similarly, we can choose the cut points for binary and ordinal categorical variables. In the rare cases that we have nominal categorical variables with many levels, it may require some pre-processing the data to reduce the possibility of all combinations.

We choose simple rectangle shape subgroups defined by less or equal to three covariates for two purposes. On one hand, we would like to keep simple and easy clinical interpretation. On the other hand, it also helps to control the model complexity to reduce over fitting the model.

When more than two treatments available, ways to divide a population are not unique, and it is not very obvious on which way is better. In addition, there are also computational challenges when the regions become more complicated. One possible solution is to borrow ideas for tree algorithms [33] to divide a population for the 4th step of our algorithms in this section. Essentially, we build a tree with the number of nodes equal to the number of treatments. For example, suppose we have five treatments and have selected three variables as X_i, X_j, X_k, as an example, one possible five regions might look like $\{X_i > c_1\}, \{X_i \leqslant c_1 \cap X_j > c_2\}, \{X_i \leqslant c_1 \cap X_j \leqslant c_2 \cap X_i > c_3\}, \{X_i \leqslant c_1 \cap X_j \leqslant c_2 \cap X_i \leqslant c_3 \cap X_k > c_4\}$, and $\{X_i \leqslant c_1 \cap X_j \leqslant c_2 \cap X_i \leqslant c_3 \cap X_k \leqslant c_4\}$.

3.2. Simulation

In this section, we design five simulation studies to evaluate our algorithm.

Speed test: The comprehensive search is computational intensive, because there are many combinations to define a subgroup. For example, if we have $p = 30$ covariates, depth is equal to 3, and each covariate has 10 split values; there will be about 32 million options to define subgroups

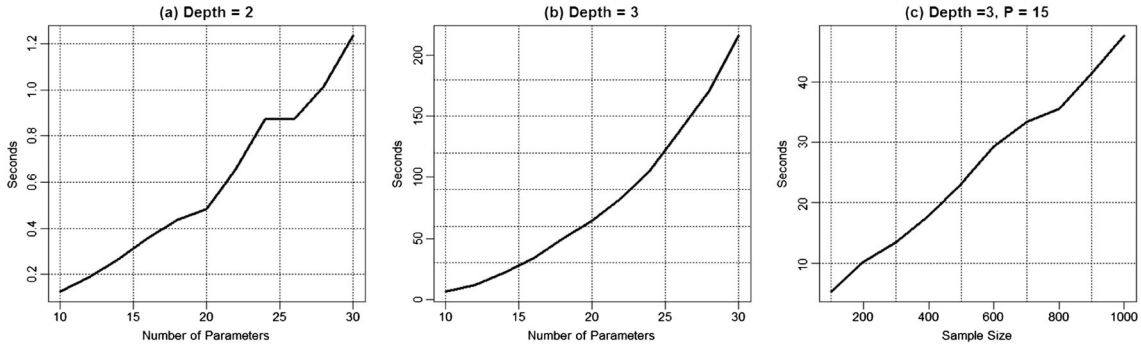


Figure 3. *Speed test for ITR algorithm.* From panel (a) and panel (b), the time spent is in an exponential shape as the number of parameters increasing. In panel (c), the time spent is in an approximately linear relationship with the sample size.

$(30 * 29 * 28/6 * 10^3 * 8)$. Our first question to our algorithm is whether this is feasible. Thanks to the `Rcpp` package, `C++` tremendously improves speed to approximate 400 fold than our the original `R` codes. Figure 3 shows the time for each situation. We ran our simulation on a regular laptop, with CPU Intel i5 M520 2.4 Ghz 3 GB RAM, 32-bit Windows XP Version 2002 Service Pack 3, and `R 3.0.0`. In this simulation, each variable has 10 split values. As we can see from panel (b), when there are 30 covariates, depth is equal to 3, our algorithm takes about 200 s to evaluate all subgroups. The time is exponentially increases with the number of covariates, but it is linearly increase with sample size as shown in panel (c).

Improve numerical stability: The second simulation is to demonstrate the procedure described in Section 2.3 can improve the numerical stability. We simulate four covariates $X_1, \ldots, X_4$. Each covariates is generated from a uniform distribution $U(0, 1)$. Response Y is generated from the following model:

$$Y = \beta_0 + \text{sign}(X_2 - 0.5) + T \cdot I_{X_1 \leqslant 0.6} + (1 - T) \cdot I_{X_1 > 0.6}.$$

We do not include random noise ϵ in this model so that the randomness are from nuisance covariates $X3, X4$. Therefore, we can better evaluate our procedure. We simulate 200 patients for our study which is close to the sample size as Phase II studies. The treatment is randomly assigned with a probability as 0.5. We change our intercept from 0 to 20, and other parts of the model keep the same through whole simulation. For different models with different intercepts, we run 1000 simulations with searching depth equal to 1. In each simulation, we count whether we can correctly select X_1 as the most important variable (i.e., with largest value function). As shown in Figure 4, as the intercept increases, fitting on original Y results in a poor performance. In particular, when $\beta_0 = 20$, the chance to select X_1 is close to 1/4 which is close to a random selection (there are four covariates in this example). However, if we use the residuals from a linear model as responses, the chance to select right variables is almost 100% correct in this simulation. From now on, all the simulations and data analysis are based on residuals instead of original Y.

Convergence in RCTs: It is natural to ask whether our algorithm can converge to the truth when sample size or signal-noise ratio becomes larger and larger. We design a simulation study to answer this question in RCTs settings. We have four covariates $X_1, \ldots, X_4$. Each covariates is generated from a uniform distribution $U(0, 1)$. Responses are generated form the following model:

$$Y = \text{sign}(X_2 - 0.5) + \theta \cdot T \cdot I_{X \in A} + \theta \cdot (1 - T) \cdot I_{X \in A^c} + \epsilon,$$

where ϵ is i.i.d. $N(0, 1)$. We define true subgroup A by different depths from 1 to 3 as

- When depth is equal to 1, $A = \{X : X_1 \leqslant 0.6\}$.
- When depth is equal to 2, $A = \{X : X_1 \leqslant 0.7 \cap X_3 > 0.3\}$.
- When depth is equal to 3, $A = \{X : X_1 \leqslant 0.8 \cap X_2 > 0.2 \cap X_3 > 0.2\}$.

Study sample sizes are tested at 100, 200, 500, and 1000. The treatment is randomly assigned with probability equal to 0.5. We use θ to change the signal-noise ratio. We range θ from 0.1 to 0.5

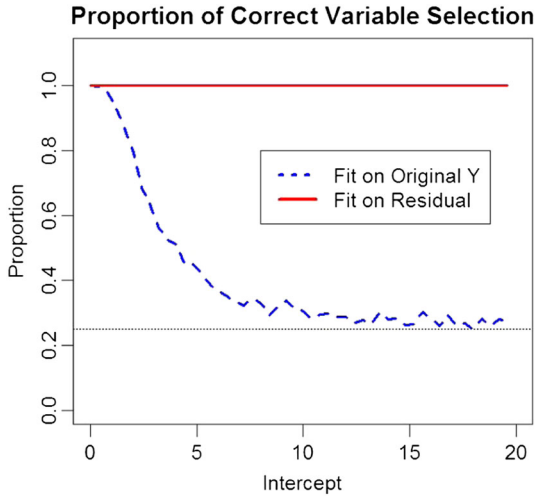


Figure 4. *Improving numerical stable*. Data are generated from $Y = \beta_0 + \text{sign}(X_2 - 0.5) + T \cdot I_{X_1 \leqslant 0.6} + (1-T) \cdot I_{X_1 > 0.6}$ where we include X_3 and X_4 as nuisance covariates. The dotted line is the proportion of corrections to select X_1 to define the treatment rule by applying ITR on the observed Y. The solid line is from applying ITR to the residuals from a linear regression with Y and $\{X_1, \ldots, X_4\}$. It shows that fitting on residuals significantly improve the accuracy.

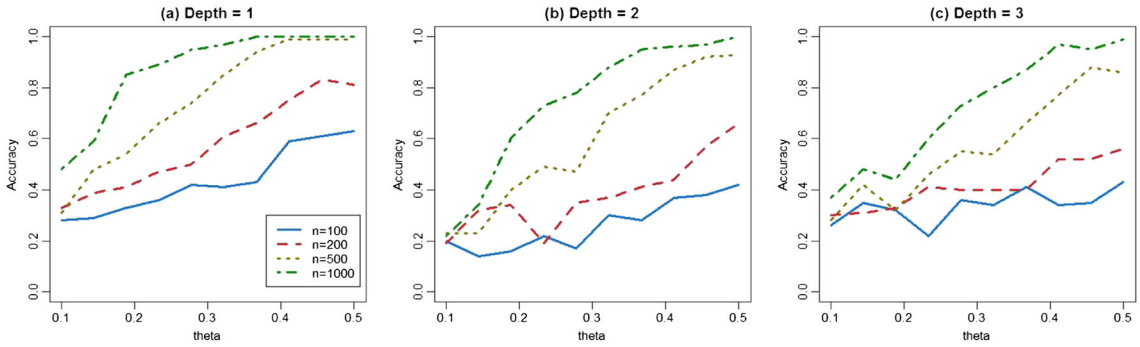


Figure 5. *Algorithm convergence*. Data are generated from simulated randomized control trials. These three plots illustrated that when sample sizes or signal-to-noise ratio go to higher, and the algorithms become more accurate and converge to the truth.

where larger θ means stronger signal. We evaluate accuracy by correctly picking out the variables to define the subgroups. The results are shown in Figure 5. This simulation demonstrates that our algorithm converges to the truth when sample size or signal-noise ratio becomes larger and larger.

Convergence in observational studies: We design this simulation study to evaluate the performance of our method to identify correct subgroup when treatment assignment depends on covariates. We have five covariates $X_1, \ldots, X_5$ which are generated from uniform distributions $U(0, 1)$. Responses are generated from the following model:

$$Y \quad = \quad 1 + 2 \cdot X_2 + \theta \cdot T \cdot I_{X_1 > 0.5} + \theta \cdot (1 - T) \cdot I_{X_1 \leqslant 0.5} + \epsilon,$$

where ϵ is i.i.d. $N(0, 1)$. The treatment assignment probability is from the following model:

$$\text{logit}(p) \quad = \quad -0.5b + bX_i,$$

where we let $b = 6.5$, so that we have 95% chance to assign treat 1 when $X_i = 0.05$. We have three different sets of simulation for $i = 1, 2, 3$. When $i = 1$, the treatment assignment is related to the predictive marker, when $i = 2$, the treatment assignment is associated with prognostic marker, and when $i = 3$, the treatment assignment is correlated with the nuisances covariate. We vary the

Table I. Algorithm convergence for simulated observational studies.

	X_1			X_2			X_3		
	$\theta = 0.25$	$\theta = 0.5$	$\theta = 1$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 1$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 1$
$n = 500$	0.282	0.805	0.998	0.235	0.723	0.966	0.240	0.675	0.968
$n = 2000$	0.839	0.999	1.000	0.678	0.972	1.000	0.648	0.978	1.000
$n = 5000$	0.989	1.000	1.000	0.913	0.999	1.000	0.921	0.999	1.000

Data are generated from simulated observational. The table shows that when sample sizes n or signal-to-noise ratio (θ) go to higher, the algorithms become more accurate and converge to the truth. The numbers in this table are the percentage to capture the true variable with the true cut point.

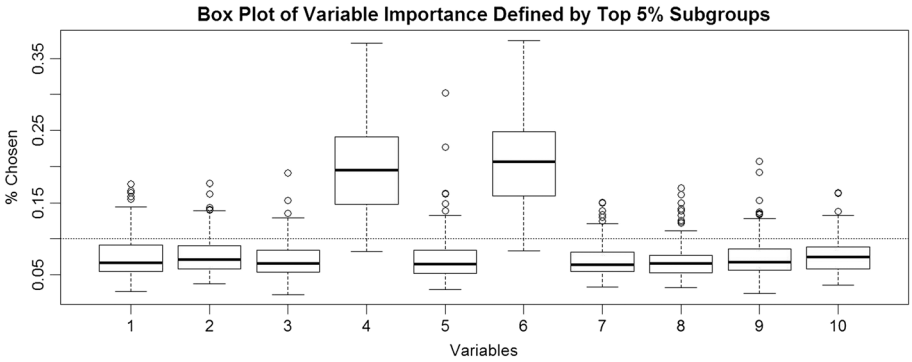


Figure 6. *Variable importance*. We have 10 variables, and the true subgroup is defined by variable 4 and 6. We have 500 simulations, and in each run, we count the relative frequency for each variable to define the 5% ITRs associated with the highest value functions.

sample size as 500, 2000, and 5000, and change $\theta = 0.25, 0.5, 1$. The accuracy is measured by the percentage times that we correctly select the subgroup as $A = \{X : X_1 > 0.5\}$. The results are shown in Table I. This simulation demonstrates that our algorithm converges to the truth for observational studies when sample size or signal-noise ratio becomes larger and larger.

Variable importance: The user of our method may ask which value of the 'depth' parameter may be most reasonable to choose. We may introduce redundancy to use depth equal to 3 when the true depth is 2, and we may miss some signal to use depth equal to 2 when the truth is equal to 3. To answer this question, we need a way to evaluate the relatively importance of the variables. This simulation demonstrates that our method is able to identify the important variables to define subgroups. We have 10 covariates $\{X_1, \ldots, X_{10}\}$ which are generated from uniform distributions $U(0, 1)$. Responses are generated from the following model:

$$Y = \text{sign}(X_2 - 0.5) + 0.5T \cdot I_{X \in A} + 0.5 \cdot (1 - T) \cdot I_{X \in A^c} + \epsilon,$$

where ϵ is i.i.d. $N(0, 1)$, $A = \{X : X_4 \leqslant 0.7 \cap X_6 \leqslant 0.7\}$. We have 500 subjects in each simulation study, and treatment is randomly assigned with probability equal to 0.5. For this simulation, we select the top 5% subgroups with the largest value function, then count the proportion of each covariates. We repeat this simulation 1000 times, and the results are shown in Figure 6. It shows that our algorithm successfully pick out covariates X_4 and X_6 as the most important variables to define the subgroups.

4. Data analysis example

Diabetes mellitus, or simply diabetes, is a disease characterized by elevated blood glucose. It is a major cause of kidney failure, nontraumatic lower-limb amputations, blindness, heart disease, and stroke. As a result, diabetes is one of the leading causes of death. Based on the data from Centers for Disease Control and Prevention (http://www.cdc.gov/diabetes/data), in year 2014, diabetes affects 29.1 million American people which is 9.3% of the US population. The newly diagnosed cases are expected to be at a rate of 1 million people per year. The goal of treating diabetes patients is to lower their blood glucose. Patients are

often first on diets and exercise then are prescribed for metformin which is the first line oral antidiabetic treatment. After patients are failed on metformin, gliclazide and pioglitazone are popular choices for the second line oral treatments [2]. Therefore, it is important to understand which patients are suitable to use gliclazide and who are suitable for pioglitazone. In our study, we use a data set from a randomized, double-blind, parallel-group comparison Phase III study to compare drug efficacy of gliclazide (control) versus pioglitazone(treatment). A total of 1270 patients with Type 2 diabetes were randomized in this study with poorly controlled HbA1c (7.5–11%). Patients were either received pioglitazone up to 45 mg once daily or gliclazide up to 160 mg two times a day. Primary efficacy endpoint was change in HbA1c from baseline to the end of the study (52 weeks). [34] provides more details on this study design and analyses including a patients demographic table. Here, we define the baseline HbA1c at randomization visit, while if HbA1c value is missing at randomization visit, we use the screening visit HbA1c value which is 2 weeks before. If both values are missing, we remove this patient from analysis. The last observation carry forward method is used to impute the last HbA1c value to calculate change from baseline HbA1c. After we processed the data, there are 593 patients on pioglitazone, and 591 patients are on gliclazide.

In our analysis, we include 22 biomarkers measured at baseline. They are as follows: age, ALT, AST, BMI, diastolic blood pressure, systolic blood pressure, cholesterol, creatinine, duration of diabetes, fasting blood glucose, fasting insulin, GGT, HbA1c, HDL, HomaB, HomaIR, HomaS, LDL, Pulse, triglycerides, waist, and weight. The reason that we use acronyms for some of these biomarkers is because they are standard names in clinical practice which could be used more often than their full names (e.g., BMI vs. body mass index).

As an initial step, we fit our data with a depth equal to 3. Therefore, we have 35,925,120 candidate ITRs to define the partition of our patient population. Each ITR is corresponding to a value function which measures the overall benefit under this rule. We select the top .01% ITRs which have the smallest value functions (smaller HbA1c change from baseline means a better efficacy). Then, we plot the frequency of the biomarkers in Figure 7. From the figure, we can see that the top two biomarker fasting insulin and baseline HbA1c are standing out. Therefore, we decide to use depth equal to two to fit the data.

We fit our final model with depth equal to 2. Then, there are 224,532 candidate ITRs. We rank all the value functions corresponding to each ITR. The smallest value function is associated with the ITR is that baseline fasting insulin $\geqslant$ 61.12 pmol/L and baseline HbA1c $\geqslant$ 8.1% (A_o^1) are recommended to take pioglitazone, otherwise (A_o^0) patients are recommended to take gliclazide. It is worth to point out that in this example, the top two biomarkers selected from the first step (depth 3) happens to be the biomarkers associated with the smallest value function in Step 2. This may not be always the case. If not, we may need to consider to comprise the best choice from Step 2 and use clinical judgement to select the ITR. The comparison results are presented in Table II. It clearly shows the value. First, if patients are simply randomly assigned to pioglitazone or gliclazide, the overall HbA1c reduction for these patients are -1.287%. However, after follow the recommended rule, the overall HbA1c reduction becomes to -1.473%. Second, for each subgroup, patients have the chance to significantly improve their outcomes. For example, in group A_o^1, if patients are assigned to gliclazide, they have -1.394% HbA1c reduction

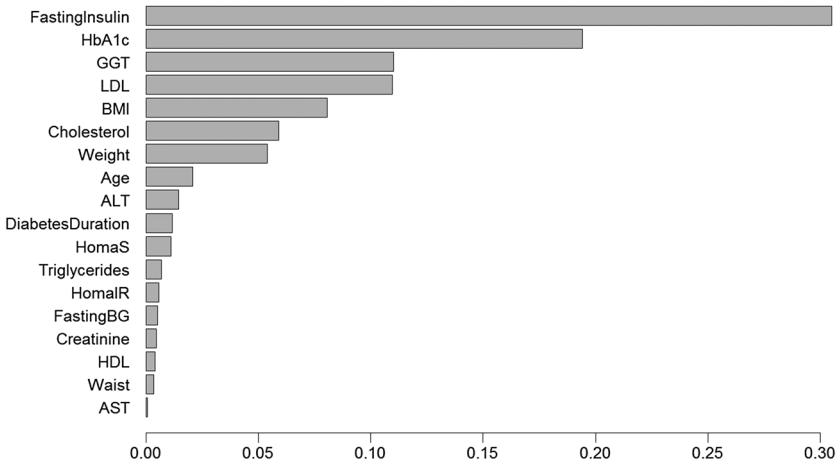


Figure 7. *Variable importance for the real data.* We apply ITR with depth equal to 3 and rank the variable relative frequency for the top 0.01% ITRs.

Table II. HbA1c reduction before and after following ITR.

	Original		Follow ITR		
	−1.287		−1.473		
	Control	Treatment		Control	Treatment
Mean	−1.271	−1.303	A_o^1	−1.394	−1.864
			A_o^0	−1.19	−0.932

Patients with baseline fasting insulin $\geqslant$ 61.12 pmol/L and baseline HbA1c $\geqslant$ 8.1% (A_o^1) are recommended to take pioglitazone, otherwise (A_o^0) patients are recommended to take gliclazide. After following ITR, the overall HbA1c reduction changes from −1.287% to −1.473%.

comparing −1.864% HbA1c reduction if they assigned on pioglitazone. In the end, our subgroup of A_o^1 contains 39.3% patients, and A_o^0 has 60.7% patients.

This simple rule is not only easy to use but also medically meaningful. It is known that pioglitazone is insulin sensitizer which makes patients more sensitive to insulin. Gliclazide belongs to a class of sulfonylurea which push patients beta cell to produce more insulin. For patients with high fasting insulin, pushing out more insulin by gliclazide may not be a solution because they already have good amount of insulin. However, pioglitazone makes patients more sensitive to insulin, and as a consequence, patients can better utilize their insulin to lower blood glucose. Patients who have higher fasting insulin and higher HbA1c are those patients who are insulin resistant while still have a good amount of insulin in their body. It is not surprising that insulin sensitizers, like pioglitazone, work better for those patients. This finding is aligned with the mechanism of action of these two drugs and consistent with other results [34].

5. Discussion

In this paper, we connect the outcome weighted learning approach by [18] and [19] with traditional subgroup identification problems. This framework is general enough to handle data from RCTs as well as observational studies with two or more than two treatments. Furthermore, we developed a searching algorithm which is particularly tailored for drug developing for pharmaceutical companies. The algorithm focuses on searching easy to interpret treatment recommendation rules with an ability to select and rank variables.

One limitation of the current algorithm is that it is not able to incorporate large number of covariates. In current clinical trials, the number of key baseline covariates including demographic and lab data are often less than 50. We can safely say that our algorithm can handle most current late phase studies. However, when we consider more comprehensive biomarkers including genetics data, our algorithm will fail. As a future work, tree algorithms [33] or bump hunting methods [35, 36] could be a solution to reduce the computational complexity.

Because our goal is to maximize the target value function, by exploring various of treatment assignment rule can lead to an over estimate of the value function. It means that the estimated average benefit could be greater than it should be. There are a few possible solutions to mitigate this issue. For example, we can split data as training data set and evaluate the results in the testing dataset. We can also permute the responses Y within each treatment group and apply our ITR method for each permuted database to obtain a distribution of the value function, so that we can compare the estimated average benefit from the original dataset versus the distribution of the value functions.

Currently, we are working on an extension to incorporate meta-analysis into this framework. Sample sizes for most of clinical trials are powered for the primary objectives of those studies, and often not for personalized medicine or subgroup identifications. Synthesising evidence from multiple studies could potentially develop more robust ITR. In particular, our current framework allows the treatment to depend on covariates, and X could incorporate study ID.

Appendix A

```
#include <Rcpp.h>
#include <iostream>
using namespace Rcpp;
```

```
NumericMatrix ITR(NumericVector Y, NumericVector Trt, NumericMatrix X,
NumericVector PS, NumericMatrix CutMatrix, NumericVector NCutPoints, int Depth) {
 // Y is the response, Trt is the treatment, X is the n by p covariates matrix
 // PS is the propensity score
 // CutMatrix is the n by p matrix for cut off values, and NcutPoints is a p
 //by 1 vector to store how many cut points for each varible
 // Depth is the max number of paramters to be used for defining the subgroups
 int nSubgroups = 0;
 // Calculate how many subgroups to evaluates
 if (Depth == 1)
 {  nSubgroups = sum(2*NCutPoints); }
 else if(Depth == 2)
 { for(int iter1=0; iter1 < NCutPoints.size(); iter1 ++)
 { for(int iter2=iter1+1; iter2 < NCutPoints.size(); iter2++)
 { nSubgroups+=4*NCutPoints(iter1)*NCutPoints(iter2); } } }
 else if(Depth != 3)
 { std::cout << "The max Depth allowed is 3.\n";
  Depth = 3;
 }
 if (Depth ==3)
 {for(int iter1=0; iter1 < NCutPoints.size(); iter1 ++)
  {for(int iter2=iter1+1; iter2 < NCutPoints.size(); iter2++)
   {for(int iter3=iter2+1; iter3 < NCutPoints.size(); iter3++)
    {nSubgroups+=8*NCutPoints(iter1)*NCutPoints(iter2)*NCutPoints(iter3);}}}}
 NumericMatrix ScoreMatrix(nSubgroups,1+Depth);
 NumericVector Index(Y.size());
 NumericVector Index1(Y.size()); //Index for the first selection variable
 NumericVector Index2(Y.size()); //Index for the second selection variable
 NumericVector Index3(Y.size()); //Index for the third selection variable
 int RowOfScoreMatrix = 0;
 if(Depth == 1)
 {for(int iter1=0;iter1 < NCutPoints.size();iter1++)
  {for(int iter_cut1=0; iter_cut1 < NCutPoints(iter1); iter_cut1++)
  {Index = (X(_,iter1) <= CutMatrix(iter_cut1,iter1));
  ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
  ScoreMatrix(RowOfScoreMatrix,1) = -((iter1+1.0)+(iter_cut1+1.0)/10);
  RowOfScoreMatrix++;
  ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*(1-Index)/PS+(1-Trt)*Y*Index/(1-PS));
  ScoreMatrix(RowOfScoreMatrix,1) = (iter1+1.0)+(iter_cut1+1.0)/10;
  RowOfScoreMatrix++;}}}
 else if(Depth == 2)
 {for(int iter1=0;iter1 < NCutPoints.size();iter1++)
  {for(int iter2=iter1+1;iter2< NCutPoints.size();iter2++)
   {for(int iter_cut1=0; iter_cut1 < NCutPoints(iter1); iter_cut1++)
    {for(int iter_cut2=0; iter_cut2 < NCutPoints(iter2); iter_cut2++)
     {Index1 = (X(_,iter1) <= CutMatrix(iter_cut1,iter1));
     Index2 = (X(_,iter2) <= CutMatrix(iter_cut2,iter2));
      Index = Index1*Index2;
 ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
 ScoreMatrix(RowOfScoreMatrix,1) = -((iter1+1.0)+(iter_cut1+1.0)/10);
 ScoreMatrix(RowOfScoreMatrix,2) = -((iter2+1.0)+(iter_cut2+1.0)/10);
 RowOfScoreMatrix++;
 Index = Index1*(1-Index2);
 ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
 ScoreMatrix(RowOfScoreMatrix,1) = -((iter1+1.0)+(iter_cut1+1.0)/10);
 ScoreMatrix(RowOfScoreMatrix,2) = ((iter2+1.0)+(iter_cut2+1.0)/10);
 RowOfScoreMatrix++;
 Index = (1-Index1)*Index2;
 ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
```

```
ScoreMatrix(RowOfScoreMatrix,1) = ((iter1+1.0)+(iter_cut1+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,2) = -((iter2+1.0)+(iter_cut2+1.0)/10);
RowOfScoreMatrix++;
Index = (1-Index1)*(1-Index2);
ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
ScoreMatrix(RowOfScoreMatrix,1) = ((iter1+1.0)+(iter_cut1+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,2) = ((iter2+1.0)+(iter_cut2+1.0)/10);
RowOfScoreMatrix++;} } } } }
else
{for(int iter1=0;iter1 < NCutPoints.size();iter1++)
 {for(int iter2=iter1+1;iter2< NCutPoints.size();iter2++)
  {for(int iter3=iter2+1;iter3<NCutPoints.size();iter3++)
   {for(int iter_cut1=0; iter_cut1 < NCutPoints(iter1); iter_cut1++)
    {for(int iter_cut2=0; iter_cut2 < NCutPoints(iter2); iter_cut2++)
     {for(int iter_cut3=0; iter_cut3 < NCutPoints(iter3); iter_cut3++)
      { Index1 = (X(_,iter1) <= CutMatrix(iter_cut1,iter1));
       Index2 = (X(_,iter2) <= CutMatrix(iter_cut2,iter2));
       Index3 = (X(_,iter3) <= CutMatrix(iter_cut3,iter3));
       Index = Index1*Index2*Index3;
ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
ScoreMatrix(RowOfScoreMatrix,1) = -((iter1+1.0)+(iter_cut1+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,2) = -((iter2+1.0)+(iter_cut2+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,3) = -((iter3+1.0)+(iter_cut3+1.0)/10);
RowOfScoreMatrix++;
Index = (1-Index1)*Index2*Index3;
ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
ScoreMatrix(RowOfScoreMatrix,1) = ((iter1+1.0)+(iter_cut1+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,2) = -((iter2+1.0)+(iter_cut2+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,3) = -((iter3+1.0)+(iter_cut3+1.0)/10);
RowOfScoreMatrix++;
Index = Index1*(1-Index2)*Index3;
ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
ScoreMatrix(RowOfScoreMatrix,1) = -((iter1+1.0)+(iter_cut1+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,2) = ((iter2+1.0)+(iter_cut2+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,3) = -((iter3+1.0)+(iter_cut3+1.0)/10);
RowOfScoreMatrix++;
Index = Index1*Index2*(1-Index3);
ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
ScoreMatrix(RowOfScoreMatrix,1) = -((iter1+1.0)+(iter_cut1+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,2) = -((iter2+1.0)+(iter_cut2+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,3) = ((iter3+1.0)+(iter_cut3+1.0)/10);
RowOfScoreMatrix++;
Index = (1-Index1)*(1-Index2)*Index3;
ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
ScoreMatrix(RowOfScoreMatrix,1) = ((iter1+1.0)+(iter_cut1+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,2) = ((iter2+1.0)+(iter_cut2+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,3) = -((iter3+1.0)+(iter_cut3+1.0)/10);
RowOfScoreMatrix++;
Index = (1-Index1)*Index2*(1-Index3);
ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
ScoreMatrix(RowOfScoreMatrix,1) = ((iter1+1.0)+(iter_cut1+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,2) = -((iter2+1.0)+(iter_cut2+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,3) = ((iter3+1.0)+(iter_cut3+1.0)/10);
RowOfScoreMatrix++;
Index = Index1*(1-Index2)*(1-Index3);
ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
ScoreMatrix(RowOfScoreMatrix,1) = -((iter1+1.0)+(iter_cut1+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,2) = ((iter2+1.0)+(iter_cut2+1.0)/10);
ScoreMatrix(RowOfScoreMatrix,3) = ((iter3+1.0)+(iter_cut3+1.0)/10);
```

```
 RowOfScoreMatrix++;
 Index = (1-Index1)*(1-Index2)*(1-Index3);
 ScoreMatrix(RowOfScoreMatrix,0) = mean(Y*Trt*Index/PS+(1-Trt)*Y*(1-Index)/(1-PS));
 ScoreMatrix(RowOfScoreMatrix,1) = ((iter1+1.0)+(iter_cut1+1.0)/10);
 ScoreMatrix(RowOfScoreMatrix,2) = ((iter2+1.0)+(iter_cut2+1.0)/10);
 ScoreMatrix(RowOfScoreMatrix,3) = ((iter3+1.0)+(iter_cut3+1.0)/10);
       RowOfScoreMatrix++;}}}}}}}
eturn ScoreMatrix;}
```

Acknowledgements

The authors are grateful to the editor, associate editor, and referees for review of this article, and they provided numerous insightful comments which significantly improve this manuscript.

References

1. Mancinelli L, Cronin M, Sadée W. Pharmacogenomics: the promise of personalized medicine. *Aaps Pharmsci* 2000; **2**(1):29–41.
2. American Diabetes Association and others. Standards of medical care in diabetes 2014. *Diabetes Care* 2014; **37**(Supplement 1):S14–S80.
3. Ellsworth RE, Decewicz DJ, Shriver CD, Ellsworth DL. Breast cancer in the personal genomics Era. *Current Genomics* 2010; **11**(3):146–161.
4. Telli ML, Hunt SA, Carlson RW, Guardino AE. Trastuzumab-related cardiotoxicity: calling into question the concept of reversibility. *Journal of Clinical Oncology* 2007; **25**(23):3525–3533.
5. Shuldiner AR, OConnell JR, Bliden KP, Gandhi A, Ryan K, Horenstein RB, Damcott CM, Pakyz R, Tantry US, Gibson, Q, *et al.* Association of cytochrome p450 2c19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *Jama* 2009; **302**(8):849–857.
6. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *Journal of Clinical Epidemiology* 2004; **57**(3):229–236.
7. Lagakos SW. The challenge of subgroup analyses-reporting without distorting. *New England Journal of Medicine* 2006; **354**(16):1667.
8. Ruberg SJ, Chen L, Wang Y. The mean does not mean as much anymore: finding sub-groups for tailored therapeutics. *Clinical Trials* 2010; **7**(5):574–583.
9. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *The Lancet* 2005; **365**(9454):176–186.
10. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine?reporting of subgroup analyses in clinical trials. *New England Journal of Medicine* 2007; **357**(21):2189–2194.
11. Su X, Tsai CL, Wang H, Nickerson DM, Li B. Subgroup analysis via recursive partitioning. *The Journal of Machine Learning Research* 2009; **10**:141–158.
12. Lipkovich I, Dmitrienko A, Denne J, Enas G. Subgroup identification based on differential effect search a recursive partitioning method for establishing response to treatment in patient subpopulations. *Statistics in Medicine* 2011; **30**(21): 2601–2621.
13. Doove LL, Dusseldorp E, Van Deun K, Van Mechelen I. A comparison of five recursive partitioning methods to find person subgroups involved in meaningful treatment—subgroup interactions. *Advances in Data Analysis and Classification* 2014; **8**(4):403–425.
14. Cai T, Tian L, Wong PH, Wei LJ. Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics* 2011; **12**(2):270–282.
15. Zhao L, Tian L, Cai T, Claggett B, Wei LJ. Effectively selecting a target population for a future comparative study. *Journal of the American Statistical Association* 2013; **108**(502):527–539.
16. Foster JC, Taylor JM, Ruberg SJ. Subgroup identification from randomized clinical trial data. *Statistics in Medicine* 2011; **30**(24):2867–2880.
17. Faries DE, Chen Y, Lipkovich I, Zagar A, Liu X, Obenchain RL. Local control for identifying subgroups of interest in observational research: persistence of treatment for major depressive disorder. *International Journal of Methods in Psychiatric Research* 2013; **22**:185–194.
18. Qian M, Murphy SA. Performance guarantees for individualized treatment rules. *Annals of Statistics* 2011; **39**(2): 1180–1210.
19. Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association* 2012; **107**(499):1106–1118.
20. Zhang B, Tsiatis AA, Laber EB, Davidian M. A robust method for estimating optimal treatment regimes. *Biometrics* 2012; **68**(4):1010–1018.
21. Zhang B, Tsiatis AA, Davidian M, Zhang M, Laber E. Estimating optimal treatment regimes from a classification perspective. *Statistic* 2012; **1**(1):103–114.
22. Xu Y, Yu M, Zhao YQ, Li Q, Wang S, Shao J. Regularized outcome weighted subgroup identification for differential treatment effects. *Biometrics* 2015; **71**(3):645–653.

23. Foster JC, Taylor JM, Kaciroti N, Nan B. Simple subgroup approximations to optimal treatment regimes from randomized clinical trial data. *Biostatistics* 2015; **16**(2):368–382.
24. Su X, Kang J, Fan J, Levine RA, Yan X. Facilitating score and causal inference trees for large observational studies. *The Journal of Machine Learning Research* 2012; **13**(1):2955–2994.
25. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**(1):41–55.
26. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 1952; **47**(260):663–685.
27. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. *Aids Epidemiology, Methodological issues* 1992; **3**:297–331.
28. Murphy SA. A generalization error for q-learning. *Journal of Machine Learning Research: JMLR* 2005; **6**:1073–1097.
29. Zhao Y, Zeng D. Recent development on statistical methods for personalized medicine discovery. *Frontiers of Medicine* 2013; **7**(1):102–110.
30. Liu Y, Wang Y, Kosorok M, Zhao Y, Zeng D. Robust hybrid learning for estimating personalized dynamic treatment regimes. *Manuscript under Review* 2014.
31. R Core Team. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing: Vienna, Austria, 2014.
32. Eddelbuettel D, François R, Allaire J, Chambers J, Bates D, Ushey K. Rcpp: Seamless r and c++ integration. *Journal of Statistical Software* 2011; **40**(8):1–18.
33. Hastie TJ, Tibshirani RJ, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer: New York, NY, USA, 2011.
34. Charbonnel BH, Matthews DR, Schernthaner G, Hanefeld M, Brunetti P. A long-term comparison of pioglitazone and gliclazide in patients with type 2 diabetes mellitus: a randomized, double-blind, parallel-group comparison trial. *Diabetic Medicine* 2005; **22**(4):399–405.
35. Kehl V, Ulm K. Responder identification in clinical trials with censored data. *Computational Statistics & Data Analysis* 2006; **50**(5):1338–1355.
36. Chen G, Zhong H, Belousov A, Devanarayan V. A prim approach to predictive-signature development for patient stratification. *Statistics in Medicine* 2015; **34**(2):317–342.

Learning Optimal Personalized Treatment Rules in Consideration of Benefit and Risk: with an Application to Treating Type 2 Diabetes Patients with Insulin Therapies

Abstract

Individualized medical decision making is often complex due to patient treatment response heterogeneity. Pharmacotherapy may exhibit distinct efficacy and safety profiles for different patient populations. An "optimal" treatment that maximizes clinical benefit for a patient may also lead to concern of safety due to a high risk of adverse events. Thus, to guide individualized clinical decision making and deliver optimal tailored treatments, maximizing clinical benefit should be considered in the context of controlling for potential risk. In this work, we propose two approaches to identify personalized optimal treatment strategy that maximizes clinical benefit under a constraint on the average risk. We derive the theoretical optimal treatment rule under the risk constraint and draw an analogy to the Neyman-Pearson lemma to prove the theorem. We present algorithms that can be easily implemented by any off-the-shelf quadratic programming package. We conduct extensive simulation studies to show satisfactory risk control when maximizing the clinical benefit. Lastly, we apply our method to a randomized trial of type 2 diabetes patients to guide optimal utilization of the first line insulin treatments based on individual patient characteristics while controlling for the rate of hypoglycemia events. We identify baseline glycated hemoglobin level, body mass index, and fasting blood glucose as three key factors among 18 biomarkers to differentiate treatment assignments, and demonstrate a successful control of the risk of hypoglycemia in both the training and testing data set.

Key Words and Phrases: Personalized Medicine, Benefit Risk Analysis, Hypoglycemia, Machine Learning, Neyman-Pearson Lemma

1 Introduction

Treatment of chronic diseases such as diabetes mellitus is often multifaceted. While maximizing clinical benefit or efficacy is the primary goal, complications and risks related to safety need to be taken into account. For example, the goal of treating patients with type 2 diabetes mellitus is to achieve a glycated hemoglobin (A1C) level of less than or equal to 7%, the level recommended by the American Diabetes Association and the European Association for the Study of Diabetes (Inzucchi et al., 2012). To attain this goal of fast and flexible control of blood glucose levels, oral hypoglycemic agents or insulin therapies are usually administered by clinicians. However, hypoglycemic events are common adverse events which may have s significant negative impact on patients. It is associated with adverse short-term and long-term outcomes, such as increased mortality, seizures, and coma. In addition, fear of hypoglycemia can lead to medication noncompliance and failure to achieve glycemic control (Cryer et al., 2003). While the progressive nature of the disease requires an escalating sequence of medications or dose of insulin, the risk of hypoglycemia and other adverse events may increase with the intensified treatment (Cryer, 2002). Hence, when choosing an optimal treatment regimen for a patient, it is necessary to consider both maximizing clinical benefit (glycemic control) and minimizing risk (hypoglycemia) at the same time.

Due to patient's differential response to a treatment, characterization of individual response heterogeneity has been a critical component of clinical decision making. Recent literature on personalized medicine has flourished with methods on using clinical and biological markers to guide the development of tailored therapies (Su et al., 2009; Lipkovich et al., 2011; Cai et al., 2011; Foster et al., 2011; Zhao et al., 2012). Several work focuses on identifying a target subgroup of subjects who are expected to gain substantial benefit from a given treatment, i.e., the right subpopulation for a given treatment (e.g., Foster et al., 2011). Some recent development aims for identifying the optimal treatment for a given subject (Qian and Murphy, 2011; Zhao et al., 2012). In particular, to estimate the optimal treatment for a patient, machine learning methods such as outcome weighted learning (Zhao et al., 2012; Liu et al., 2014) and

alternatives building on double robust estimation in a semiparametric modeling framework (Zhang et al., 2012) are proposed.

The above referenced work on estimating optimal treatment strategy targets one side of clinical decision making – the clinical benefit and efficacy, but they ignore the potential safety and risk due to the "optimal" and potentially aggressive treatments. The important issue of controlling risk while maximizing benefit has long been recognized in the clinical community where safety concerns for medications often arise, since the most efficacious medication for a patient or a subpopulation may also lead to a higher risk. For example, the recent controversy regarding the safety of thiazolidinediones for treating diabetes led the new guidance issued by the Food and Drug Administration on the need to evaluate the cardiovascular risk of diabetes medications (`http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071627.pdf`). This new guideline calls for attention to adverse events to make accurate conclusions about the efficacy and safety of the medication. In addition, for chronic diseases such as diabetes, the long duration of treatment may expose patients to higher risk of adverse events such as hypoclycemia (Cryer et al., 2003).

Similar to the phenomenon of the heterogeneous treatment effect measured by the efficacy outcome, patients may exhibit heterogeneity in their risk profiles depending on the subject-specific characteristics. For example, Sinclair et al. (2015) reported that the risk of hypoglycemic events is higher among elder individuals. As another example, it is well-known that the relative abundance of drug-metabolizing enzymes such as cytochrome P450 (CYP450) varies from person to person, and genetic polymorphisms associated with CYP450 have been identified (Belle and Singh, 2008). Therefore, patients' adverse reactions to the same drug dosage can show between-subject heterogeneity. In addition, the efficacy outcome and safety or toxicity outcome for an individual is correlated. Increasing a subject's dosage of a medication leads to an increase in the efficacy, but may at the same time increase the risk of adverse events. Thus, in the context of identifying optimal individualized treatment rules, the goal of maximizing efficacy outcome (reward function) needs to be considered in conjunction with controlling for

the risk.

An existing body of literature has considered both efficacy and safety outcomes for personalized medicine through joint modeling of the two outcomes (Houede et al., 2010; Thall et al., 2008; Thall, 2012; Lee et al., 2015). These Bayesian approaches construct a utility function or posterior criterion based on bivariate models of the efficacy and safety outcomes, and estimate an optimal dosing strategy to maximize the utility function or criterion. Subject-specific covariates are introduced to the bivariate model to estimate personalized optimal dosing strategy. Other authors propose to handle multiple outcomes while maximizing clinical benefit through a conditional regression as in Q-learning. See for example, Lizotte et al. (2012); Laber et al. (2014), and (Kosorok and Moodie, 2015, Chapter 15). Solutions were obtained by grid search or iterative methods, and no general theorem was given for a unified optimal solution. Another related work (Kosorok and Moodie, 2015, Chapter 14) proposed outcome weighted learning with a rejection option to reserve the selection of treatment and leave the actual assignment open to other considerations (including risk outcomes). Luedtke et al. (2016) considered to restrict treatment options instead of a risk outcome.

In contrast to the above existing approaches, here we aim to estimate the optimal personalized treatment rule to maximize the clinical benefit while directly impose a constraint such that the average risk under the optimal treatment assignment is lower than a pre-determined, and clinically meaningful threshold. Our approach does not require joint modeling of the efficacy and safety outcome. In addition, by directly controlling the risk while maximizing the efficacy, there is no need to examine the trade-off between the benefit and risk to form a composite outcome or utility function, as done in the benefit-risk analysis (Guo et al., 2010) or as required in some existing work (e.g., Houede et al., 2010). How to weight benefit and risk in a manner that addresses the complexity of the clinical contexts in which a medical decision is made is a separate issue that deserves attention in its own right (Moore et al., 2008). In our approach, the problem of how to compare dissimilar outcomes is avoided.

We translate the scientific goal of maximizing efficacy while controlling for the risk to a

constrained optimization problem, where the resulting optimal rule is expected to maximize the clinical reward function and satisfy the risk constraint. We propose two approaches to solve the risk-constrained learning: one based on regression model with an additional constraint placed to bound the average risk (model-based benefit-risk learning, BR-M), and the other based on outcome weighted learning (benefit-risk O-learning, BR-O). For the latter, the zero-one loss in the risk constraint is approximated by a shifted ramp loss (Huang et al., 2014) instead of the usual hinge loss to allow more precise control of the risk bound. The difference of convex functions algorithm (DC algorithm, Tao and An (1998)) is applied to solve the optimization problem. We derive the theoretical optimal treatment rule under the risk constraint and show a natural connection with finding the optimal rejection region while controlling for the type I error rate as given by the familiar Neyman-Pearson lemma. We perform extensive simulation studies to examine the performance and stability of both algorithms and compare with utility function based methods. Lastly, we apply our approach to the real world motivating study to develop personalized treatment rules to guide the administration of the first line insulin treatment based on individual patient characteristics while restricting the rate of hypoglycemia events.

2 DURABLE Study: A Type 2 Diabetes Randomized Clinical Trial

Type 2 diabetes mellitus (T2DM) is a progressive disease, and the current treatment strategy is one of gradual regimen intensification. When lifestyle intervention and oral anti-diabetic agents fail to achieve adequate glycemic control, insulin treatments are often appropriate next steps. There are a variety of insulin initiation options, such as adding basal insulin or starting with twice-daily premixed insulin. To compare the efficacy, safety, and durability of two common starter insulin regimens, a randomized control clinical trial, DURABLE, was conducted to compare once-daily basal insulin glargine versus twice-daily insulin lispro mix 75/25 (Fahrbach et al., 2008). This study enrolled 2,091 patients from 242 centers in 11 countries (Argentina,

Australia, Brazil, Canada, Greece, Hungary, India, the Netherlands, Romania, Spain, and the U.S.) between December 2005 and July 2007. The trial was designed with a 6-month initiation phase to compare safety and efficacy, and a subsequent 24-month maintenance phase to compare durability. The last patient visit of this trial occurred in December 2009. Because the patients had to be re-consented for the maintenance phase, the second phase was not a randomized trial any more. Therefore, we focus on the data from the first 6-month initiation phase, which consisted of 965 patients on lispro mix and 980 patients on insulin glargine.

While reducing A1C is an important efficacy goal, controlling hypoglycemia is also particularly crucial for optimizing insulin treatment regimen and patient care management. Overmedication with insulin, delays in mealtimes, or insufficient carbohydrate intake to match insulin dose are some common causes of hypoglycemia. The Diabetes Complications and Controls Trial (DCCT) found that intensive therapy for type 1 diabetes patients caused a 3-fold increase in hypoglycemic event rates compared with less aggressive treatment strategy (Control et al., 1997). Similarly, tight glycemic control in type 2 diabetes led to a significant increase in the incidence of hypoglycemia in the UK Prospective Diabetes Study (UKPDS) (Group et al., 1998). Hypoglycemia is a major limiting factor in the management of type 1 and type 2 diabetes. Patients having hypoglycemic events have symptoms ranging from hunger, sweating, to severe cases such as seizure, coma, and even death. Hypoglycemia management also relates to significant health care utilization. The mean costs for hypoglycemia visits were $17,564 for an inpatient admission, $1,387 for an emergency room visit, and $394 for an outpatient visit (Quilliam et al., 2011). As pointed out by Fidler et al. (2011), if we can successfully avoid the hypoglycemic events, glycemia targets would be much easier to achieve. Therefore, it is important to have a treatment algorithm to maximize patients benefit from insulin treatment while controlling for the risk of hypoglycemia.

Our primary efficacy measure (benefit outcome) was A1C at the study end point, and the safety measure (risk outcome) was daily hypoglycemic event rate. The study enrolled patients with baseline A1C $> 7\%$ with a median of 8.8%. The duration of diabetes at baseline ranged

from 0.3 to 39 years with a median of 8 years. Our preliminary data analysis indicated that patients' responses to the treatments were heterogeneous. For example, for patients with baseline A1C $> 8.8\%$ versus $\leq 8.8\%$, their A1C reductions were 2.44 versus 1.08 (p-value < 0.01), and daily hypoglycemic rates were 0.061 versus 0.078 (p-value < 0.01); and for patients with baseline duration of diabetes > 8 years versus ≤ 8 years, their A1C reductions were 1.752 versus 1.746 (p-value $= 0.82$), and daily hypoglycemia rates were 0.082 versus 0.058 (p-value < 0.01). These pooled analyses show that the efficacy and safety endpoints are associated with some covariates but whether these covariates could further interact with the treatment assignment is largely unknown. Hence, it is important to understand the underlying heterogeneity of treatment responses for both benefit and risk outcomes so as to derive the optimal personalized treatment rules to maximize the A1c reduction while controlling for the risk of hypoglycemia.

3 Personalized Treatment Rules Maximizing Benefit While Controlling for Risk

3.1 Statistical framework

Let Y denote the benefit outcome and R denote the risk outcome. Thus, a large Y and small R is desirable. Consider a dichotomous treatment option denoted by $A \in \{-1, 1\}$. Let X denote subject-specific feature variables. A treatment rule is then a map from X to the treatment option domain. For any given treatment rule, say $\mathcal{D}$, the expected benefit and risk are $E^{\mathcal{D}}[Y]$ and $E^{\mathcal{D}}[R]$, respectively, where $E^{\mathcal{D}}[\cdot]$ is the expectation under probability measure $\mathcal{P}^{\mathcal{D}}$ for (Y, R, A, X) given that $A = \mathcal{D}(X)$. Our goal is to estimate a personalized treatment rule that maximizes the expected benefit while controlling the overall expected risk to be below a given threshold.

Specifically, we aim to find an optimal treatment rule, denoted by $\mathcal{D}^*$, such that $\mathcal{D}^*$ solves

$$\begin{cases} \max_{\mathcal{D}} & E^{\mathcal{D}}(Y), \\ \text{subject to} & E^{\mathcal{D}}(R) \leq \tau, \end{cases} \tag{3.1}$$

where τ is a pre-specified value denoting the maximal tolerance of the overall risk. For example,

in the DURABLE study, Y was the reduction in HbA1C at the end point, R was the rate of hypoglycemic events, and τ was chosen to be 0.065 hypoglycemic event/day (or approximately 2 hypoglycemic events/month).

Consider data collected from a randomized trial such as DURABLE, so (Y, R, A, X) follows a non-degenerate distribution $\mathcal{P}$. Furthermore, we assume that the randomization probability $P(A = a|X)$ is bounded strictly away from zero for $a \in \{-1, 1\}$. Under these conditions, from the Randon-Nikodym theorem, since

$$\frac{d\mathcal{P}^{\mathcal{D}}}{d\mathcal{P}} = \frac{I_{a=\mathcal{D}(x)}}{p(a|x)},$$

we obtain that the expected value of treatment benefit under the rule $\mathcal{D}$ is

$$E^{\mathcal{D}}(Y) = \int Y d\mathcal{P}^{\mathcal{D}} = \int Y \frac{d\mathcal{P}^{\mathcal{D}}}{d\mathcal{P}} d\mathcal{P} = E\left\{\frac{I\left(A = \mathcal{D}(X)\right)}{p(A|X)} Y\right\},$$

and the expected risk under the same rule is given as

$$E^{\mathcal{D}}(R) = E\left\{\frac{I\left(A = \mathcal{D}(X)\right)}{p(A|X)} R\right\}.$$

Since

$$E\left\{\frac{Y}{p(A|X)}\right\} - E\left\{\frac{I\left(A = \mathcal{D}(X)\right)}{p(A|X)} Y\right\} = E\left\{\frac{I\left(A \neq \mathcal{D}(X)\right)}{p(A|X)} Y\right\},$$

maximizing the benefit function is equivalent to minimizing $E\left\{I(A \neq \mathcal{D}(X))Y/p(A|X)\right\}$. Treatment rule $\mathcal{D}(X)$ is usually given as the sign of a decision function, i.e., $\mathcal{D}(X) = \text{sign}(f(X))$, for some function $f(X)$. Therefore, estimating $\mathcal{D}^*$ using randomized trial data is equivalent to solving

$$\begin{cases} \min_f & E\left\{\frac{I(Af(X)<0)}{p(A|X)} Y\right\}, \\ \text{subject to} & E\left\{\frac{I(Af(X)\geq 0)}{p(A|X)} R\right\} \leq \tau. \end{cases} \tag{3.2}$$

As a note, $E^{\mathcal{D}}[Y]$ is termed the value function associated with the rule $\mathcal{D}$. Some existing work (e.g., Zhao et al., 2012) aims to maximize this value function to find the optimal treatment rule but impose no risk control, i.e., setting $\tau = \infty$. Here f is identifiable up to a scale. To estimate the optimal rule, only the sign of f is required. To ensure identifiability of f, a constraint on the norm of f (e.g., $\|f\| = 1$) can be imposed.

3.2 Theoretical optimal treatment rule

In this section, we will derive an explicit solution for $\mathcal{D}^*$ that solves (3.2). First, we let $\delta_Y(X) = E[Y|X, A = 1] - E[Y|X, A = -1]$ and $\delta_R(X) = E[R|X, A = 1] - E[R|X, A = -1]$. Note that (3.2) is equivalent to

$$\max_f E\left\{\delta_Y(X)I(f(X) > 0)\right\} \quad \text{subject to } E\{\delta_R(X)I(f(X) > 0)\} \leq \alpha,$$

where $\alpha = \tau - E\left\{E(R|A = -1, X)\right\}$. Clearly, from the constraint, we require that $\alpha \geq E\{\delta_R(X)I(\delta_R(X) < 0)\}$ since otherwise, no f exists. In other words, τ cannot be too small so that no treatment rule can induce risk below τ. Additionally, we assume X to be continuous.

To derive the optimal treatment rule, we consider two different domains of X: $\mathcal{M} = \{X : \delta_Y(X)\delta_R(X) \leq 0\}$ and $\mathcal{M}^c$. Clearly, for $X \in \mathcal{M}$, treatment options yielding higher benefit also reduce risk or treatments yielding lower benefit also increase risk; while it is the opposite for $X \in \mathcal{M}^c$. Hence, for $X \in \mathcal{M}$, we choose $f^*(X) = \text{sign}(\delta_Y(X))$ which gives the largest benefit and smallest risk.

It only remains to determine $f^*(X)$ for $X \in \mathcal{M}^c$. The objective function to be maximized becomes

$$E\left\{\delta_Y(X)I(f(X) > 0)\Big|X \in \mathcal{M}^c\right\}, \text{ subject to } E\{\delta_R(X)I(f(X) > 0)|X \in \mathcal{M}^c\} \leq \alpha^*,$$

where

$$\begin{aligned}\alpha^* &= \frac{\alpha - E\{\delta_R(X)I(\delta_Y(X) > 0, X \in \mathcal{M})\}}{P(X \in \mathcal{M}^c)} \\ &= \frac{\alpha - E\{\delta_R(X)I(\delta_Y(X) > 0, \delta_R(X) \leq 0)\}}{1 - P(\delta_Y(X) > 0, \delta_R(X) \leq 0) - P(\delta_Y(X) \leq 0, \delta_R(X) > 0)}.\end{aligned}$$

The optimal rule is given in the following theorem.

Theorem 1 *The optimal treatment rule is* $\mathcal{D}^*(X) = sign(f^*(X))$, *where*

$$\mathcal{D}^*(X) = sign\left(\delta_Y(X) - \lambda^*\delta_R(X)\right),$$

where $\lambda^* = 0$ *if* $E\left[\delta_R^+(X)|X \in \mathcal{M}^c\right] \leq \alpha^*$*; otherwise,* λ^* *solves equation*

$$E\left[\delta_R(X)I\{\delta_R(X) > 0, \delta_Y(X)/\delta_R(X) > \lambda\}|X \in \mathcal{M}^c\right]$$

$$+E\left[\delta_R(X)I\{\delta_R(X) < 0, \delta_Y(X)/\delta_R(X) < \lambda\}|X \in \mathcal{M}^c\right] = \alpha^*.$$

The proof of Theorem 1 is given in Appendix A.

Remark 1. The proof of Theorem 1 shows that estimating the optimal treatment rule is analogous to finding the optimal rejection region at a given type I error rate as in the Neyman-Pearson lemma, which aims to maximize the power under an alternative hypothesis while controlling the type I error under the null. Therefore, if X is not continuous, there may not exist an exact solution λ^* to satisfy the last equation. In this case, similar to the Neyman-Pearson lemma, we propose to adopt a probability distribution to assign the optimal treatment. Analogous to the fact that the optimal test in Neyman-Pearson lemma is the likelihood ratio test, our optimal treatment rule depends on the benefit-risk ratio, $\delta_Y(X)/\delta_R(X)$.

Remark 2. When there is no treatment heterogeneity on safety outcomes in the population, Theorem 1 still applies with $\delta_R(X) = c$, which is a special case of our general scenario allowing $\delta_R(\cdot)$ to depend on X.

4 Estimation of the Optimal Treatment Rules

In this section, we propose two learning algorithms to estimate the optimal treatment rules using data collected from a randomized trial: regression-based learning and O-learning (abbreviated for outcome weighted learning) in the benefit-risk framework. In the subsequent development, we assume that data consist of (Y_i, R_i, A_i, X_i) for $i = 1, ..., n$.

4.1 Regression-model-based learning algorithm (BR-M learning)

Our first algorithm is based on explicitly estimating the solution as given in Theorem 1. We use the empirical observations to fit regression models for Y given (A, X) and for R given (A, X),

respectively. From the fitted model, we can estimate $\delta_Y(X)$ and $\delta_R(X)$ by $\widehat{\delta}_Y(X)$ and $\widehat{\delta}_R(X)$. Let $\widehat{\mathcal{M}} = \left\{x : \widehat{\delta}_Y(x)\widehat{\delta}_R(x) \leq 0\right\}$ and

$$\widehat{\alpha}^* = \frac{\tau - \bar{R}_1 - n^{-1}\sum_{i=1}^n [\widehat{\delta}_R(X_i) I(\widehat{\delta}_Y(X_i) > 0, X_i \in \widehat{\mathcal{M}})]}{n^{-1}\sum_{i=1}^n I(X_i \in \widehat{\mathcal{M}}^c)},$$

where $\bar{R}_1$ is the average risk outcome in subjects with $A = 1$. Theorem 1 implies that the optimal treatment rule can be estimated by $\widehat{D}(X) = \text{sign}(\widehat{f}(X))$, where

$$\widehat{f}(X) = \begin{cases} \text{sign}(\widehat{\delta}_Y(X)), & X \in \widehat{\mathcal{M}} \\ \text{sign}\left(\widehat{\delta}_Y(X) - \widehat{\lambda}\widehat{\delta}_R(X)\right), & X \in \widehat{\mathcal{M}}^c \end{cases}$$

with $\widehat{\lambda} = 0$ if

$$\frac{\sum_{i=1}^n \widehat{\delta}_R^+(X_i) I(X_i \in \widehat{\mathcal{M}}^c)}{\sum_{i=1}^n I(X_i \in \widehat{\mathcal{M}}^c)} \leq \alpha^*,$$

and otherwise, $\widehat{\lambda}$ is the solution to

$$\frac{\sum_{i=1}^n \widehat{\delta}_R(X_i) I(X_i \in \widehat{\mathcal{M}}^c, \widehat{\delta}_R(X_i) > 0, \widehat{\delta}_Y(X_i)/\widehat{\delta}_R(X_i) > \lambda)}{\sum_{i=1}^n I(X_i \in \widehat{\mathcal{M}}^c)}$$

$$+ \frac{\sum_{i=1}^n \widehat{\delta}_R(X_i) I(X_i \in \widehat{\mathcal{M}}^c, \widehat{\delta}_R(X_i) < 0, \widehat{\delta}_Y(X_i)/\widehat{\delta}_R(X_i) < \lambda)}{\sum_{i=1}^n I(X_i \in \widehat{\mathcal{M}}^c)} = \alpha^*.$$

The regression models can be fitted by linear regression with a sparse penalty when the number of covariates is large.

4.2 Outcome-weighted learning algorithm (BR-O learning)

Our second algorithm directly solves the empirical version of (3.2) using machine learning approaches. This approach is a non-trivial extension of the non-constrained O-learning where maximizing benefit without risk control is the goal. Denote $p_i = p(A_i|X_i)$. We propose to solve the following optimization problem:

$$\begin{cases} \min_f & n^{-1}\sum_{i=1}^n \frac{Y_i}{p_i} I\left(A_i f(X_i) < 0\right), \\ \text{subject to} & n^{-1}\sum_{i=1}^n \frac{R_i}{p_i} I\left(A_i f(X_i) \geq 0\right) \leq \tau. \end{cases} \tag{4.1}$$

However, due to the discontinuity of the indicator function corresponding to the zero-one loss, solving the above constrained optimization problem in (4.1) is an NP-hard problem (Natarajan,

1995). Thus, instead of using the zero-one loss in the optimization, we consider other types of surrogate loss functions and propose feasible algorithms to obtain solutions.

First, following Liu et al. (2014), we modify the above optimization problem to (a) reduce the variability of weights (Y_i/p_i) in the objective function, and (b) handle the situation that the weights in the objective function can be negative, especially after (a). Specifically, for (a), instead of using the original Y as outcomes, we first regress Y on X and let the residuals be the outcomes. The validity of using the residual is because

$$\arg\max_{\mathcal{D}\in\Delta} E\left\{\frac{I(A=\mathcal{D}(X))}{p(A|X)}Y\right\} = \arg\max_{\mathcal{D}\in\Delta} E\left[\frac{I(A=\mathcal{D}(X))}{p(A|X)}\{Y-m(X)\}\right],$$

where $m(X)$ is any function of X. For (b), we note

$$\begin{aligned} E\left\{\frac{I(A=\mathcal{D}(X))}{p(A|X)}Y\right\} + E\left\{\frac{Y^-}{p(A|X)}\right\} &= E\left\{\frac{I(A=\mathcal{D}(X))}{p(A|X)}Y^+\right\} + E\left\{\frac{I(A\neq\mathcal{D}(X))}{p(A|X)}Y^-\right\} \\ &= E\left\{\frac{I(A\cdot\text{sign}(Y)=\mathcal{D}(X))}{p(A|X)}|Y|\right\}. \end{aligned}$$

This suggests that the original optimization problem can be modified as follows. Define $Y_i^* = |Y_i - m(X_i)|$, $A_i^* = \text{sign}\{Y_i - m(X_i)\}A_i$, equation (4.1) is equivalent to

$$\begin{cases} \min_f & \sum_{i=1}^n \frac{Y_i^*}{p_i} I(A_i^* f(X_i) < 0), \\ \text{subject to} & \sum_{i=1}^n \frac{R_i}{p_i} I(A_i f(X_i) \geq 0) \leq n\tau. \end{cases} \tag{4.2}$$

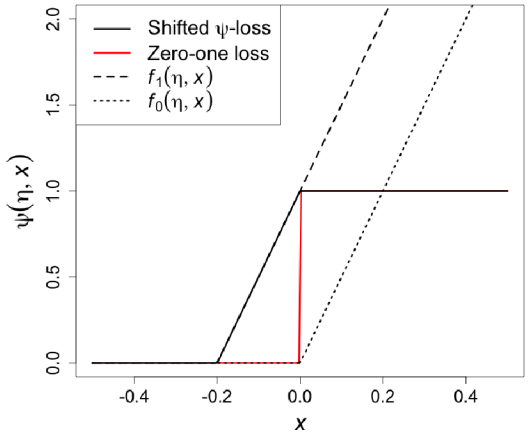
Since the indicator function in the objective function is discontinuous, we introduce a slack variable as the usual support vector machine, and adopt the kernel tricks to estimate f. Specifically, we let $f(X|\beta) = \beta_0 + \sum_{j=1}^n \beta_j K(X, X_j)$ where $K(\cdot,\cdot)$ is a reproducing kernel and introduce slack variables, $\xi_i \geq 1 - A_i^* f(X|\beta)$, and $\xi_i \geq 0$. Then, the optimization problem becomes

$$\begin{cases} \min_f & C\sum_{i=1}^n \frac{Y^*}{p_i}\xi_i + \frac{1}{2}\|f\|_{\mathcal{H}}, \\ \text{subject to} & \sum_{i=1}^n \frac{R_i}{p_i} I(A_i f(X_i) \geq 0) \leq n\tau, \\ & \xi_i \geq 1 - A_i^*\{\beta_0 + \sum_{j=1}^n \beta_j K(X_i, X_j)\}, \xi_i \geq 0 \quad \forall i. \end{cases} \tag{4.3}$$

Here, $\|f\|_{\mathcal{H}}^2$ is the norm for the reproducing kernel Hilbert space which is equivalent to $\beta_{(0)}^T \mathbf{K} \beta_{(0)}$ with $\mathbf{K}$ being the kernel matrix and $\beta_{(0)} = \{\beta_1, \ldots, \beta_n\}^T$.

In the above optimization, the non-convex indicator function in the constraint makes computation difficult. To avoid this discontinuity, we approximate $I\{Af(X) \geq 0\}$ by the following

Figure 1: Comparison of zero-one loss, hinge loss, and shifted $\psi-$loss*



*: Shifted ψ-loss is defined as $\psi(\eta, x) = f_1(\eta, x) - f_0(\eta, x) = \eta^{-1}(x+\eta)_+ - \eta^{-1}(x)_+$.

shifted ramp loss (Huang et al., 2014, see also Figure 1):

$$\psi(\eta, x) \quad = \quad f_1(\eta, x) - f_0(\eta, x) = \eta^{-1}(x+\eta)_+ - \eta^{-1}(x)_+.$$

As we can see that $\psi(\eta, x) \geq I(x \geq 0)$, thus the shifted ψ-loss serves as an upper bound of the zero-one loss function. Therefore, if we can control the risk under loss function $\psi(\eta, x)$, we can also control the risk under the zero-one loss. In summary, the final solution for f^* solves

$$\begin{cases} \min_f & C\sum_{i=1}^n \frac{Y_i^*}{p_i}\xi_i + \frac{1}{2}\beta_{(0)}^T \mathbf{K}\beta_{(0)}, \\ \text{subject to} & \sum_{i=1}^n \frac{R_i}{p_i}\left[\eta^{-1}\{A_i f(X_i)+\eta\}_+ - \eta^{-1}\{A_i f(X_i)\}_+\right] \leq n\tau, \\ & \xi_i \geq 1 - A_i^*\left\{\beta_0 + \sum_{j=1}^n \beta_j K(X_i, X_j)\right\}, \xi_i \geq 0 \quad \forall i. \end{cases} \tag{4.4}$$

We present the detailed algorithm for solving the above optimization problem in the Appendix. Briefly speaking, by expressing the non-convex loss function as the difference of two convex functions, the DC algorithm (Tao and An, 1998) can be applied for optimization, and the tuning parameters are chosen by cross-validation. To stabilize optimization, all covariates will be standardized before fitting the DC algorithm.

5 Simulation Studies

We simulated 10 i.i.d. covariates, $X_1, \ldots, X_{10}$, from a Uniform$(0, 1)$ distribution. The efficacy responses Y and safety responses R were generated from two continuous distributions,

$$Y = 1 - 2X_1 + X_2 - X_3 + h_Y(X, A) + \epsilon_Y,$$
$$R = 2 + X_1 + h_R(X, A) + \epsilon_R,$$

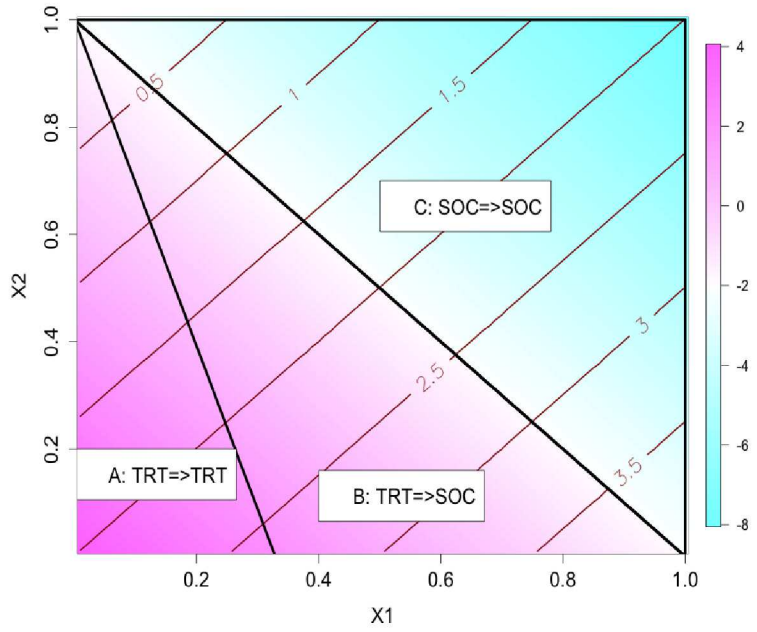
where ϵ_Y follows $N(0, 1)$, ϵ_R follows a truncated standard normal distribution (truncated at 1), $h_Y(X, A)$ and $h_R(X, A)$ are the interaction terms between covariates and treatment for the efficacy and risk outcome, respectively. Here A takes two values: 1 (experimental therapy) and -1 (standard of care), and the randomization probability is 0.5. In the first simulation setting, we considered the decision rules for both risk and efficacy to be linear, that is,

$$h_Y(X, A) = 2(1 - X_1 - X_2)A, \quad h_R(X, A) = (1 + X_1 - X_2)A.$$

In this setting, regardless of whether all subjects are given the experimental therapy or standard treatment, the average efficacy values do not vary greatly and they are both close to 0. However, the risk of receiving the experimental therapy is much higher (3.503) as compared to the standard treatment (1.496). If ignoring the risk outcomes, the optimal treatment strategy based on the efficacy alone results in a maximal efficacy of 0.670 while the average risk reaches 2.495. Therefore, if the safety outcome above some level, say 2.0, is of great concern, this optimal treatment strategy is highly risky and not acceptable.

To illustrate how the risk constraint may affect the optimal treatment decision boundary, in Figure 2 we plot these boundaries and present the partition of optimal treatment regions. Without the risk constraint, the optimal treatment decision boundary is a linear function, $X_1 + X_2 = 1$, so the optimal treatment for subjects in regions A and B is the experimental therapy, TRT, because the benefit function $\delta_Y(X)$ is positive in these regions. Similarly, the optimal treatment for subjects in region C is the standard of care, SOC, because there is no gain in terms of the efficacy for TRT. However, when considering risk constraint, subjects in

Figure 2: Regions of optimal treatment rules with $\tau = 1.75$ in the first simulation setting.



region B originally with TRT the as the optimal treatment should be changed to receiving SOC instead, since the benefit gain from TRT is moderate but the average risk is higher for this group as demonstrated by the contour lines of the risk differences. Subjects in region A maintain TRT as their optimal treatment taking into account of risk because the benefit is higher in this group. Likewise, subjects in region C maintain SOC as the optimal treatment since SOC shows a higher benefit and the risk is always lower than TRT. This region mimics the real world scenario where a safer medication is preferred when the efficacy is similar.

For each simulated data, we applied the proposed two methods to estimate the optimal treatment rules, where we varied the risk threshold τ from 1.75 to 2.25. In the BR-M learning, we estimated both $\delta_Y(X)$ and $\delta_R(X)$ using linear regression models with interactions between X and A included. We then estimated the optimal treatment rule using the expression in Section 4.1, where $\widehat{\lambda}$ was calculated using a grid search algorithm. In the BR-O learning, we chose the kernel function to be a linear kernel, i.e., $K(x, x) =$

$x^T x$. The DC algorithm was used to solve the optimization problem, where at each iteration we used an existing quadratic programming package ("ipop" in **R** package *kernlab* `https://cran.r-project.org/web/packages/kernlab/kernlab.pdf`) for the optimization. The tuning parameters, C and δ, for BR-O were chosen by 2-fold cross validation. The data was split into a training set and a testing set. On the training set, the optimal treatment rule was computed with the risk constraint. On the testing set, the empirical average of the efficacy outcome under the derived rule was computed and used as the criterion for the cross validation. We let C vary from 0.5 to 2 and δ between 0.01 and 0.05. To compare different methods, we independently generated a validation data set of size 20,000 and applied the estimated decision rules to the validation data in order to assess the predicted efficacy and risk on this large testing set. We also calculated the efficacy of the theoretical optimal treatment rule using Theorem 1.

The simulation results from 100 replicates for this setting are shown in Table 1. BR-M and BR-O perform similarly: with a more lenient threshold τ, the average risk increases and average efficacy also increases. Both methods control the risk to be close to the pre-specified level and the estimated optimal treatment agrees with the true optimal treatment for most subjects (proportion of agreement greater than 85%). With a smaller sample size, BR-M provides a slightly larger efficacy and a higher probability of identifying the correct optimal treatment. With increasing sample size, both methods show improved performance. The median efficacy and risk is similar to the mean. Since the true optimal treatment boundary is linear, the model-free BR-O may not greatly improve the performance as compared to the model-based BR-M.

Next, we examined a procedure to rank variable importance using the magnitude of the coefficients of the fitted linear optimal treatment rule. In this simulation setting, the first two feature variables were equally informative (same coefficient on the optimal treatment rule) and the other variables were noise. The average estimated rank based on the absolute value of the fitted coefficient is 1.46 (sd = 0.50) for X_1 and 1.53 (sd = 0.50) for X_2, where the true rank for both variables is 1.5. The eight noise variables have an average rank of 6.50 across 100

simulations, while the true rank is also 6.5. These results suggest that using the magnitude of the coefficients of the optimal rule to determine variable importance is effective when the true optimal rule is linear.

Table 1: Estimated average risk and optimal benefit in the first simulation setting†

τ	Efficacy	n	Method	M-risk (sd)	M-effic (sd)	P-risk (dev)	P-effic (dev)	Accuracy
1.75	0.370	200	BR-M	1.751 (0.062)	0.338 (0.062)	1.750 (0.039)	0.337 (0.037)	93% (1.6%)
			BR-O	1.776 (0.136)	0.321 (0.125)	1.763 (0.094)	0.335 (0.085)	89% (2.5%)
		400	BR-Q	1.747 (0.041)	0.345 (0.042)	1.743 (0.028)	0.349 (0.027)	95% (1.2%)
			BR-O	1.773 (0.111)	0.331 (0.102)	1.768 (0.072)	0.344 (0.068)	90% (1.7%)
2.00	0.556	200	BR-M	2.004 (0.070)	0.525 (0.040)	2.009 (0.041)	0.530 (0.025)	92% (1.8%)
			BR-O	2.005 (0.154)	0.487 (0.091)	1.993 (0.107)	0.492 (0.066)	87% (3.1%)
		400	BR-M	2.001 (0.049)	0.535 (0.028)	1.999 (0.037)	0.535 (0.020)	95% (1.4%)
			BR-O	2.006 (0.113)	0.505 (0.067)	2.000 (0.082)	0.512 (0.049)	90% (2.2%)
2.25	0.648	200	BR-M	2.261 (0.070)	0.617 (0.019)	2.260 (0.051)	0.622 (0.012)	92% (1.9%)
			BR-O	2.216 (0.154)	0.576 (0.055)	2.225 (0.091)	0.592 (0.034)	88% (3.6%)
		400	BR-M	2.251 (0.050)	0.626 (0.015)	2.254 (0.037)	0.626 (0.011)	94% (1.3%)
			BR-O	2.242 (0.102)	0.604 (0.029)	2.236 (0.063)	0.604 (0.021)	91% (2.2%)

†: "Efficacy" is the theoretical efficacy value under the risk constraint; "M-risk" is the mean of the predicted risk in the validation data; "M-effic" is the mean of the predicted efficacy using an independent validation data set of size 20,000; "sd" is the empirical standard deviation; "P-risk" is the median of the predicted risk in the validation data; "P-effic" is the median of the predicted efficacy using an independent validation data set of size 20,000; "dev" is the median of the absolute value of the deviation from the median; "Accuracy" is the proportion of subjects whose predicted optimal treatments agree with the true optimal treatments. The numbers in the parentheses are the median absolute deviation from 100 replicates.

In the second simulation setting, we considered a nonlinear efficacy and safety boundaries by letting

$$Y = 1 - 2X_1 + X_2 - X_3 + h_Y(X, A) + \epsilon_Y,$$

$$R = 2 + X_1 + h_R(X, A) + \epsilon_R,$$

where ϵ_Y and ϵ_R were generated the same as in the previous setting but

$$h_Y(X, A) = 8(1 - X_1^2 - X_2^2)A, \quad h_R(X, A) = (X_1 + X_2 - 0.3)A.$$

The covariates and treatment assignments were the same as setting 1. Figure 3 illustrates 3 regions partitioned by the nonlinear optimal treamtent boundaries. The optimal treatment for subjects in Region C is SOC due to a negative benefit function. However, the risk of treating subjects in region B with TRT is too high ($\delta_R(X) > 0$ with large magnitude). Thus, for region B, the optimal treatment should be switched to SOC when considering risk regardless of a slightly inferior efficacy on SOC. This scenario is similar to setting 1 except that the optimal decision boundaries are nonlinear. The average risk is 2.661 and the theoretical maximal efficacy without risk constraint is 3.602.

The results from setting 2 are summarized in Table 2. Both methods adequately control the risk since the risk outcome is simulated from a linear model. For small sample size, BR-M has slightly higher efficacy and a smaller variability in estimating benefit function since it is model-based. The variability of BR-O is larger than BR-M, leading to some outliers and a slightly smaller mean benefit. When considering median benefit, BR-O performs as well as BR-M even for small sample size: for $\tau = 2.0$ and $n = 200$, the median efficacy for BR-O and BR-M are: 1.801 and 1.834, respectively. With larger sample size, the variability of BR-O is reduced and it has a greater mean benefit (and a greater median benefit) than BR-M under all values of τ.

In the above two simulation settings, we also estimated the optimal rules using the BR-O under a Gaussian kernel. For the Gaussian kernel, following Jaakkola et al. (1999) and Wu et al. (2010), we used a heuristic method to choose the spread parameter σ as $\sigma = 1/d_m^2$, where d_m was the median pairwise Euclidean distance defined as median$\{\|X_i - X_j\| : A_i \neq A_j\}$. The performance was very similar to using the linear kernel with a slightly larger efficacy in the nonlinear setting. However, the computational time using the Gaussian kernel was much more intensive. Therefore, a linear kernel for BR-O may be sufficient in practice.

As a comparison, we also implemented a utility based method inspired by Thall et al. (2008). Let $U(Y_i, R_i) = Y_i - \beta R_i$ denote a utility function, where β reflects the equivalent

Figure 3: Regions of optimal treatment rules with $\tau = 2.0$ in the second simulation setting.

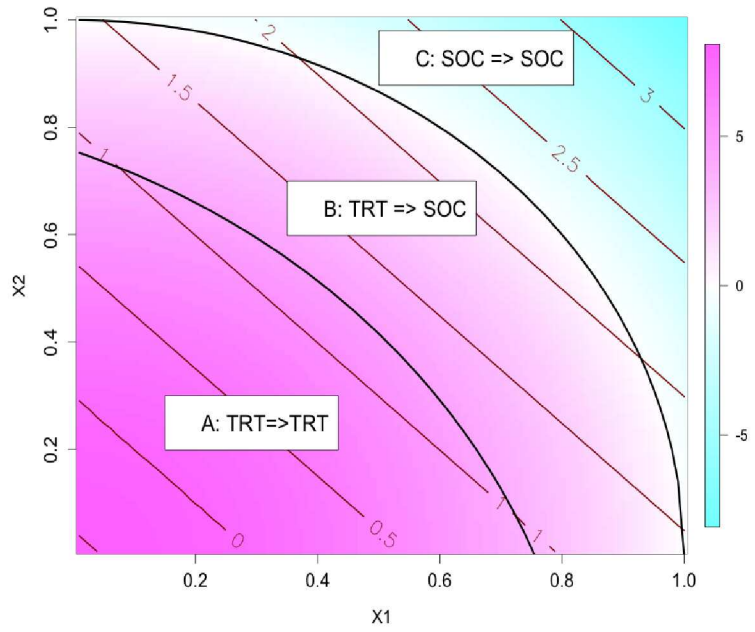


Table 2: Estimated average risk and optimal benefit in the second simulation setting†

τ	Efficacy	n	Method	M-risk (sd)	M-effic (sd)	P-risk (dev)	P-effic (dev)	Accuracy
2.0	1.843	200	BR-M	2.008 (0.051)	1.758 (0.402)	2.008 (0.030)	1.801 (0.226)	94% (2.0%)
			BR-O	2.052 (0.139)	1.737 (1.064)	2.026 (0.101)	1.834 (0.666)	88% (6.0%)
		400	BR-M	2.005 (0.037)	1.785 (0.274)	2.009 (0.027)	1.823 (0.184)	95% (1.0%)
			BR-O	2.037 (0.094)	1.858 (0.701)	2.036 (0.052)	1.944 (0.350)	91% (4.2%)
2.1	2.467	200	BR-M	2.107 (0.051)	2.363 (0.267)	2.107 (0.033)	2.399 (0.138)	93% (1.1%)
			BR-O	2.135 (0.137)	2.296 (0.689)	2.131 (0.107)	2.438 (0.455)	88% (3.9%)
		400	BR-M	2.106 (0.042)	2.380 (0.207)	2.107 (0.026)	2.402 (0.128)	94% (0.9%)
			BR-O	2.134 (0.099)	2.408 (0.488)	2.124 (0.066)	2.462 (0.299)	91% (3.2%)
2.2	2.902	200	BR-M	2.202 (0.053)	2.766 (0.197)	2.202 (0.031)	2.778 (0.109)	92% (0.9%)
			BR-O	2.221 (0.157)	2.642 (0.602)	2.221 (0.100)	2.789 (0.334)	88% (4.4%)
		400	BR-M	2.202 (0.041)	2.781 (0.146)	2.207 (0.024)	2.814 (0.082)	93% (0.4%)
			BR-O	2.232 (0.111)	2.797 (0.390)	2.230 (0.078)	2.852 (0.249)	90% (2.6%)

†: See Table 1.

benefit loss for one unit increase in risk and is computed as the regression coefficient of Y_i on R_i using the population data. Specifically, $\beta = 0.091$ and 0.093 for the two simulation settings, respectively. Following Thall et al. (2008), we fitted a bivariate copula model to estimate the joint distribution of efficacy and risk outcome including treatment by covariate interactions, and computed the expected utility given the joint distribution for each patient. The optimal treatment was estimated as the one leading to a higher utility value. Under simulation setting 1, the mean efficacy (sd) was 0.622 (0.015) and 0.634 (0.008) for $n = 200$ and $n = 400$, respectively. The mean risk (median absolute deviation) was 2.620 (0.078) and 2.629 (0.053), respectively. The median efficacy and risk were similar to the mean. For setting 2, the mean (sd) of the efficacy was 3.521 (0.017) and 3.526 (0.012), for the two sample sizes, respectively. The corresponding mean (sd) for the risk outcome was 2.563 (0.023) and 2.558 (0.017), respectively. These results show that the utility based method captures the optimal efficacy, but does not provide control over the risk at a particular level. Thus, the higher efficacy is achieved by allowing a higher risk, where the latter may potentially be over the tolerance bound.

Next, we explored a modified utility-based method that provides more control over individual risk outcome by imposing a safety admission rule: for each individual, we examine whether the previously estimated utility-based treatment rule will lead to a risk outcome (or estimated risk outcome) exceeding the threshold τ; if so, we select the safer treatment with a lower risk regardless of the efficacy. The results under the first simulation setting show that utility-based approach with safely admission rule leads to the risk outcome being controlled well below the threshold, at the price of a lower efficacy than BR-M and BR-O. For example, for setting 1, the mean efficacy across simulations was 0.016 and the mean risk was 1.49 when $\tau = 2.0$ and $n = 200$. When $\tau = 2.25$, the mean risk and efficacy was 1.50 and 0.0164, respectively. The efficacy is close to that of the non-personalized, "one-size-fits-all" rule (treatment effect of 0.003). This may be due to that the risk constraint is imposed at the individual level for this utility-based approach, while for BR-M and BR-O the risk is controlled on average (across patients). Thus, when strictly enforcing each individual's risk to be below τ is desirable, utility-based

method with a safety admission rule is preferable.

6 Application to DURABLE Trial

We applied our method to DURABLE study (Fahrbach et al., 2008) introduced in Section 2. A more detailed description of the study design was previously published (Fahrbach et al., 2008; Buse et al., 2009). A major objective of treating diabetes patients was to lower patients' blood glucose measured by A1C. Similar to the original report in Fahrbach et al. (2008), our efficacy endpoint (the benefit) was A1C change from baseline at 24 weeks (last observation carried forward [LOCF] to 24 weeks). Our safety (the risk) endpoint was measured by daily hypoglycemic event rate. We considered 18 relevant covariates measured at the baseline, including baseline A1C, fasting blood glucose, fasting insulin, adiponectin, blood pressure, 7 points self monitored blood glucose, duration of diabetes, weight, height, blood pressure, body mass index (BMI). The covariates were standardized before fitting the model and the tuning parameters for BR-O were chosen by 2-fold cross validation similar to the simulation studies.

We included randomized patients who received at least one treatment, which consisted of 965 patients on lispro mix and 980 patients on insulin glargine. Within a reasonable range, we examined different threshold values for daily hypoglecemic rates ($\tau = 0.063, 0.064, 0.065, 0.066, 0.067$) as our safety constraint. To compare the performance of different methods and evaluate the importance of the variables in this analysis, we randomly split the full data into a training set and a testing set, so the performance can be assessed using the testing set. Among 1,945 patients, we randomly selected 300 patients as a training set, and used the rest 1,645 patients as a testing set. We applied the proposed algorithms (BR-M, BR-O) to the training set, and obtained the average efficacy and safety outcomes under the estimated optimal treatment assignments on the testing set. Furthermore, to minimize influence on the variability due to selecting training and testing set by chance, we repeated this procedure 100 times.

We show in Table 3 the average benefit and risk obtained by BR-M and BR-O under risk constraints, and compare with results obtained by O-learning without controlling for the risk

(Zhao et al., 2012). We see that the average risks are reasonably controlled for both BR-M and BR-O at all constraint value of τ. BR-O produces a more conservative result in terms of a lower rate of hypoglycemia on the testing set as compared to BR-M, and thus a slightly lower average benefit. The average benefit increases as a function of threshold τ, indicating that allowing a more lenient risk control leads to gain in efficacy. The maximal benefit for the optimal treatment rules under O-learning without constraint ($\tau = \infty$) is the highest (1.738).

Table 3: DURABLE study analysis results. We randomly selected 300 patients as a training dataset, used the rest 1,645 patients as a testing dataset, and repeated this 100 times. The average benefit, risk and their standard deviations are shown.

Risk(τ)	Method	Risk-Training	Risk-Testing	Benefit-Training	Benefit-Testing
0.063	BR-M	0.0639(0.005)	0.0689(0.004)	1.8668(0.142)	1.7201(0.049)
	BR-O	0.0626(0.003)	0.0640(0.006)	1.7824(0.142)	1.6980(0.042)
0.064	BR-M	0.0644(0.006)	0.0690(0.004)	1.8682(0.141)	1.7209(0.050)
	BR-O	0.0632(0.003)	0.0650(0.006)	1.7905(0.135)	1.7004(0.050)
0.065	BR-M	0.0652(0.006)	0.0692(0.004)	1.8736(0.142)	1.7228(0.050)
	BR-O	0.0638(0.003)	0.0650(0.006)	1.7983(0.135)	1.7030(0.051)
0.066	BR-M	0.0657(0.006)	0.0694(0.004)	1.8780(0.146)	1.7241(0.051)
	BR-O	0.0644(0.003)	0.0655(0.006)	1.8048(0.135)	1.7021(0.046)
0.067	BR-M	0.0667(0.006)	0.0696(0.004)	1.8827(0.148)	1.7250(0.052)
	BR-O	0.0654(0.003)	0.0660(0.006)	1.8273(0.131)	1.7093(0.048)
∞	BR-M	0.0756(0.010)	0.0712(0.003)	1.9392(0.153)	1.7378(0.048)
	BR-O	0.0769(0.010)	0.0714(0.005)	1.9895(0.146)	1.7360(0.052)

Since a linear kernel was used for the BR-O algorithm, the treatment decision boundary can be expressed as a linear combination of the covariates. The standardized effects can be used to compare influences of the covariates on the decision boundary and rank the importance of covariates (absolute value of the standardized coefficient). The ranking based on covariates' average effects from 100 random splits indicates that the top 3 most important variables are, in turn, baseline fasting blood glucose, BMI and A1C. The variables with highest importance are not sensitive to the threshold of the risk constraint. The less important variables can change

order in the presence and absence of risk constraint (e.g., fasting insulin and heart rate).

To determine an appropriate risk bound for the final model, we control the rate of hypoglycemic event to be below the observed incidence rate in the low risk treatment arm (insulin glargine) in DURABLE trial, an incidence rate of 0.5 per week for each patient (Buse et al., 2009, Table 2, a total of 530 incidences per week for 1,046 patients), which corresponds to 2 hypoglycemic events per month (i.e., $\tau = 0.065$). To further illustrate an empirical method to choose the threshold, we adapt the net clinical benefit (NCB) which was proposed to weigh benefits against harms of a treatment based on the difference between expected benefit and expected harm (Lynd, 2006; Sutton et al., 2005). Specifically, the NCB at each threshold was computed as $Y - 1.5R$, where Y was the average efficacy obtained on the testing set using BR-O, and R was the corresponding average risk on the testing set (Table 3). The scaled association coefficient of 1.5 was obtained from the full data, similar to the method used for constructing the utility function. The increase in NCB with increasing threshold was 0.0026, 0.0029, -0.0024, and 0.0098 for τ changing from 0.063 to 0.067 at an increment of $\Delta\tau = 0.001$. Thus, when $\tau = 0.065$ the change in NCB is negative for the first time, i.e., the change in benefit does not outweigh the change in risk. This analysis further supports the choice of $\tau = 0.065$ as the threshold.

Next, we present the estimated optimal treatment rule obtained by BR-O and O-learning using all 1,945 patients with $\tau = 0.065$. To visualize the influence of top ranking variables on the personalized treatment rules, in Figure 4 we plot the decision boundary as a function of two covariates with A1C values fixed at pre-specified levels and all other covariates fixed at sample means. The proportion of patients recommended to take mix 75/25 based on BR-O (40%, 33%, 23%) is decreasing with increasing A1C (7.8%, 8.8%, 10.3%). Similar trend was also found for O-learning (62%, 55%, 44%). From Figure 4, we observe that for patients with a higher BMI, mix 75/25 is recommended. This finding is consistent with clinical knowledge on the mechanism of these two treatments: because patients with a higher BMI often have more food intake, a treatment that can more efficiently lower the post prandial blood glucose, namely mix 75/25,

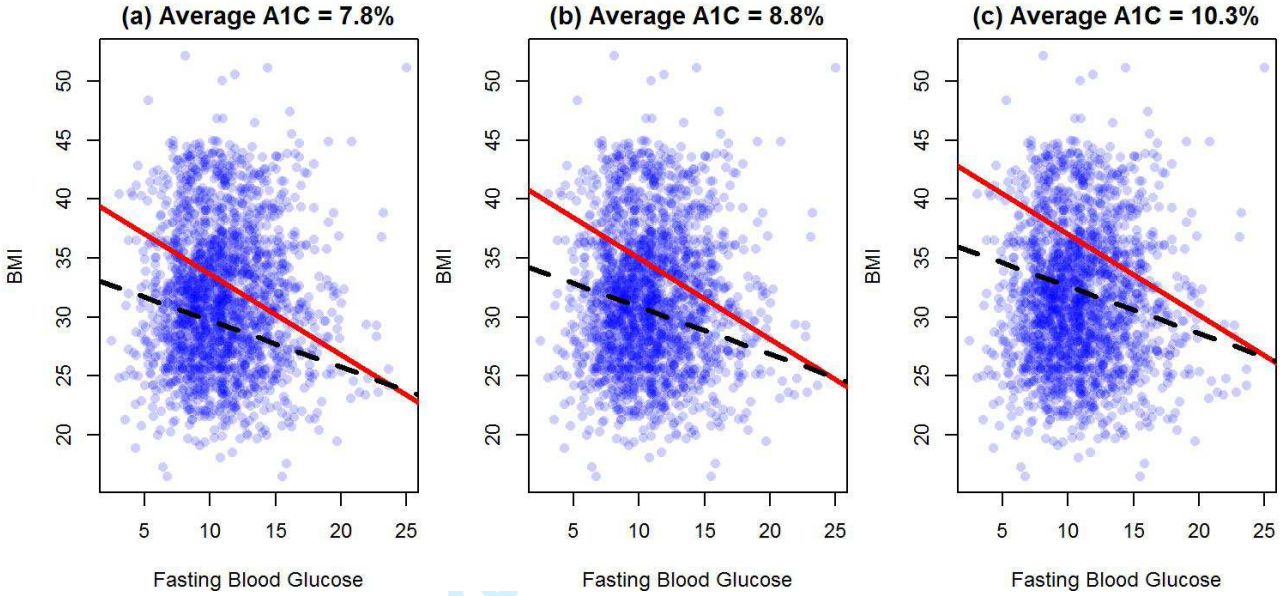
is more desirable. Comparing the change of slopes for the decision boundary with and without risk constraint, Figure 4 indicates that when considering hypoglycemic events, fasting blood glucose plays a more important role. In addition, the two decision boundaries become more divergent when the fasting blood glucose is low. This observation reflects the fact that patients with a low blood glucose level are at a higher risk of experiencing hypoglycemia events, which is consistent with the diabetes treatment guidance (American Diabetes Association and others, 2014).

Table 4: DURABLE study analyses results: ranking of baseline biomarkers based on average standardized effects over 100 repetitions.

$\tau =$	0.063	0.064	0.065	0.066	0.067	∞
Baseline A1C	1	1	1	1	1	1
BMI	2	2	2	2	2	2
Fasting Blood Glucose	3	3	3	3	3	3
Height	4	4	4	4	4	4
Adiponectin	5	5	5	5	5	5
Duration of diabetes	6	6	6	6	6	6
Body Weight	7	7	7	7	7	7
Diastolic blood pressure	8	8	8	8	8	8
Fasting Insulin	9	9	9	9	9	10
Heart rate	10	10	10	10	10	9
Systolic blood pressure	11	11	11	11	11	11
Glucose:Morning before meal	12	12	12	12	12	12
Glucose: 3am at night	13	13	13	13	13	14
Glucose:Evening before meal	14	14	14	14	14	13
Glucose:Morning 2 hours after meal	15	15	16	15	15	16
Glucose:Evening after meal	16	16	15	16	16	15
Glucose:Noon before meal	17	17	18	17	17	18
Glucose:Noon 2 hours after meal	18	18	17	18	18	17

Lastly, when comparing with the utility based method, the utility function was computed as $U(Y, R) = Y - \beta R$, where $\beta = 0.0041$ is the regression coefficient of Y on R using the full data. Thus, on average one unit increase in the risk is reflected as 0.0041 unit decrease in the benefit. Following Thall et al. (2008), we first estimated the joint distribution of Y and R using a copula model including treatment and covariate interactions. Using this joint distribution, we then

Figure 4: Estimated optimal treatment decision boundaries (based on all subjects) stratified by baseline A1C†.



†: Except BMI, fasting blood glucose and A1C, all other covariates are fixed at sample mean level. Red solid line: BR-O ($\tau = 0.065$). Black dahsed line: O-learning without risk constraint. Patients above the lines are recommended to take mix 75/25 and patients below the lines are recommended to take insulin glargine.

calculated the expected utility function and maximized it to find the optimal treatment rule. The estimated optimal treatment rule achieves an average efficacy of 1.742 (sd=0.049), similar to O-learning without the risk constraint. However, similar to that observed in the simulation studies, the average risk is also higher (0.0704, sd=0.003) than BR-M or BR-O. Therefore, the utility function based method leads to a higher efficacy, but at the price of a higher risk.

7 Discussion

7.1 Concluding remarks

In this work, we introduce a risk constraint to the estimation of optimal personalized treatment rule, so that the identified rule not only maximizes efficacy but also controls the average risk to be below a pre-specified threshold. We have proposed two methods to the constrained optimization, of which BR-M relies on valid models for both efficacy and safety outcomes while BR-O directly maximizes an approximation to the objective function without modelling. In

our simulation and data analysis, we used linear models in BR-M, so if the linear model is misspecified, BR-M may not always control the risk in a testing dataset. In contrast, as seen from our numerical studies, BR-O controls the risk at the pre-specified level on both the training and testing data, and maintains robustness against the model misspecification of $\delta_R(X)$. From a computational perspective, BR-M only requires fitting two regression models and solving a single-parameter monotone optimization problem, which is fast. For BR-O, the computation can be improved by using reasonable initial values, for example, those estimated under the hinge loss.

In our approaches, the choice of threshold value for controlling the risk is important. Ideally, the threshold should be a clinically meaningful safety/risk bound specified by clinicians or policy makers. When such a clinically meaningful bound is not available, our method provides a complete picture of the trade-off between benefit and risk for a range of threshold values. In our application example, several thresholds were examined, and the final bound was controlled at the average level observed in the low risk arm. Furthermore, we illustrate using the change in NCB (Lynd, 2006; Sutton et al., 2005) as an empirical guide to determine the threshold. A limitation of these proposals to find the threshold is that they do not take into account the variability rising from estimating the optimal rule. A better alternative is to use bootstrap resampling to examine the probabilistic behavior of the risk outcome (for example, whether the mean or median risk across bootstrap samples is below the threshold bound). A Bayesian approach may also be suitable. When two optimal treatment rules have similar efficacy but different safety which are both under the bound, we propose to select the treatment strategy that achieves the smallest threshold.

Alternative Bayesian approaches to handle multivariate benefit and risk outcomes include defining a meaningful utility function as exemplified in Houede et al. (2010); Thall et al. (2008); Thall (2012), although the definition of a utility function may be difficult in cases without a consensus. As demonstrated in the simulation study and application to our motivating example, using utility function may yield a higher benefit but at a price of an increased risk potentially

higher than the threshold. A solution is to impose a more strict safety admission rule at the individual level as in Section 5. However, as observed in the simulations, it may result in a conservative treatment rule with a low risk but at the price of a lower efficacy.

There is a link between the utility based approaches and BR-O: by re-defining the value function as negative infinity when the average risk is greater than the bound, the risk constraint can be directly enforced into the value function. However, the re-defined value function will be non-smooth with a singleton component. Thus, computation may be more difficult and potentially causing numeric instability.

7.2 Extension to multiple risk constraints

Patients' perspective on acceptable risk threshold may depend on the goal of the treatment they are receiving (e.g., disease prevention, chronic treatment), their genetic risk factors, or their perception of susceptibility of a disease. Thus, it is desirable to include multiple thresholds or risk constraints in estimating the optimal treatment depending on patient-specific features or preferences. Extensions to handle multiple constraints can be incorporated in BR-O learning framework, although it requires determination of thresholds for multiple risk outcomes.

Consider K types of risk outcomes, denoted by $R^{(1)}, ..., R^{(K)}$, each with a threshold value $\tau^{(1)}, ..., \tau^{(K)}$, respectively. Then our method can be extended to solve the following optimization problem:

$$\max_f E\left\{\frac{I(Af(X)\geq 0)Y}{p(A|X)}\right\},$$
$$\text{subject to } E\left\{\frac{I(Af(X)\geq 0)}{p(A|X)}R^{(k)}\right\}\leq \tau_k, \quad k=1,...,K.$$

Following the same derivation as in Theorem 1, by introducing K Lagrange-multipliers, we can show that the optimal decision rule takes a similar form as described in Theorem 1 with λ^* replaced by outcome-specific λ_k^*. BR-O as given in (4.4) can also be extended to include multiple constraints in the optimization.

Finally, we also note that the same procedure is applicable to control group-specific risks, in

which different risk bounds may be imposed on difference subgroups of patients. For example, patients who are more susceptible to adverse events may require tighter control on their risk outcomes. In this case, the above procedure is applicable when each constraint is conditioned on the corresponding subgroup.

7.3 Extension to multiple treatments

The same idea in this paper can be extended to the applications with multiple treatment arms. Assume that A has m treatment levels so that the decision function $\mathcal{D}(X)$ maps X to one of these m levels. The optimization problem in (3.2) becomes

$$\max_{\mathcal{D}} \sum_{j=1}^{m} E\left\{E[Y|A=j,X]I(\mathcal{D}(X)=j)\right\},$$

subject to the constraint

$$\sum_{j=1}^{m} E\left\{E[R|A=j,X]I(\mathcal{D}(X)=j)\right\} \leq \tau.$$

Hence, BR-M can be extended by solving an empirical version of a constrained equation. To extend BR-O, we can replace the objective function in the optimization by a consistent continuous loss for multicategory learning (Lee et al., 2004; Liu and Yuan, 2011), where $\mathcal{D}(X)$ is replaced by a vector of decision functions $(f_1, ..., f_m)$ with each component corresponding to each level of A, and the indicator function in the constraint, which is equivalent to $(I(A=j)f_j(X) \geq 0)$, can be approximated by the shifted ψ-loss. A computational algorithm similar to BR-O can be carried out but will be more involved. Further investigation is warranted for implementation.

A Appendix: Proof of Theorem 1

If $E[\delta_R(X)^+|X \in \mathcal{M}^c] \leq \alpha^*$, then every $f(x)$ satisfies the constraint so the optimal $f^*(X) = \text{sign}(\delta_Y(X))$. Thus, for the following discussion, we assume $E[\delta_R(X)^+|X \in \mathcal{M}^c] > \alpha^*$. Suppose $f^*(x)$ to be the optimal solution in this region. We claim $E[\delta_R(X)I(f^*(X)>0)|X \in \mathcal{M}^c] = \alpha^*$. To see this, we note that $f^*(X)$ cannot have the same sign as $\delta_R(X)$ because otherwise,

$E[\delta_R(X)I(f(X) > 0)|X \in \mathcal{M}^c] = E[\delta_R^+(X)|X \in \mathcal{M}^c] > \alpha^*$. Thus, if $E[\delta_R(X)I(f^*(X) > 0)|X \in \mathcal{M}^c] < \alpha^*$, then we can change sign of $f^*(X)$ in a small region of X where $f^*(X)$ and $\delta_R(X)$ have the opposite signs so that $E[\delta_R(X)I(f^*(X) > 0)]$ is closer to α^*. However, since $\delta_Y(X)\delta_R(X) > 0$, this change will only increase the overall benefit. Therefore, in region $\mathcal{M}^c$, $f^*(x)$ solves

$$\max_f E\left\{\delta_Y(X)I(f(X) > 0)|X \in \mathcal{M}^c\right\} \quad \text{subject to } E[\delta_R(X)I(f(X) > 0)|X \in \mathcal{M}^c] = \alpha^*.$$

After introducing the Lagrange multiplier, $f^*(x)$ maximize,

$$E\left[\{\delta_Y(X) - \lambda\delta_R(X)\}I(f(X) > 0)|X \in \mathcal{M}^c\right]$$

subject to $E[\delta_R(X)I(f(X) > 0)|X \in \mathcal{M}^c] = \alpha^*$. Clearly,

$$I\{f^*(X) > 0\} = I\left[\{\delta_Y(X) - \lambda\delta_R(X)\} > 0\right],$$

where λ solves equation,

$$E\left[\delta_R(X)I\left(\{\delta_Y(X) - \lambda\delta_R(X)\} > 0\right)|X \in \mathcal{M}^c\right] = \alpha^*.$$

The last equation is equivalent to,

$$E\left[\delta_R(X)I\{\delta_R(X) > 0, \delta_Y(X)/\delta_R(X) > \lambda\}|X \in \mathcal{M}^c\right]$$
$$+E\left[\delta_R(X)I\{\delta_R(X) < 0, \delta_Y(X)/\delta_R(X) < \lambda\}|X \in \mathcal{M}^c\right] = \alpha^*.$$

The left-hand side of the equation is strictly decreasing in λ which is equal to $E[\delta_R(X)^+]$ if $\lambda = 0$ and $E[\delta_R(X)I\{\delta_R(X) < 0\}|X \in \mathcal{M}^c] \leq \alpha^*$ if $\lambda = \infty$. Thus, this equation has a unique solution denoted by λ^*. Consequently, the optimal treatment regime $f^*(X)$ has the same sign as sign $(\delta_Y(X) - \lambda^*\delta_R(X))$.

Interestingly, when $X \in \mathcal{M}$ where $\delta_Y(X)$ and $\delta_R(X)$ take opposite signs, we note that the sign of $\delta_Y(X)$ is the same as the sign of $\delta_Y(X) - \lambda^*\delta_R(X)$. So we have the results of Theorem 1.

B Appendix: Detailed algorithm for BR-O-learning

Here we describe the detailed computational algorithm for solving (4.4). At each iteration of the DC algorithm, after introducing additional slack variable $\zeta_i \geq A_i f(X_i) + \delta$ and $\zeta_i \geq 0, \forall i$, we have,

$$\begin{cases} \beta^{(l+1)} \in \arg\min_\beta & \sum_{i=1}^n \left(C\frac{Y_i^*}{p_i}\xi_i + C/n\zeta_i \right) + \frac{1}{2}\beta_{(0)}^T \mathbf{K}\beta_{(0)}, \\ \text{subject to} & \sum_{i=1}^n \frac{R_i}{p_i}\left\{\zeta_i - \widehat{v}_i(\beta, \beta^{(l)})\right\} \leq \delta n\tau, \\ & \xi_i \geq 1 - A_i^*\big\{\beta_0 + \sum_{j=1}^n \beta_j K(X_i, X_j)\big\}, \xi_i \geq 0 \\ & \zeta_i \geq \delta + A_i\big\{\beta_0 + \sum_{j=1}^n \beta_j K(X_i, X_j)\big\}, \zeta_i \geq 0, \quad \forall i, \end{cases}$$

where $\widehat{v}_i(\beta, \beta^{(l)}) = \{A_i f(X_i|\beta)\} I_i^{(l)}$ and $I_i^{(l)} = I\{A_i f(X_i|\beta^{(l)}) > 0\}$. Note that the presence of ξ_i in the objective function guarantees that the optimal ζ_i should be equal to

$$\left(\delta + A_i\big\{\beta_0 + \sum_{j=1}^n \beta_j K(X_i, X_j)\big\} \right)_+ .$$

However, since we only weigh the summation of ζ_i by a small weight C/n. The objective in the above optimization is approximately equivalent to (4.4).

The Lagrange (primal) function is,

$$\begin{aligned} L_P \;=\; & C\sum_{i=1}^n \left(\frac{Y_i^*}{p_i}\xi_i + n^{-1}\zeta_i \right) + \frac{1}{2}\beta_{(0)}^T \mathbf{K}\beta_{(0)} + \pi\left[\sum_{i=1}^n \frac{R_i}{p_i}\left\{\zeta_i - \widehat{v}_i(\beta, \beta^{(l)})\right\} - \delta n\tau \right] \\ & - \sum_{i=1}^n \alpha_i \left[\xi_i - 1 + A_i^*\big\{\beta_0 + \sum_{j=1}^n \beta_j K(X_i, X_j)\big\} \right] - \sum_{i=1}^n \mu_i \xi_i \\ & - \sum_{i=1}^n \kappa_i \left[\zeta_i - \delta - A_i\big\{\beta_0 + \sum_{j=1}^n \beta_j K(X_i, X_j)\big\} \right] - \sum_{i=1}^n \rho_i \zeta_i. \end{aligned}$$

Taking the derivative with respect to ξ_i, ζ_i, β_0 and $\beta_{(0)}$, and set them equal to zero, we have,

$$\begin{aligned} C\frac{Y_i^*}{p_i} - \alpha_i - \mu_i &= 0, \\ \frac{C}{n} + \pi\frac{R_i}{p_i} - \kappa_i - \rho_i &= 0, \\ \pi\sum_{i=1}^n \frac{R_i}{p_i} A_i I_i^{(l)} + \sum_{i=1}^n \alpha_i A_i^* - \sum_{i=1}^n \kappa_i A_i &= 0, \\ \beta_i - \pi\frac{R_i}{p_i} A_i I_i^{(l)} - \alpha_i A_i^* + \kappa_i A_i &= 0, \quad \forall i = 1, \ldots, n. \end{aligned}$$

Let $\boldsymbol{\theta} = \left(\pi\frac{R_1}{p_1}A_1 I_1^{(l)} + \alpha_1 A_1^* - \kappa_1 A_1, \ldots, \pi\frac{R_n}{p_n}A_n I_n^{(l)} + \alpha_n A_n^* - \kappa_n A_n\right)^T$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$, $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)^T$, $C_* = \left(C\frac{Y_1^*}{p_1}, \ldots, C\frac{Y_n^*}{p_n}\right)^T$, $R_p = (R_1/p_1, \ldots, R_n/p_n)^T$, $A = (A_1, \ldots, A_n)^T$, $A^* = (A_1^*, \ldots, A_n^*)^T$, and $R_{pA}^I = (R_1/p_1 A_1 I_1^{(l)}, \ldots, R_n/p_n A_n I_n^{(l)})^T$. By substituting the equations and removing constants, the Lagrange (dual) problem for (4.4) is

$$\max_{\pi,\boldsymbol{\alpha},\boldsymbol{\kappa}} \quad -\frac{1}{2}\boldsymbol{\theta}^T\mathbf{K}\boldsymbol{\theta} + (\boldsymbol{\alpha} + \delta\boldsymbol{\kappa})^T\mathbf{1} - \delta\pi n\tau$$
$$\text{subject to} \quad \begin{cases} \mathbf{0} \preceq \boldsymbol{\alpha} \preceq C_*, \\ \boldsymbol{\kappa} - \pi R_p \preceq n^{-1}C\mathbf{1}, 0 \leq \pi, \mathbf{0} \preceq \boldsymbol{\kappa}, \\ \pi\mathbf{1}^T R_{pA}^I + \boldsymbol{\alpha}^T A^* - A^T\boldsymbol{\kappa} = 0, \end{cases}$$

where $\preceq$ is the by element inequality. Let $\boldsymbol{\omega} = (\pi, \boldsymbol{\alpha}^T, \boldsymbol{\kappa}^T)^T$, $\mathbf{1}^\delta = \{-\delta n\tau, \mathbf{1}^T, \delta\mathbf{1}\}^T$, $\mathbf{H} = \{R_{pA}^I, \text{diag}(A^*), \text{diag}(-A)\}$ which is a $n \times (2n+1)$ matrix,

$$\mathbf{W} = \begin{bmatrix} R_p & \text{diag}(\mathbf{0}_n) & \text{diag}(-\mathbf{1}_n) \\ \mathbf{1}^T R_{Ap}^I & (A^*)^T & -A^T \end{bmatrix}_{(n+1)\times(2n+1)} \quad \mathbf{b} = \begin{bmatrix} -C\mathbf{1}_n/n \\ 0 \end{bmatrix}_{(n+1)\times 1} \quad \mathbf{r} = \begin{bmatrix} \infty\mathbf{1}_n \\ 0 \end{bmatrix}_{(n+1)\times 1},$$

and $\mathbf{u} = (\infty, C_*^T, \infty\mathbf{1}_n^T)^T$. We end up to solve

$$\max_{\boldsymbol{\omega}} \quad -\frac{1}{2}\boldsymbol{\omega}^T(\mathbf{H}^T\mathbf{K}\mathbf{H})\boldsymbol{\omega} + \boldsymbol{\omega}^T\mathbf{1}^\delta$$
$$\text{subject to} \quad \begin{cases} \mathbf{b} \preceq \mathbf{W}\boldsymbol{\omega} \preceq \mathbf{b} + \mathbf{r}, \\ \mathbf{0}_{2n+1} \preceq \boldsymbol{\omega} \preceq \mathbf{u}. \end{cases}$$

In particular, package ipop can be used to obtain the solution.

References

American Diabetes Association and others (2014), "Standards of medical care in diabetes2014," *Diabetes Care*, 37, S14–S80.

Belle, D. J. and Singh, H. (2008), "Genetic factors in drug metabolism." *American family physician*, 77, 1553–1560.

Buse, J. B., Wolffenbuttel, B. H., Herman, W. H., Shemonsky, N. K., Jiang, H. H., Fahrbach, J. L., Scism-Bacon, J. L., and Martin, S. A. (2009), "DURAbility of basal versus lispro mix 75/25 insulin efficacy (DURABLE) trial 24-week results: safety and efficacy of insulin lispro

mix 75/25 versus insulin glargine added to oral antihyperglycemic drugs in patients with type 2 diabetes," *Diabetes Care*, 32, 1007–13.

Cai, T., Tian, L., Wong, P. H., and Wei, L. (2011), "Analysis of randomized comparative clinical trial data for personalized treatment selections," *Biostatistics*, 12, 270–282.

Control, D., Group, C. T. R., et al. (1997), "Hypoglycemia in the diabetes control and complications trial," *Diabetes*, 46, 271–286.

Cryer, P. E. (2002), "Hypoglycaemia: the limiting factor in the glycaemic management of Type I and Type II diabetes," *Diabetologia*, 45, 937–48.

Cryer, P. E., Davis, S. N., and Shamoon, H. (2003), "Hypoglycemia in diabetes," *Diabetes care*, 26, 1902–1912.

Fahrbach, J., Jacober, S., Jiang, H., and Martin, S. (2008), "The DURABLE trial study design: comparing the safety, efficacy, and durability of insulin glargine to insulin lispro mix 75/25 added to oral antihyperglycemic agents in patients with type 2 diabetes," *Journal of diabetes science and technology*, 2, 831–838.

Fidler, C., Elmelund Christensen, T., and Gillard, S. (2011), "Hypoglycemia: an overview of fear of hypoglycemia, quality-of-life, and impact on costs," *Journal of medical economics*, 14, 646–655.

Foster, J. C., Taylor, J. M., and Ruberg, S. J. (2011), "Subgroup identification from randomized clinical trial data," *Statistics in medicine*, 30, 2867–2880.

Group, U. P. D. S. U. et al. (1998), "Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33)," *The Lancet*, 352, 837–853.

Guo, J. J., Pandey, S., Doyle, J., Bian, B., Lis, Y., and Raisch, D. W. (2010), "A review of quantitative risk–benefit methodologies for assessing drug safety and efficacyreport of the ISPOR Risk–Benefit Management Working Group," *Value in Health*, 13, 657–666.

Houede, N., Thall, P. F., Nguyen, H., Paoletti, X., and Kramar, A. (2010), "Utility-Based Optimization of Combination Therapy Using Ordinal Toxicity and Efficacy in Phase I/II Trials," *Biometrics*, 66, 532–540.

Huang, X., Shi, L., and Suykens, J. A. (2014), "Ramp loss linear programming support vector machine," *The Journal of Machine Learning Research*, 15, 2185–2211.

Inzucchi, S. E., Bergenstal, R. M., Buse, J. B., Diamant, M., Ferrannini, E., Nauck, M., Peters, A. L., Tsapas, A., Wender, R., and Matthews, D. R. (2012), "Management of hyperglycemia in type 2 diabetes: a patient-centered approach position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD)," *Diabetes care*, 35, 1364–1379.

Jaakkola, T., Diekhans, M., and Haussler, D. (1999), "Using the Fisher kernel method to detect remote protein homologies." in *ISMB*, vol. 99, pp. 149–158.

Kosorok, M. R. and Moodie, E. E. (2015), *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*, vol. 21, SIAM.

Laber, E. B., Lizotte, D. J., and Ferguson, B. (2014), "Set-valued dynamic treatment regimes for competing outcomes," *Biometrics*, 70, 53–61.

Lee, J., Thall, P. F., Ji, Y., and Müller, P. (2015), "Bayesian dose-finding in two treatment cycles based on the joint utility of efficacy and toxicity," *Journal of the American Statistical Association*, 110, 711–722.

Lee, Y., Lin, Y., and Wahba, G. (2004), "Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data," *Journal of the American Statistical Association*, 99, 67–81.

Lipkovich, I., Dmitrienko, A., Denne, J., and Enas, G. (2011), "Subgroup identification based on differential effect search a recursive partitioning method for establishing response to treatment in patient subpopulations," *Statistics in medicine*, 30, 2601–2621.

Liu, Y., Wang, Y., Kosorok, M., Zhao, Y., and Zeng, D. (2014), "Robust Hybrid Learning for Estimating Personalized Dynamic Treatment Regimens," *arXiv preprint arXiv:1611.02314*.

Liu, Y. and Yuan, M. (2011), "Reinforced multicategory support vector machines," *Journal of Computational and Graphical Statistics*, 20, 901–919.

Lizotte, D. J., Bowling, M., and Murphy, S. A. (2012), "Linear fitted-q iteration with multiple reward functions," *Journal of Machine Learning Research*, 13, 3253–3295.

Luedtke, A. R., Van Der Laan, M. J., et al. (2016), "Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy," *The Annals of Statistics*, 44, 713–742.

Lynd, L. (2006), "Quantitative Methods for Therapeutic Risk-Benefit Analysis," in *11th International Meetings of the International Society for Pharmacoeconomics and Outcomes Research*, Issues Panel: Health outcomes approaches to risk-benefit analysis: how ready are they.

Moore, R. A., Derry, S., McQuay, H. J., and Paling, J. (2008), "What do we know about communicating risk? A brief review and suggestion for contextualising serious, but rare, risk, and the example of cox-2 selective and nonselective NSAIDs," *Arthritis Research and Therapy*, 10, R20.

Natarajan, B. K. (1995), "Sparse approximate solutions to linear systems," *SIAM journal on computing*, 24, 227–234.

Qian, M. and Murphy, S. A. (2011), "Performance guarantees for individualized treatment rules," *Annals of statistics*, 39, 1180.

Quilliam, B. J., Simeone, J. C., Ozbay, A. B., and Kogut, S. J. (2011), "The incidence and costs of hypoglycemia in type 2 diabetes." *The American journal of managed care*, 17, 673–680.

Sinclair, A., Dunning, T., and Rodriguez-Mañas, L. (2015), "Diabetes in older people: new insights and remaining challenges," *The Lancet Diabetes & Endocrinology*, 3, 275–285.

Su, X., Tsai, C.-L., Wang, H., Nickerson, D. M., and Li, B. (2009), "Subgroup analysis via recursive partitioning," *The Journal of Machine Learning Research*, 10, 141–158.

Sutton, A. J., Cooper, N. J., Abrams, K. R., Lambert, P. C., and Jones, D. R. (2005), "A Bayesian approach to evaluating net clinical benefit allowed for parameter uncertainty," *Journal of clinical epidemiology*, 58, 26–40.

Tao, P. D. and An, L. T. H. (1998), "A DC optimization algorithm for solving the trust-region subproblem," *SIAM Journal on Optimization*, 8, 476–505.

Thall, P. F. (2012), "Bayesian adaptive dose-finding based on efficacy and toxicity," *J Statistical Research*, 14, 187–202.

Thall, P. F., Nguyen, H. Q., and Estey, E. H. (2008), "Patient-Specific Dose Finding Based on Bivariate Outcomes and Covariates," *Biometrics*, 64, 1126–1136.

Wu, Y., Zhang, H. H., and Liu, Y. (2010), "Robust model-free multiclass probability estimation," *Journal of the American Statistical Association*, 105.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012), "A robust method for estimating optimal treatment regimes," *Biometrics*, 68, 1010–1018.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012), "Estimating individualized treatment rules using outcome weighted learning," *Journal of the American Statistical Association*, 107, 1106–1118.