Stat

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.11

Estimating optimal treatment regimes from a classification perspective

Baqun Zhang[a,*], Anastasios A. Tsiatis[b], Marie Davidian[b], Min Zhang[c] and Eric Laber[b]

Received 8 October 2012; Accepted 29 October 2012

A treatment regime maps observed patient characteristics to a recommended treatment. Recent technological advances have increased the quality, accessibility, and volume of patient-level data; consequently, there is a growing need for powerful and flexible estimators of an optimal treatment regime that can be used with either observational or randomized clinical trial data. We propose a novel and general framework that transforms the problem of estimating an optimal treatment regime into a classification problem wherein the optimal classifier corresponds to the optimal treatment regime. We show that commonly employed parametric and semi-parametric regression estimators, as well as recently proposed robust estimators of an optimal treatment regime can be represented as special cases within our framework. Furthermore, our approach allows any classification procedure that can accommodate case weights to be used without modification to estimate an optimal treatment regime. This introduces a wealth of new and powerful learning algorithms for use in estimating treatment regimes. We illustrate our approach using data from a breast cancer clinical trial.

Keywords: classification; doubly robust estimator; inverse probability weighting; personalized medicine; potential outcomes; propensity score

1 Introduction

The goal of personalized medicine is to inform clinical interventions using individual patient characteristics. These characteristics may include patient demographics, genetic or genomic information, treatment and outcome history, ability to cope with side-effect burden, and so on. Personalized medicine has the potential to increase the quality of patient care while reducing cost by reducing over-treatment and making efficient use of all existing information. There is currently a great deal of interest among clinical and intervention scientists in the development of evidence-based personalized treatment strategies, also known as treatment regimes. With the increasing volume, accessibility, and quality of patient level data, statistics has an important role to play in the estimation and evaluation of treatment regimes.

Formally, a treatment regime is a rule that assigns a treatment, from among a set of possible treatments, to a patient based on his/her observed characteristics. Deducing optimal treatment regimes using data from a clinical trial or observational study can be informed, for example, in traditional regression-based methods, by identifying patient

[a]**Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, USA**
[b]**Department of Statistics, North Carolina State University, Raleigh, NC, 27695-8203, USA**
[c]**Department of Biotatistics, University of Michigan, Ann Arbor, MI 48109-2029, USA**
***Email: baqun.zhang@northwestern.edu**

covariates that exhibit a qualitative interaction with treatment assignment; i.e., an interaction in which the treatment effect changes direction depending on the covariates (Gunter et al., 2011).

Recently, there has been vigorous research on estimating optimal treatment regimes involving a single decision or a series of decisions based on data from clinical trials or observational studies (Murphy, 2003; Robins, 2004; Moodie et al., 2007; Robins et al., 2008; Brinkley et al., 2009; Zhao et al., 2009; Henderson et al., 2010; Orellana et al., 2010; Gunter et al., 2011). Much of this work involves postulating a model for the regression of outcome on treatment assignment and covariates, and then inferring from the model the best treatment assignment given patient covariates. Zhang et al. (2012a,b) proposed a robust method that maximizes across all regimes in a prespecified class a doubly robust augmented inverse probability weighted estimator (AIPWE) of the population mean outcome. This method achieves comparable performance to methods based on direct outcome regression modeling and is more robust to misspecification of regression models. With this method, as well as the recent inverse probability weighted estimator of Zhao et al. (2012) and methods based on outcome regression, the parametric form of regimes has to be pre-specified, either by practical considerations or through ad hoc preliminary data analysis.

In this article, we present a novel and general framework that facilitates flexible estimation of optimal treatment regimes in the single decision point setting. Specifically, we recast the original problem of finding the optimal treatment regime as a weighted classification problem and estimate the optimal treatment regime by estimating the Bayes classifier, i.e., the one that minimizes the expected weighted misclassification error. This framework allows the estimation of mean outcomes under a regime using any existing methods, e.g., regression estimator, inverse probability weighted estimator (IPWE) or AIPWE, and is separated from the subsequent optimization for identifying the form of treatment regimes, giving rise to its flexibility. Within this framework, the class of treatment regimes does not need to prespecified and can instead be identified in a data-driven way by minimizing an expected weighted misclassification error. Importantly, our approach allows for existing classification algorithms to be used without modification to estimate an optimal treatment regime. This introduces a wealth of new and powerful learning algorithms for use in estimating optimal treatment regimes.

The remainder of this paper is organized as follows. In Section 2, we formalize the problem of estimating the optimal treatment regime using potential outcomes and review existing methods. In Section 3, we present a general classification framework for identification of the optimal treatment regime. We conduct a small empirical study of the proposed method in Section 4. We illustrate the proposed method using data from the National Adjuvant Breast and Bowel Project (NSABP) in Section 5. Concluding remarks are made in Section 6.

2 Framework and methods

Consider a clinical trial or observational study where n subjects from a population of interest received one of two treatment options, denoted by $A = 0$ or 1. Let Y denote the observed outcome of interest and, without loss of generality, assume that larger values of Y are preferred. Let X denote the vector of patient characteristics collected prior to treatment. The observed data are then (X_i, A_i, Y_i), $i = 1, \ldots, n$, which are assumed to be independent and identically distributed (i.i.d.) across i.

A treatment regime, g, is a map from the domain of X to the domain of A. The goal is to use the data to estimate the optimal treatment regime, defined as the one that maximizes the expected outcome if used to assign treatments to all patients in the population of interest. To precisely define and identify the optimal treatment regime, we adopt the potential outcome framework (Rubin, 1978). Let $Y^*(0)$ and $Y^*(1)$ denote the potential outcomes for a subject that would be observed had the subject received treatment 0 or 1, respectively. We assume that the actual observed outcome is connected to the potential outcomes through $Y = Y^*(1)A + Y^*(0)(1 - A)$; this is usually referred to as

(wileyonlinelibrary.com) DOI: 10.1002/sta4.11

the consistency assumption and states that the observed outcome is the same as the potential outcome under the treatment actually received. We assume that there is no interference among units, also known as the stable unit treatment assumption (SUTVA). We further assume $\{Y^*(0), Y^*(1)\} \perp\!\!\!\perp A|X$, where $\perp\!\!\!\perp$ denotes statistical independence; this states that there are no unmeasured confounders, and that treatment A, conditional on X, can be viewed as being randomly assigned. In a randomized clinical trial, this assumption is trivially true. Under these assumptions, it is straightforward to show that the overall population mean were all patients in the population to receive treatment a, $E\{Y^*(a)\}$, is equal to $E_X[E\{Y|A=a,X\}]$, where $E_X(\cdot)$ denotes expectation with respect to the marginal distribution of X. Thus, for an arbitrary treatment regime g, the potential outcome for a subject randomly chosen from the population, if he/she were to receive treatment according to g, can be defined as

$$Y^*(g) = Y^*(1)g(X) + Y^*(0)\{1-g(X)\}.$$

The optimal regime, g^{opt}, is defined as the one yielding the largest value of $E\{Y^*(g)\}$ among the class of all potential regimes, $\mathcal{G}$; i.e., $g^{opt} = \arg\max_{g\in\mathcal{G}} E\{Y^*(g)\}$. Writing $\mu(a,X) = E(Y|A=a,X)$, it is straightforward to show that

$$E\{Y^*(g)\} = E_X[\,\mu(1,X)g(X) + \mu(0,X)\{1-g(X)\}\,],$$

and hence the optimal treatment regime is given by

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

An intuitive approach to estimating the optimal treatment regime, which we refer to as the regression method, is to posit a parametric regression model for $\mu(A,X) = E(Y|A,X)$, say $\mu(A,X;\beta)$. If the model is correctly specified, then $\mu(A,X) = \mu(A,X;\beta_0)$ for some β_0, and the optimal regime is therefore $g(X,\beta_0)$, where $g(X,\beta) = I\{\mu(1,X,\beta) > \mu(0,X,\beta)\}$. Hence, it is natural to estimate the optimal treatment regime by $\hat{g}^{opt}_{reg}(X) = I\{\mu(1,X,\hat{\beta}) > \mu(0,X,\hat{\beta})\}$, where $\hat{\beta}$ is an estimator of β. Clearly, if the model for $\mu(A,X)$ is incorrectly specified, $\hat{g}^{opt}_{reg}(X)$ may not be a good estimator of $g^{opt}(X)$.

Alternatively, a semiparametric version of the regression method, G-estimation (Robins, 2004), considers a semiparametric model for $\mu(A,X)$, exploiting the fact that the optimal treatment regime $g^{opt}(X)$ only depends on the contrast function $C(X) = \mu(1,X) - \mu(0,X)$ through $g^{opt}(X) = I\{C(X) > 0\}$. Specifically, G-estimation posits a semiparametric model $\mu(A,X) = h_1(X) + AC_G(X;\psi)$, where ψ is a finite dimensional vector and $h_1(X)$ is unspecified. The estimator $\hat{\psi}$ for ψ can be found by solving appropriate estimating equations involving a known or estimated propensity score $\pi(X) = \mathrm{pr}(A=1|X)$ (Robins, 2004; see also Schulte et al., 2012). The optimal treatment regime is estimated by $\hat{g}^{opt}_G(X) = I\{C_G(X;\hat{\psi}) > 0\}$. Like $\hat{g}^{opt}_{reg}(X)$, the quality of the G-estimation estimator $\hat{g}^{opt}_G(X)$ depends on how close $C_G(X;\psi)$ is to the true contrast function.

Recognizing that the posited regression model may be misspecified, Zhang et al. (2012a) instead considered such a posited regression model as a mechanism for defining a class of induced treatment regimes. They estimated the optimal regime within a pre-specified class by directly maximizing a doubly robust AIPWE of the population mean outcome across all regimes in the class. The class of regimes $\mathcal{G}_\eta$, indexed by parameter η, can be derived from a regression model $\mu(A,X;\beta)$, in which case η is a many-to-one function of β (see Zhang et al., 2012a, for details) or directly specified as depending on a key subset of elements of X based on practical considerations. The value $\eta^{opt} = \arg\max_\eta E\{Y^*(g_\eta)\}$, $g_\eta \in \mathcal{G}_\eta$, defines the optimal regime in $\mathcal{G}_\eta$, i.e., $g^{opt}_\eta(X) = g(X,\eta^{opt})$, which equals $g^{opt}(X)$ if $\mathcal{G}_\eta$ contains $g^{opt}(X)$ and, although not the same as $g^{opt}(X)$ if $g^{opt}(X)$ is not in $\mathcal{G}_\eta$, is still of considerable interest when we focus our attention on the feasible class $\mathcal{G}_\eta$. For fixed η, the AIPWE for $E\{Y^*(g_\eta)\}$ is given by

$$AIPWE(\eta) = n^{-1}\sum_{i=1}^{n}\left\{\frac{\mathcal{C}_{\eta,i}Y_i}{\pi_c(X_i;\eta,\hat{\gamma})} - \frac{\mathcal{C}_{\eta,i}-\pi_c(X_i;\eta,\hat{\gamma})}{\pi_c(X_i;\eta,\hat{\gamma})}m(X_i;\eta,\hat{\beta})\right\}, \quad (1)$$

where $\mathcal{C}_\eta = Ag(X,\eta) + (1-A)\{1-g(X,\eta)\}$, $\pi(X;\gamma)$ is a posited model for the propensity score $\pi(X)$; $\hat{\gamma}$ is the maximum likelihood (ML) estimator for γ; $\pi_c(X;\eta,\hat{\gamma}) = \pi(X;\hat{\gamma})g(X,\eta) + \{1-\pi(X;\hat{\gamma})\}\{1-g(X,\eta)\}$; $m(X;\eta,\beta) = \mu(1,X,\beta)g(X,\eta) + \mu(0,X,\beta)\{1-g(X,\eta)\}$ is a model for $E\{Y^*(g_\eta)|X\} = \mu(1,X)g(X,\eta) + \mu(0,X)\{1-g(X,\eta)\}$; $\mu(A,X;\beta)$ is a model for $E(Y|A,X)$; and $\hat{\beta}$ is an estimator for β. Denoting the value that maximizes $AIPWE(\eta)$ by $\hat{\eta}^{opt}_{AIPWE}$, which estimates η^{opt}, one can then estimate $g^{opt}_\eta(X)$ by $\hat{g}^{opt}_{\eta,AIPWE}(X) = g(X,\hat{\eta}^{opt}_{AIPWE})$.

As an alternative to the AIPWE estimator, Zhang et al. (2012a) also discussed the inverse probability weighted estimator (IPWE) for $E\{Y^*(g_\eta)\}$, given by

$$IPWE(\eta) = n^{-1}\sum_{i=1}^{n}\frac{\mathcal{C}_{\eta,i}Y_i}{\pi_c(X_i;\eta,\hat{\gamma})}.$$

The definitions of the estimators $\hat{\eta}^{opt}_{IPWE}$ and $\hat{g}^{opt}_{\eta,IPWE}(X)$ based on $IPWE(\eta)$ follow immediately. The method of Zhao et al. (2012) estimates the optimal treatment regime by maximizing a concave relaxation of the above IPWE estimator. This relaxation is analogous to the use of surrogate or proxy loss functions in classification (see, for example, Hastie et al., 2009); this relaxation provides numerical stability and facilitates efficient computation of the maximum.

3 Treatment regimes and classification

In this section, we describe a general framework for transforming the problem of estimating an optimal treatment regime into weighted classification problem. Using the previous notation, we see that

$$\begin{aligned} E\{Y^*(g)\} &= E_X[\,\mu(1,X)g(X) + \mu(0,X)\{1-g(X)\}] \\ &= E_X[\,g(X)\{\mu(1,X)-\mu(0,X)\} + \mu(0,X)] \\ &= E\{g(X)C(X)\} + E\{\mu(0,X)\}, \end{aligned}$$

so that $g^{opt} = \arg\max_{g\in\mathcal{G}} E\{Y^*(g)\} = \arg\max_{g\in\mathcal{G}} E\{g(X)C(X)\}$. A natural strategy for estimating the optimal treatment regime is to first construct an estimator $\hat{C}$ for C using the observed data, and estimate g^{opt} by $\hat{g}^{opt} = \arg\max_{g\in\mathcal{G}} n^{-1}\sum_{i=1}^{n} g(X_i)\hat{C}_i(X_i)$. Next we will see that all of the estimators discussed in the preceding section can be seen as following this approach.

The regression method posits a model $E(Y|A,X) = \mu(A,X;\beta)$, that defines the class of treatment regimes, $\mathcal{G}_\beta$, indexed by β, with elements of the form $g(X,\beta) = I\{\mu(1,X,\beta) > \mu(0,X,\beta)\}$. It then estimates $g^{opt}(x)$ by $\hat{g}^{opt}_{reg}(x) = I\{\mu(1,x,\hat{\beta}) > \mu(0,x,\hat{\beta})\}$, which is equivalent to $\arg\max_{g\in\mathcal{G}_\beta} n^{-1}\sum_{i=1}^{n} g(X_i)\hat{C}_{reg}(X_i)$, where $\hat{C}_{reg}(x) = \mu(1,x,\hat{\beta}) - \mu(0,x,\hat{\beta})$ is a regression estimator of $C(x)$.

G-estimation directly models the contrast function, which subsequently defines the class of treatment regimes, $\mathcal{G}_\psi$, indexed by ψ, with elements of the form $g(X,\psi) = I\{C_G(X,\psi) > 0\}$. The resulting estimator $\hat{g}^{opt}_G(x) = I\{C_G(x,\hat{\psi}) > 0\}$ is thus equal to $\arg\max_{g\in\mathcal{G}_\psi} n^{-1}\sum_{i=1}^{n} g(X_i)\hat{C}_G(X_i)$, where $\hat{C}_G(x) = C_G(x,\hat{\psi})$.

The robust method in Zhang et al. (2012a) considers the a priori specified class of regimes $\mathcal{G}_\eta$. The AIPWE for $E\{Y^*(g_\eta)\}$ with a fixed η can be rewritten as

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.11

$$
\begin{aligned}
AIPWE(\eta) &= n^{-1}\sum_{i=1}^{n}\left\{\frac{C_{\eta,i}Y_i}{\pi_c(X_i;\eta,\hat{\gamma})} - \frac{C_{\eta,i}-\pi_c(X_i;\eta,\hat{\gamma})}{\pi_c(X_i;\eta,\hat{\gamma})}m(X_i;\eta,\hat{\beta})\right\}\\
&= n^{-1}\sum_{i=1}^{n}\left\{\frac{C_{\eta,i}}{\pi_c(X_i;\eta,\hat{\gamma})}Y_i - \frac{C_{\eta,i}-\pi_c(X_i;\eta,\hat{\gamma})}{\pi_c(X_i;\eta,\hat{\gamma})}\left[\mu(1,X_i,\hat{\beta})g(X,\eta)+\mu(0,X_i,\hat{\beta})\{1-g(X_i,\eta)\}\right]\right\}\\
&= n^{-1}\sum_{i=1}^{n}\left\{g(X_i,\eta)\hat{C}_{AIPWE}(X_i)\right\} + n^{-1}\sum_{i=1}^{n}\left\{\frac{1-A_i}{1-\pi(X_i,\hat{\gamma})}Y_i - \frac{A_i-\pi(X_i,\hat{\gamma})}{1-\pi(X_i,\hat{\gamma})}\mu(0,X_i,\hat{\beta})\right\},
\end{aligned}
$$

where

$$
\hat{C}_{AIPWE}(X_i) = \frac{A_i}{\pi(X_i,\hat{\gamma})}Y_i - \frac{1-A_i}{1-\pi(X_i,\hat{\gamma})}Y_i - \frac{A_i-\pi(X_i,\hat{\gamma})}{\pi(X_i,\hat{\gamma})}\mu(1,X_i,\hat{\beta}) - \frac{A_i-\pi(X_i,\hat{\gamma})}{1-\pi(X_i,\hat{\gamma})}\mu(0,X_i,\hat{\beta}). \tag{2}
$$

The method of Zhang et al. (2012a) estimates g_η^{opt} by $\hat{g}_{\eta,AIPWE}^{opt} = \arg\max_{g\in\mathcal{G}_\eta} AIPWE(\eta)$, which is equal to $\arg\max_{g\in\mathcal{G}_\eta} n^{-1}\sum_{i=1}^{n} g(X_i)\hat{C}_{AIPWE}(X_i)$. Note that the AIPWE estimator of the contrast function borrows information from, but is not completely determined by, the specified parametric regression model for the outcome. In contrast to this, estimators of the contrast function using regression or G-estimation methods are completely determined by the specified regression models, and the IPWE estimator given below makes no use of a regression model for the outcome given covariates. The AIPWE contrast estimates strike a balance between the foregoing two extremes and this balance partly explains the improved empirical performance in Section 4.

The inverse probability weighted estimator (IPWE) for $E\{Y^*(g_\eta)\}$, is equivalent to estimating the contrast function at each of the observed data points by

$$
\hat{C}_{IPWE}(X_i) = \frac{A_i}{\pi(X_i,\hat{\gamma})}Y_i - \frac{1-A_i}{1-\pi(X_i,\hat{\gamma})}Y_i. \tag{3}
$$

Unlike other methods for estimating the contrast function, which either are completely determined by or incorporates information from a semiparametric or parametric outcome regression model, the IPWE estimator of $C(X_i)$ for each i is completely determined by the observed outcome, weighted by an estimated propensity. The IPWE estimates of the contrast values may be too noisy to successfully inform the class of treatment regimes, as demonstrated by our simulations.

From the above discussion, we see that estimating the optimal treatment regime in the class $\mathcal{G}$ or the restricted class $\mathcal{G}_\eta$ can be separated into two steps: constructing an estimator $\hat{C}(X_i)$ of the contrast function $C(X_i)$ for $i = 1,\ldots,n$, and subsequently estimating g^{opt} by $\hat{g}^{opt} = \arg\max n^{-1}\sum_{i=1}^{n} g(X_i)\hat{C}(X_i)$, where the maximization is across all regimes in the class considered. Note that in the regression and G-estimation methods, the class of regimes is dictated by either $\mu(A,X;\beta)$ or $C_G(X;\psi)$; and the estimation of the contrast function and the maximization of the objective function are carried out simultaneously by fitting the corresponding regression models. As a result, if the posited regression model, $\mu(A,X;\beta)$ or $C_G(X;\psi)$, is correctly specified, the corresponding estimator of the contrast function is consistent; in such a case the estimator of the optimal treatment regime is a consistent estimator of g^{opt}, as $\mathcal{G}_\beta$ or $\mathcal{G}_\psi$ contains g^{opt}. If the posited model is misspecified, however, $\hat{g}_{reg}^{opt}$ or $\hat{g}_G^{opt}$ may be far from the optimal treatment regime in $\mathcal{G}$ or even the optimal regime in the corresponding restricted class $\mathcal{G}_\beta$ or $\mathcal{G}_\psi$, and thus, may perform poorly. In contrast, in the robust AIPWE-based method of Zhang et al. (2012), the estimation of the contrast function and the maximization of the objective function across $\mathcal{G}_\eta$ are separated. An advantage of this separation is that even if $\mathcal{G}_\eta$ does not contain g^{opt}, the resulting estimator may still be the optimal one in $\mathcal{G}_\eta$. In the method of Zhang et al. (2012a), the class of treatment regimes under consideration, $\mathcal{G}_\eta$, indexed by a finite-dimensional parameter, is either entirely determined by the model for the outcomes Y or is pre-specified based on practical considerations. In practice, one can inform the class of treatment regimes by using standard model building techniques for the regression model of outcome on

(wileyonlinelibrary.com) DOI: 10.1002/sta4.11

The ISI's Journal for the Rapid Dissemination of Statistics Research

treatment and patient characteristics. However, these model building techniques target identifying a good model for the outcome, but not necessarily a high-quality treatment regime.

We now introduce a general framework that can address the issues discussed above. Specifically, the problem of estimating the optimal treatment regime is reformulated as a weighted classification problem, where the optimal treatment regime minimizes a weighted misclassification error.

Because $C(X) = I\{C(X) > 0\}|C(X)| - I\{C(X) \leqslant 0\}|C(X)|$, we can rewrite $g(X)C(X)$ as

$$\begin{aligned} g(X)C(X) &= g(X)I\{C(X) > 0\}|C(X)| - g(X)I\{C(X) \leqslant 0\}|C(X)| \\ &= I\{C(X) > 0\}|C(X)| - |C(X)|[\,\{1 - g(X)\}I\{C(X) > 0\} + g(X)I\{C(X) \leqslant 0\}]\,. \end{aligned}$$

As $g(X)$ takes values $\{0, 1\}$, it is easy to check that

$$\{1 - g(X)\}I\{C(X) > 0\} + g(X)I\{C(X) \leqslant 0\} = [\,I\{C(X) > 0\} - g(X)]^2\,.$$

Combining these results, $g(X)C(X)$ can be rewritten as

$$g(X)C(X) = I\{C(X) > 0\}|C(X)| - |C(X)|[\,I\{C(X) > 0\} - g(X)]^2\,.$$

Therefore, we can define the optimal treatment regime as

$$\begin{aligned} g^{opt} &= \arg\max_{g\in G}[\,E\{g(X)C(X)\}] \\ &= \arg\max_{g\in G}\left[E[\,I\{C(X) > 0\}|C(X)|] - E\left\{|C(X)|[\,I\{C(X) > 0\} - g(X)]^2\right\}\right] \\ &= \arg\min_{g\in G}\left[E\left\{|C(X)|[\,I\{C(X) > 0\} - g(X)]^2\right\}\right]. \end{aligned}$$

That is, the optimal treatment regime, g^{opt}, is the one that minimizes $E(|C(X)|[\,I\{C(X) > 0\} - g(X)]^2)$. This identity is what allows us to recast the problem of estimating an optimal treatment regime as a weighted classification problem.

We view each subject as belonging to one of the two classes defined by $Z = I\{C(X) > 0\}$. That is, the class $Z = 1$ is composed of those subjects who would benefit more from treatment 1 compared to treatment 0; i.e., those who have $\mu(1, X) > \mu(0, X)$, and should therefore be treated with treatment option 1. Each subject is also given a weight $W = |C(X)|$, which represents the loss that would be incurred if the subject were misclassified. In this way, we separate the information contained in $C(X)$ into two parts: the class label Z, containing the information about the sign of $C(X)$; and the weight W, containing the information about the magnitude of $C(X)$. Hence, $E(|C(X)|[\,I\{C(X) > 0\} - g(X)]^2)$ can be regarded as the expected weighted misclassification error under the classification rule $g(X)$.

In practice, the contrast function $C(X)$ and hence the class label Z and weight W for each subject are not available in the observed data. As discussed previously, the contrast values $C(X_i)$ for each i can be estimated from the data, for example, using $\hat{C}_{reg}$, $\hat{C}_G$, $\hat{C}_{IPWE}$, or $\hat{C}_{AIPWE}$. Once the estimates $\hat{C}(X_i)$, $i = 1, \ldots, n$, are obtained, we can construct a class label $\hat{Z}_i = I\{\hat{C}(X_i) > 0\}$, and a weight $\hat{W}_i = |\hat{C}(X_i)|$ for each subject, and g^{opt} can be estimated subsequently by $\arg\min_{g\in\mathcal{G}} \sum_{i=1}^n [\,\hat{W}_i\{\hat{Z}_i - g(X_i)\}^2]$. The minimization of $\sum_{i=1}^n [\,\hat{W}_i\{\hat{Z}_i - g(X_i)\}^2]$ can then be viewed as a typical classification problem with $\hat{Z}_i$ as the binary "response," X_i as the "predictor," $\hat{W}_i$ as the "weight," and g is the "classification rule." By reformulating the problem of estimating the optimal treatment regime as a classification problem, existing classification techniques can be used, for example, classification and regression trees (CART, Breiman et al., 1984)

(wileyonlinelibrary.com) DOI: 10.1002/sta4.11

or support vector machines (SVM, Cortes and Vapnik, 1995), to minimize the classification error across a broad class of regimes. Therefore, the parametric form of treatment regimes does not need to be pre-specified and instead can be selected using classification techniques.

We comment that the method of Zhao et al. (2012) can be viewed as a special case within our framework, with the contrast function at each of the observed data points estimated by the IPWE estimator $\hat{C}_{IPWE}(X_i)$. To see this, it is straightforward to show that, corresponding to the IPWE estimator of the contrast function, the class label $\hat{Z}_i = I\{\hat{C}_{IPWE}(X_i) > 0\}$ is equal to A_i as Y is assumed to be positive in Zhao et al. (2012), and that the weight is equal to

$$\hat{W}_i = \left| \frac{A_i}{\pi(X_i, \hat{\gamma})} Y_i - \frac{1 - A_i}{1 - \pi(X_i, \hat{\gamma})} Y_i \right| = \frac{Y_i}{A_i \pi(X_i, \hat{\gamma}) + (1 - A_i)\{1 - \pi(X_i, \hat{\gamma})\}}.$$

Thus, within our framework, the weighted misclassification error rate under treatment rule g is

$$n^{-1} \sum_{i=1}^{n} \hat{W}_i \{\hat{Z}_i - g(X_i)\}^2 = n^{-1} \sum_{i=1}^{n} \frac{Y_i}{A_i \pi(X_i, \hat{\gamma}) + (1 - A_i)\{1 - \pi(X_i, \hat{\gamma})\}} I\{A_i \neq g(X_i)\},$$

which is exactly the approximated weighted classification error used by Zhao et al. (2012). Zhao et al. (2012) minimize the above weighted classification error using support vector machines. The method of Zhao et al. (2012) is predicated on an IPWE estimator of the expected outcome. However, the classification framework we propose is more general and allows estimation of the contrast function by any method, e.g., the AIPWE, as well as the data-driven selection of the class of treatment regimes using the estimated class labels and observation weights.

In the proposed classification framework, we disentangle two critical steps: (i) constructing a suitable estimator of the contrast function, and (ii) finding estimated optimal treatment rules with an interpretable form using classification techniques. This allows both greater flexibility in modeling the outcome or contrast functions and the ability to use any classification technique to inform the class of treatment regimes. We comment that, in our framework, for each subject there is a corresponding "weight" and "label," which do not depend on a treatment regime. Therefore, exploratory analysis and model diagnostics can be used in the classification step by a skilled data-analyst to build high-quality scientifically defensible models. This added benefit, however, is not available in the classification method of Zhao et al. (2012) or the previous work on robust estimation by Zhang et al. (2012a). Also, notice that with the classification approach the interpretability of the final decision rule does not require a parsimonious estimator of the contrast function. Consequently, we can use flexible models for the contrast function, e.g. support vector regression (Vapnik et al., 1997), boosting (Freund & Schapire, 1997), etc., and still produce an interpretable decision rule. In addition, the selection of the form of treatment regimes by classification techniques in the proposed framework is directly targeting the problem of finding the optimal treatment regime by minimizing weighted misclassification error. In the next section, for illustration of the proposed methods, we use classification and regression trees (CART) to produce interpretable decision rules.

4 Simulation studies

To evaluate the performance of the proposed methods, we have carried out two simulation studies, each involving 1000 Monte Carlo data sets. For definiteness, we use CART to minimize the expected weighted misclassification; other methods developed in the area of classification could also be used.

In the first scenario, for each data set, we generated $n = 200$, 500, and 1000 observations (Y_i, A_i, X_i), $i = 1, \ldots, n$, where $X_i = (X_{i1}, \ldots, X_{i5})^T$ and $X_{i1}, \ldots, X_{i5}$ were independent standard normal; given X_i, A_i was Bernoulli with success probability satisfying $\text{logit}\{\text{pr}(A = 1|X)\} = -0.1 + 0.5X_1 + 0.5X_2$, $\text{logit}(u) = \log\{u/(1-u)\}$; and outcomes were

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.11

The ISI's Journal for the Rapid Dissemination of Statistics Research

generated as $Y_i = \mu(A_i, X_i) + \epsilon_i$ for ϵ_i standard normal and $\mu(A, X) = \exp\{2.0 + 0.25X_1 + 0.25X_2 - 0.25X_5 - 0.5(a - g^{opt}(X))^2\}$, where $g^{opt}(X) = I(X_1 > -0.545)I(X_2 < 0.545)$.

For the proposed methods, to estimate the contrast function $C(X)$, we considered the regression estimator $\hat{C}_{reg}$, the AIPWE estimator $\hat{C}_{AIPWE}$, and the IPWE estimator $\hat{C}_{IPWE}$. We considered a working regression model $\mu(A, X; \beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + A(\beta_6 + \beta_7 X_1 + \beta_8 X_2 + \beta_9 X_3 + \beta_{10} X_4 + \beta_{11} X_5)$, which is misspecified, and estimated $\beta = (\beta_1, \ldots, \beta_{11})$ in the model using least squares. We considered both a correctly specified propensity model $\pi(X; \gamma) = \text{expit}(\gamma_0 + \gamma_1 X_1 + \gamma_2 X_2)$, and an incorrectly specified model $\pi(X; \gamma) = \gamma$; these models were fit using ML. Note that when the propensity model is constant, $\hat{C}_{reg}$ and $\hat{C}_G$ are equivalent, thus we omit $\hat{C}_G$.

Once we obtained the estimated contrast function for each subject, e.g., $\hat{C}_{AIPWE}(X_i, A_i, Y_i; \hat{\gamma}, \hat{\beta})$, we defined the binary responses, e.g., $\hat{Z}_i = I\{\hat{C}_{AIPWE}(X_i, A_i, Y_i; \hat{\gamma}, \hat{\beta}) > 0\}$, and the case weights, e.g., $\hat{W}_i = |\hat{C}_{AIPWE}(X_i, A_i, Y_i; \hat{\gamma}, \hat{\beta})|$ for each subject, so that the classification dataset becomes $\{\hat{Z}_i, X_i, \hat{W}_i\}$. We input this new data set into the CART algorithm to find the estimated optimal treatment regime. We used the `R` function `rpart` with default settings, except that we set the weights as the estimated weight $\hat{W}$.

For the second scenario, the data generation was the same as in the first scenario except that $g^{opt}(X) = I(X_1 > X_2)$. Note that here, in contrast to the first scenario, the class of treatment regimes with simple tree form does not contain g^{opt}. Thus, this scenario examines whether or not CART can find a regime close to the optimal treatment regime g^{opt}.

We also estimated the optimal treatment regime using the usual regression (RG) method which models $\mu(A, X; \beta)$ and the robust AIPWE-based method of Zhang et al. (2012a) which involves modeling both $\mu(A, X; \beta)$ and propensities; models for $\mu(A, X; \beta)$ and propensities are as those used in the proposed methods. In the method of Zhang et al. (2012a), we consider optimizing over the class of treatment regimes defined by the outcome regression model, i.e., $\mathcal{G}_\eta = \{I(\beta_6 + \beta_7 X_1 + \beta_8 X_2 + \beta_9 X_3 + \beta_{10} X_4 + \beta_{11} X_5 > 0)\}$.

Table I. Results for the first simulation scenario using 1000 Monte Carlo data sets. $E\{Y^*(g^{opt})\} = 8.12$. $\hat{E}(\hat{g}^{opt})$ shows the Monte Carlo average and standard deviation of estimated values of the true $E(\hat{g}^{opt})$ using (1). $E(\hat{g}^{opt})$ shows the Monte Carlo average and standard deviation of values $E\{Y^*(\hat{g}^{opt})\}$ obtained using 10^6 Monte Carlo simulations for each data set. PS correct and PS incorrect indicate whether the specified propensity score model is correct or not.

	n=200		n=500		n=1000	
Estimator	$\hat{E}(\hat{g}^{opt})$	$E(\hat{g}^{opt})$	$\hat{E}(\hat{g}^{opt})$	$E(\hat{g}^{opt})$	$\hat{E}(\hat{g}^{opt})$	$E(\hat{g}^{opt})$
	PS correct					
RG	7.56 (0.35)	7.49 (0.08)	7.56 (0.21)	7.53 (0.05)	7.55 (0.16)	7.54 (0.04)
Zhang et al.	7.83 (0.37)	7.53 (0.08)	7.76 (0.22)	7.59 (0.06)	7.73 (0.17)	7.62 (0.04)
$\hat{C}_{AIPWE}$	8.08 (0.37)	8.07 (0.06)	8.10 (0.22)	8.11 (0.02)	8.11 (0.16)	8.12 (0.01)
$\hat{C}_{IPWE}$	7.18 (0.53)	7.02 (0.42)	7.66 (0.40)	7.57 (0.38)	7.93 (0.26)	7.90 (0.22)
$\hat{C}_{reg}$	7.51 (0.36)	7.43 (0.18)	7.52 (0.23)	7.49 (0.10)	7.52 (0.17)	7.50 (0.08)
	PS incorrect					
RG	7.50 (0.30)	7.49 (0.08)	7.50 (0.18)	7.53 (0.05)	7.50 (0.13)	7.54 (0.04)
Zhang et al.	7.68 (0.30)	7.52 (0.09)	7.62 (0.19)	7.57 (0.06)	7.59 (0.14)	7.59 (0.05)
$\hat{C}_{AIPWE}$	7.95 (0.31)	8.08 (0.04)	7.98 (0.19)	8.11 (0.02)	7.99 (0.14)	8.12 (0.01)
$\hat{C}_{IPWE}$	6.98 (0.34)	6.93 (0.25)	7.08 (0.22)	7.04 (0.16)	7.12 (0.16)	7.08 (0.12)
$\hat{C}_{reg}$	7.44 (0.31)	7.43 (0.18)	7.45 (0.20)	7.49 (0.10)	7.44 (0.14)	7.50 (0.08)

(wileyonlinelibrary.com) DOI: 10.1002/sta4.11

Table II. Results for the second simulation scenario using 1000 Monte Carlo data sets. $E\{Y^*(g^{opt})\} = 8.12$. Entries as Table I.

	n=200		n=500		n=1000	
Estimator	$\hat{E}(\hat{g}^{opt})$	$E(\hat{g}^{opt})$	$\hat{E}(\hat{g}^{opt})$	$E(\hat{g}^{opt})$	$\hat{E}(\hat{g}^{opt})$	$E(\hat{g}^{opt})$
	PS correct					
RG	7.82 (0.32)	7.78 (0.13)	7.84 (0.21)	7.83 (0.10)	7.85 (0.16)	7.84 (0.08)
Zhang et al.	8.09 (0.32)	8.01 (0.08)	8.11 (0.20)	8.07 (0.03)	8.11 (0.15)	8.09 (0.02)
$\hat{C}_{AIPWE}$	7.96 (0.33)	7.68 (0.13)	8.03 (0.20)	7.83 (0.06)	8.04 (0.15)	7.89 (0.04)
$\hat{C}_{IPWE}$	7.11 (0.41)	6.96 (0.31)	7.39 (0.28)	7.28 (0.23)	7.55 (0.22)	7.47 (0.17)
$\hat{C}_{reg}$	7.67 (0.34)	7.57 (0.17)	7.73 (0.21)	7.68 (0.11)	7.73 (0.16)	7.70 (0.08)
	PS incorrect					
RG	7.79 (0.30)	7.78 (0.13)	7.82 (0.19)	7.83 (0.10)	7.82 (0.14)	7.84 (0.08)
Zhang et al.	8.01 (0.30)	8.02 (0.06)	8.03 (0.18)	8.07 (0.03)	8.03 (0.14)	8.09 (0.02)
$\hat{C}_{AIPWE}$	7.87 (0.30)	7.70 (0.11)	7.95 (0.18)	7.85 (0.05)	7.96 (0.13)	7.89 (0.04)
$\hat{C}_{IPWE}$	7.11 (0.37)	7.09 (0.26)	7.22 (0.24)	7.19 (0.16)	7.25 (0.18)	7.22 (0.12)
$\hat{C}_{reg}$	7.63 (0.31)	7.57 (0.17)	7.70 (0.20)	7.68 (0.11)	7.70 (0.15)	7.70 (0.08)

Results for the two scenarios are shown in Tables I and II, respectively. Under scenario 1 (Table I) where the true g^{opt} is in the form of a tree, it is clear that the proposed method using the AIPWE estimator of the contrast function, i.e., $\hat{C}_{AIPWE}$, achieves the best performance overall, with expected outcomes under the chosen regimes very close to the expectation under the true optimal regime. This good performance, we believe, is for two reasons. First, $\hat{C}_{AIPWE}$ estimates the contrast function using the AIPWE, which, as discussed previously, is robust and efficient relative to competing methods. Second, it exploits a flexible classification method for optimization, without having to prespecify the parametric form of the class of regimes under consideration. All other methods lack either one or both of these two features. We note that $\hat{C}_{IPWE}$ has the worst performance of the considered methods across all scenarios, which may be due to instability of the IPWE. Under scenario 2 (shown in Table II) the true g^{opt} is linear and not well-approximated by a tree with splits along the coordinate axes; the method of Zhang et al. (2012a) has the best performance, which is expected since the specified class $\mathcal{G}_\eta$ contains the true optimal regime. Nevertheless, the proposed method using $\hat{C}_{AIPWE}$ estimates regimes with near optimal performance.

5 Application to the NSABP trial

As an illustration, consider data from a trial conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP) comparing L-phenylalanine mustard and 5-fluorouracil (PF) to PF plus tamoxifen (PFT) in patients with primary operable breast cancer and positive nodes (Fisher et al., 1983). The study investigators found that heterogeneity in response to PFT exists and the response depends on age (years) and progesterone receptor level (PR, fmol). Gail & Simon (1985) analyzed these data using a test for qualitative interaction between treatment and covariates. Their results support the regime proposed by Fisher et al. (1983), which recommends that subjects with age < 50 and PR < 10 fmol should receive PF, with all others receiving PFT.

We analyzed data from $n = 1276$ patients with complete information on age and PR. Because the distribution of PR is very skewed, following Zhang et al. (2012a), we make the log transformation, i.e., $\text{LPR} = \log(\text{PR} + 1)$. We denote age and LPR by X_1 and X_2, respectively. The outcome of interest is binary with $Y = 1$ if a subject survived disease-free

(wileyonlinelibrary.com) DOI: 10.1002/sta4.11

The ISI's Journal for the Rapid Dissemination of Statistics Research

to three years from baseline, and $Y = 0$ otherwise. Indicator variable A denotes treatment with $A = 1$ if a subject was randomized to PFT and 0 if PF.

We implemented the proposed method using the AIPWE estimator of the contrast function and CART for the classification step. The simple form of a decision tree yielded from CART allow us to make direct comparison with the regime of Fisher et al. (1983) and Gail & Simon (1985). To calculate $\hat{C}_{AIPWE}(X, A, Y; \hat{\gamma}, \hat{\beta})$, one needs to build both outcome regression and propensity score models. For the outcome regression, we postulated the logistic regression model

$$\mu(A, X; \beta) = \text{expit}\{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)\} \tag{4}$$

for $E(Y|A, X) = \text{pr}(Y = 1|A, X)$, where $\text{expit}(u) = e^u/(1 + e^u)$. The propensity score $\pi(X)$ was estimated directly by the sample proportion i.e., $\sum_{i=1}^{n} A_i/n$ for all X, as this was a randomized study. Constructing weights and labels based on the estimated contrasts, the estimated optimal treatment regime given by CART is $\hat{g}^{opt}_{C,AIPWE}(X) = 1 - I(\text{age} < 59.5 \text{ and PR} < 16.5)$, under which a patient should receive PF if she is younger than 59.5 and has PR less than 16.5 and should receive PFT otherwise. Note that the estimated regime has the same form as that of Fisher et al. (1983) and Gail & Simon (1985) but differs a bit in the cutoff values. The estimated mean outcomes using (1) under the estimated regime is 0.681 $(95\%CI : 0.646, 0.717)$.

Considering a restricted class of regimes with a form $1 - I(\text{age} < \eta_1 \text{ and PR} < \eta_2)$, the robust AIPWE-based method of Zhang et al. (2012a) yields an estimated regime given by $1 - I(\text{age} < 60 \text{ and PR} < 9)$, with estimated mean outcomes under the regime 0.686 $(0.651, 0.722)$. The results are virtually identical as in our methods. However, in the method of Zhang et al. (2012a), the form of regime has to be determined a priori, which can be challenging in practice.

6 Discussion

We proposed a novel framework within which the optimal treatment regime at a single decision point can be estimated using off-the-shelf classification methods. This framework allows the separation of two critical steps. In the first step, estimated contrast functions are constructed for each subject independently without the need to specify a class of treatment regimes. Based on the estimated contrasts, a "weight" and a binary "response" are created for each subject which are then used as input to a classification algorithm to identify the optimal treatment regime by minimizing a weighted misclassfication error. This separation creates flexibility and allows the use of existing classification algorithms on this new class of problems. As in Zhang et al. (2012a), the class of treatment regimes does not have to be dictated by a regression model for the outcome and can therefore be more robust and flexible.

The proposed framework is general enough to include both the work of Zhao et al. (2012), and Zhang et al. (2012a) as special cases. Nonetheless, there are a number of interesting directions for future research; we mention two that are of particular interest. The first is the incorporation of variable selection methods both in the modeling of the contrast function (say, through the outcome regression model) and the subsequent classification algorithm. One approach would be to perform model selection separately for the estimation of the contrast function and the estimation of the optimal treatment regime. However, in high dimensions, the selected outcome regression model may have a significant impact on the quality of the estimated optimal treatment regime, and it is preferable that the two model selection steps be done in concert. A second direction is to extend this framework to include the multiple decision setting. In this setting, personalized treatment is operationalized as sequence of treatment regimes, one for each stage of clinical intervention, that adapt to the patients evolving health status. Zhang et al. (2012b) derive an AIPWE method for estimating an optimal sequence of treatment regimes but a general framework is lacking.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.11

Acknowledgement

This research was partially supported by NIH grants P01CA142538, R01CA085848, and R37AI031789.

References

Breiman, L, Freidman, JH, Olshen, RA & Stone, CJ (1984), *Classification and Regression Trees*, *Wadsworth*, Belmont, CA.

Brinkley, J, Tsiatis, AA & Anstrom, KJ (2009), 'A generalized estimator of the attributable benefit of an optimal treatment regime', *Biometrics*, **21**, 512–522.

Cortes, C & Vapnik, V (1995), 'Support-vector networks', *Machine Learning*, **20**, 273–297.

Fisher, B, Redmond, C, Brown, A, Wickerham, DL, Wolmark, N, Allegra, J, Escher, G, Lippman, M, Savlov, E & Wittliff, J (1983), 'Influence of tumor estrogen and progesterone receptor levels on the response to Tamoxifen and chemotherapy in primary breast cancer', *Journal of Clinical Oncology*, **1**, 227–241.

Freund, Y & Schapire, RE (1997), 'A decision-theoretic generalization of on-line learning and an application to boosting', *Journal of Computer and System Sciences*, **55**, 119–139.

Gail, M & Simon, R (1985), 'Testing for qualitative interactions between treatment effects and patient subsets', *Biometrics*, **41**, 361–372.

Gunter, L, Zhu, J & Murphy, SA (2011), 'Variable selection for qualitative interactions', *Statistical Methodology*, **8** (1), 42–55.

Hastie, T, Tibshirani, R & Friedman, J (2009), *The Elements of Statistical Learning*, *Springer-Verlag*, New York, NY.

Henderson, R, Ansell, P & Alshibani, D (2010), 'Regret-regression for optimal dynamic treatment regimes', *Biometrics*, **66**, 1192–1201.

Moodie, EEM, Richardson, TS & Stephens, DA (2007), 'Demystifying optimal dynamic treatment regimes', *Biometrics*, **63**, 447–455.

Murphy, SA (2003), 'Optimal dynamic treatment regimes (with discussion)', *Journal of the Royal Statistical Society, Series B*, **58**, 331–366.

Orellana, L, Rotnitzky, A & Robins, J (2010), 'Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: main content', *International Journal of Biostatistics*, **6**(2), Article 8, DOI 10.2202/1557-4679.1200.

Robins, JM. (2004). 'Optimal structured nested models for optimal sequential decisions', in Lin, DY & Heagerty, PJ (eds.), Proceedings of the Second Seattle Symposium on Biostatistics, Springer, New York, pp. 189–326.

Robins, J, Orellana, L & Rotnitzky, A (2008), 'Estimation and extrapolation of optimal treatment and testing strategies', *Statistics in Medicine*, **27**, 4678–4721.

Rubin, DB (1978), 'Bayesian inference for causal effects: the role of randomization', *Annals of Statistics*, **6**, 34–58.

Schulte, PJ, Tsiatis, AA, Laber, EB & Davidian, M (2012), 'Q- and A-learning methods for estimating optimal dynamic treatment regimes', *Statistical Science*. in revision, http://arxiv.org/abs/1202.4177.

Vapnik, V, Golowich, S & Smola, A (1997), 'Support vector method for function approximation, regression estimation, and signal processing', *Advances in Neural Information Processing Systems*, **9**, 281–287.

Zhang, B, Tsiatis, AA, Laber, EB & Davidian, M (2012a), 'A robust method for estimating optimal treatment regimes', *Biometrics*, in press. DOI: 10.1111/j.1541-0420.2012.01763.x.

Zhang, B, Tsiatis, AA, Laber, EB & Davidian, M (2012b), 'Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions', *Biometrika*, in revision.

Zhao, Y, Kosorok, MR & Zeng, D (2009), 'Reinforcement learning design for cancer clinical trials', *Statistics in Medicine*, **28**, 3294–3315.

Zhao, Y, Zeng, D, Rush, AJ & Kosorok, MR (2012), 'Estimating individualized treatment rules using outcome weighted learning', *Journal of the American Statistical Association*, **107**, 1106–1118.