Using Decision Lists to Construct Interpretable and Parsimonious Treatment Regimes

Yichi Zhang[*], **Eric B. Laber**[**], **Anastasios Tsiatis**[***], and **Marie Davidian**[****]

Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA

Summary

A treatment regime formalizes personalized medicine as a function from individual patient characteristics to a recommended treatment. A high-quality treatment regime can improve patient outcomes while reducing cost, resource consumption, and treatment burden. Thus, there is tremendous interest in estimating treatment regimes from observational and randomized studies. However, the development of treatment regimes for application in clinical practice requires the long-term, joint effort of statisticians and clinical scientists. In this collaborative process, the statistician must integrate clinical science into the statistical models underlying a treatment regime and the clinician must scrutinize the estimated treatment regime for scientific validity. To facilitate meaningful information exchange, it is important that estimated treatment regimes be interpretable in a subject-matter context. We propose a simple, yet flexible class of treatment regimes whose members are representable as a short list of if-then statements. Regimes in this class are immediately interpretable and are therefore an appealing choice for broad application in practice. We derive a robust estimator of the optimal regime within this class and demonstrate its finite sample performance using simulation experiments. The proposed method is illustrated with data from two clinical trials.

Keywords

Decision lists; Exploratory analyses; Interpretability; Personalized medicine; Treatment regimes

1. Introduction

Treatment regimes formalize clinical decision making as a function from patient information to a recommended treatment. Proponents of personalized medicine envisage the widespread clinical use of evidence-based, i.e., data-driven, treatment regimes. The potential benefits of applying treatment regimes are now widely recognized. By individualizing treatment, a treatment regime may improve patient outcomes while reducing cost and the consumption of resources. This is important in an era of growing medical costs and an aging global

[*]yzhang52@ncsu.edu

[**]laber@stat.ncsu.edu

[***]tsiatis@ncsu.edu

[****]davidian@ncsu.edu

population. However, the widespread integration of data-driven treatment regimes into clinical practice is, and should be, an incremental process wherein: (i) data are used to generate hypotheses about optimal treatment regimes; (ii) the generated hypotheses are scrutinized by clinical collaborators for scientific validity; (iii) new data are collected for validation and new hypothesis generation, and so on. Within this process, it is crucial that estimated treatment regimes be interpretable to clinicians. Nevertheless, optimality, not interpretability, has been the focal point in the statistical literature on treatment regimes.

A treatment regime said to be optimal if it maximizes the expectation of a pre-specified clinical outcome when used to assign treatment to a population of interest. There is a large literature on estimating optimal treatment regimes using data from observational or randomized studies. Broadly, these estimators can be categorized as regression-based or classification-based estimators. Regression-based estimators model features of the conditional distribution of the outcome given treatment and patient covariates. Examples include estimators of the regression of an outcome on covariates, treatment, and their interactions (e.g., Su et al., 2009; Qian and Murphy, 2011; Tian et al., 2014), and estimators of point treatment effects given covariates (e.g., Robins, 1994; Vansteelandt et al., 2014). Regression-based methods rely on correct specification of some or all of the modeled portions of the conditional distribution of the outcome. Thus, a goal of many regression-based estimators is to ensure correct model specification under a large class of generative models (Zhao et al., 2009; Qian and Murphy, 2011; Moodie et al., 2013; Laber et al., 2014; Taylor et al., 2014). However, as flexibility is introduced into the model, interpretability tends to diminish, and in the extreme case the estimated treatment regime becomes an unintelligible black box.

Classification-based estimators, also known as policy-search or value-search estimators, estimate the marginal mean of the outcome for every treatment regime within a pre-specified class and then take the maximizer as the estimated optimal regime. Examples include marginal structural mean models (Robins et al., 2008; Orellana et al., 2010); robust marginal mean models (Zhang et al., 2012); and outcome weighted learning (Zhang et al., 2012; Zhao et al., 2012, 2015). Classification-based estimators often rely on fewer assumptions about the conditional distribution of the outcome given treatment and patient information and thus may be more robust to model misspecification than regression-based estimators (Zhang et al., 2012,?). Furthermore, because classification-based methods estimate an optimal regime within a pre-specified class, it is straightforward to impose structure on the estimated regime, e.g., interpretability, by restricting this class. We use robust marginal mean models with a highly interpretable yet flexible class of regimes to estimate a high-quality regime that can be immediately understood by clinical and intervention scientists.

To obtain an interpretable and parsimonious treatment regime, we use the concept of decision list, which was developed in the computer science literature for representing flexible but interpretable classifiers (Rivest, 1987; Clark and Niblett, 1989; Marchand and Sokolova, 2005; Letham et al., 2012; Wang and Rudin, 2014); see Freitas (2014) for a recent position paper on the importance of interpretability in predictive modeling and additional references on interpretable classifiers. As a treatment regime, a decision list comprises a sequence of "if-then" clauses that map patient covariates to a recommended treatment.

Figure 1 shows a decision list for patients with chronic depression (see Section 4.2). This decision list recommends treatments as follows: if a patient presents with somatic anxiety score above 1 and retardation score above 2, the list recommends nefazodone; otherwise, if the patient has Hamilton anxiety score above 23 and sleep disturbance score above 2, the list recommends psychotherapy; and otherwise the list recommends nefazodone + psychotherapy (combination). Thus, a treatment regime represented as a decision list can be conveyed as either a diagram or text and is easily understood, in either form, by domain experts. Indeed, decision lists have frequently been used to display estimated treatment regimes (Shortreed et al., 2011; Moodie et al., 2012; Shortreed et al., 2014; Laber and Zhao, 2015) or to describe theory-based, i.e., not data-driven, treatment regimes (Shiffman, 1997; Marlowe et al., 2012).

Another important attribute of a decision list is that it "short circuits" measurement of patient covariates; e.g., in Figure 1, the Hamilton anxiety score and sleep disturbance score do not need to be collected for patients with somatic anxiety score above 1 and retardation score above 2. This is important in settings where patient covariates are expensive or burdensome to collect (e.g., Gail et al., 1999; Gail, 2009; Baker et al., 2009; Huang et al., 2015). We provide an estimator of the treatment regime that minimizes an expected cost among all regimes that optimize the marginal mean outcome.

2. Methodology

2.1 Framework

We assume that the observed data are $\{(X_i, A_i, Y_i)\}_{i=1}^{n}$, which comprise n independent identically distributed observations, one for each subject in an experimental or observational study. Let (X, A, Y) denote a generic observation. Then $X \in \mathbb{R}^p$ are baseline patient covariates; $A \in A = \{1, \ldots, m\}$ is the treatment assigned; and $Y \in \mathbb{R}$ is the outcome, coded so that higher values are better. A treatment regime, π, is a function from $\mathbb{R}^p$ into A, so that under π a patient presenting with $X = x$ is recommended treatment $\pi(x)$.

The value of a regime π is the expected outcome if all patients in the population of interest are assigned treatment according to π. To define the value, we use the set of potential outcomes $\{Y^*(a)\}_{a \in A}$, where $Y^*(a)$ is the outcome that would be observed if a subject were assigned treatment a. Define $Y^*(\pi) = \sum_{a \in A} Y^*(a) I\{\pi(X) = a\}$ to be the potential outcome under regime π, and $R(\pi) = E\{Y^*(\pi)\}$ to be the value of regime π. An optimal regime, say π^{opt}, satisfies $R(\pi^{\text{opt}}) \geq R(\pi)$ for all π. Let $\prod$ denote a class of regimes of interest. Classification-based estimation methods form an estimator of $R(\pi)$, say $\hat{R}(\pi)$, and then estimate π^{opt} using $\hat{\pi} = \arg\max_{\pi \in \prod} \hat{R}(\pi)$. The success of this approach requires: (i) a high-quality estimator of $R(\pi)$; (ii) a sufficiently rich class $\prod$; and (iii) an efficient algorithm for maximizing $\hat{R}(\pi)$ over $\prod$. We discuss these topics in the next three sections.

2.2 Estimation of R(π)

We make several standard assumptions: (A1) consistency: $Y = Y^*(A)$; (A2) no unmeasured confounders: $\{Y^*(a)\}_{a \in A}$ are conditionally independent of A given X; and (A3) positivity: there exists $\delta > 0$ so that $\mathbb{P}(A = a|X) \geq \delta$ for all $a \in A$. Assumption (A2) is automatically

satisfied in a randomized study but is untestable in observational studies (Robins et al., 2000). Under (A1)–(A3), it can be shown (Tsiatis, 2006) that

$$R(\pi) = E\left(\sum_{a=1}^{m}\left[\frac{I(A=a)}{\omega(X,a)}\{Y - \mu(X,a)\} + \mu(X,a)\right]I\{\pi(X)=a\}\right), \quad (1)$$

where $\omega(x, a) = \mathbb{P}(A = a|X = x)$ and $\mu(x, a) = E(Y|X = x, A = a)$. Alternate expressions for $R(\pi)$ exist (Zhang et al., 2012); however, estimators based on (1) possess a number of desirable properties (see below).

To construct an estimator of $R(\pi)$ from (1) we replace $\omega(x, a)$ and $\mu(x, a)$ with estimated working models and replace the expectation with its sample analog. If treatment is randomly assigned independently of subject covariates, then $\omega(x, a)$ can be estimated by $n^{-1}\sum_{i=1}^{n}I(A_i=a)$. Otherwise, we posit a multinomial logistic regression model of the form $\omega(x,a) = exp(u^{\mathrm{T}}\gamma_a)/\left\{1 + \sum_{j=1}^{m-1}exp(u^{\mathrm{T}}\gamma_j)\right\}$, $a = 1, \ldots, m - 1$, where $u = u(x)$ is a known feature vector, and $\gamma_1, \ldots, \gamma_{m-1}$ are unknown parameters. Let $\hat{\omega}(x, a)$ denote the maximum likelihood estimator of $\omega(x, a)$, where $\gamma_1, \ldots, \gamma_{m-1}$ are replaced by maximum likelihood estimators $\hat{\gamma_1}, \ldots, \hat{\gamma_{m-1}}$. We posit a generalized linear model for $\mu(x, a)$, $g\{\mu(x, a)\} = z^{\mathrm{T}}\beta_a$, where $g(\cdot)$ is a known link function, $z = z(x)$ is a known feature vector constructed from x, and $\beta_1, \ldots, \beta_m$ are unknown parameters. We use $\hat{\mu}(x, a) = g^{-1}(z^{\mathrm{T}}\hat{\beta_a})$ as our estimator of $\mu(x, a)$, where $\hat{\beta_1}, \ldots, \hat{\beta_m}$ are the maximum likelihood estimators of $\beta_1, \ldots, \beta_m$.

Given estimators $\hat{\omega}(x, a)$ and $\hat{\mu}(x, a)$, an estimator of $R(\pi)$ based on (1) is

$$\hat{R}(\pi) = \frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{m}\left[\frac{I(A_i=a)}{\hat{\omega}(X_i,a)}\{Y_i - \hat{\mu}(X_i,a)\} + \hat{\mu}(X_i,a)\right]I\{\pi(X_i)=a\}. \quad (2)$$

For any fixed π, $\hat{R}(\pi)$ is doubly robust in the sense that it is a consistent estimator of $R(\pi)$ if either the model for $\omega(x, a)$ or $\mu(x, a)$ is correctly specified (Tsiatis, 2006; Zhang et al., 2012). As a direct consequence, $\hat{R}(\pi)$ is guaranteed to be consistent in a randomized study, as $\omega(x, a)$ is known by design. Furthermore, if both models are correctly specified, then $\hat{R}(\pi)$ is semiparametric efficient; i.e., it has the smallest asymptotic variance among the class of regular, asymptotically linear estimators (Tsiatis, 2006).

2.3 Regimes representable as decision lists

Gail and Simon (1985) present an early example of a treatment regime using data from the NSABP clinical trial. The treatment regime they propose is

> **If** $age \leq 50$ and $PR \leq 10$ **then** $chemotherapy$ $alone$;
> **else** $chemotherapy$ $with$ $tamoxifen$,

where age (in years) denotes the age of the patient and PR denotes the progesterone receptor level (in fmol). The simple if-then structure of the foregoing treatment regime makes it immediately interpretable.

Formally, a treatment regime, π, that is representable as a decision list of length L is described by $\{(c_1, a_1), \ldots, (c_L, a_L), a_0\}$, where c_j is a logical condition that is true or false for each $x \in \mathbb{R}^p$, and $a_j \in A$ is a recommended treatment, $j = 0, \ldots, L$. As a special case, $L = 0$ is allowed. The corresponding treatment regime $\{a_0\}$ gives the same treatment a_0 to every patient. Hereafter, let $\prod$ denote the set of regimes that are representable as a decision list. Clearly, the regime proposed by Gail and Simon (1985) is a member of $\prod$.

Define $T(c_j) = \{x \in \mathbb{R}^p : c_j \text{ is true for } x\}$, $j = 1, \ldots, L$; $R_1 = T(c_1)$, $R_j = \{\cap_{\ell < j} T(c_\ell)^c\} \cap T(c_j)$, $j = 2, \ldots, L$; and $R_0 = \cap_{\ell=1}^L T(c_\ell)^c$, where S^c is the complement of the set S. Then a regime $\pi \in \prod$ can be written as $\pi(x) = \sum_{\ell=0}^L a_\ell I(x \in R_\ell)$, which has structure

$$
\begin{aligned}
&\textbf{If } c_1 \textbf{ then } a_1; \\
&\textbf{else if } c_2 \textbf{ then } a_2; \\
&\qquad \cdots \\
&\textbf{else if } c_L \textbf{ then } a_L; \\
&\qquad \textbf{else } a_0.
\end{aligned}
\qquad (3)
$$

In principle, the conditions c_j, and hence the sets $T(c_j)$, can be arbitrary. To ensure parsimony and interpretability, we restrict c_j so that $T(c_j)$ is one of the following sets:

$$
\begin{aligned}
&[1]: &&\{x \in \mathbb{R}^p : x_{j_1} \leq \tau_1\}, &&[6]: &&\{x \in \mathbb{R}^p : x_{j_1} \leq \tau_1 \text{ or } x_{j_2} \leq \tau_2\}, \\
&[2]: &&\{x \in \mathbb{R}^p : x_{j_1} \leq \tau_1 \text{ and } x_{j_2} \leq \tau_2\}, &&[7]: &&\{x \in \mathbb{R}^p : x_{j_1} \leq \tau_1 \text{ or } x_{j_2} > \tau_2\}, \\
&[3]: &&\{x \in \mathbb{R}^p : x_{j_1} \leq \tau_1 \text{ and } x_{j_2} > \tau_2\}, &&[8]: &&\{x \in \mathbb{R}^p : x_{j_1} > \tau_1 \text{ or } x_{j_2} \leq \tau_2\}, \\
&[4]: &&\{x \in \mathbb{R}^p : x_{j_1} > \tau_1 \text{ and } x_{j_2} \leq \tau_2\}, &&[9]: &&\{x \in \mathbb{R}^p : x_{j_1} > \tau_1 \text{ or } x_{j_2} > \tau_2\}, \\
&[5]: &&\{x \in \mathbb{R}^p : x_{j_1} > \tau_1 \text{ and } x_{j_2} > \tau_2\}, &&[10]: &&\{x \in \mathbb{R}^p : x_{j_1} > \tau_1\},
\end{aligned}
\qquad (4)
$$

where $j_1 < j_2 \in \{1, \ldots, p\}$ are indices and $\tau_1, \tau_2 \in \mathbb{R}$ are thresholds. We believe that the conditions that dictate the sets in (4), e.g., $x_{j_1} \quad \tau_1$ and $x_{j_2} \quad \tau_2$, are more easily interpreted than those dictated by linear thresholds, e.g., $a_1 x_{j_1} + a_2 x_{j_2} \quad a_3$, as the former are more commonly seen in clinical practice.

In the proposed setup, at most two variables are involved in any single condition. Having a small number of variables in each clause yields two important properties. First, the resulting treatment regime is parsimonious, and the most important variables for treatment selection are automatically identified. Second, application of the treatment regime allows for patient measurements to be taken in sequence so that the treatment recommendation can be short-circuited. For example, consider a decision list described by $\{(c_1, a_1), (c_2, a_2), a_0\}$. It is necessary to measure the variables that compose c_1 on all subjects, but the variables composing c_2 need only be measured for those who do not satisfy c_1.

2.3.1 Uniqueness and minimal cost of a decision list—For a decision list π described by $\{(c_1, a_1), \ldots, (c_L, a_L), a_0\}$, let N_ℓ denote the cost of measuring the covariates required to check logical conditions $c_1, \ldots, c_\ell$. Hereafter, for simplicity, we assume that this cost is equal to the number of covariates needed to check $c_1, \ldots, c_\ell$, but it can be extended easily to a more complex cost function reflecting risk, burden, and availability. The expected

cost of applying treatment rule $\pi(x) = \sum_{\ell=0}^{L} a_\ell I(x \in R_\ell)$ is

$N(\pi) = \sum_{\ell=1}^{L} N_\ell \, \mathbb{P}(X \in R_\ell) + N_L \, \mathbb{P}(X \in R_0)$, which is smaller than

$N_L = N_L \sum_{\ell=0}^{L} \mathbb{P}(X \in R_\ell)$, the cost of measuring all covariates in the treatment regime. This observation reflects the benefit of the short-circuit property.

A decision list π described by $\{(c_1, a_1), \ldots, (c_L, a_L), a_0\}$ need not be unique in that there may exist an alternative decision list π' described by $\left\{(c_1', a_1'), \ldots, (c_{L'}', a_{L'}'), a_0'\right\}$ such that $\pi(x) = \pi'(x)$ for all x but $L \neq L'$ or $L = L'$ but $c_j \neq c_j'$ or $a_j \neq a_j'$ for some $j \in \{1, \ldots, L\}$. This is potentially important because the expected costs $N(\pi)$ and $N(\pi')$ might differ substantially. Figure 2 shows two representations, π and π', of the same decision list both with $L = L' = 2$ but with different clauses. The cost of the decision list in the middle panel, π, is $N(\pi) = N_1 \mathbb{P}(X_1 > \tau_1) + N_2 \mathbb{P}(X_1 \leq \tau_1)$, whereas the cost of the decision list in the right panel, π' is $N(\pi') = N_2 \geq N(\pi)$ with strict inequality if $N_2 > N_1$ and $\mathbb{P}(X_1 > \tau_1) > 0$. Thus, π is preferred to π' in settings where X_2 is a biomarker that is expensive, burdenome, or potentially harmful to collect (e.g., Huang et al., 2015, and references therein).

Therefore, among all decision lists achieving the value $R(\pi^{\mathrm{opt}})$, where π^{opt} is an optimal regime as defined previously, we seek to estimate the one that minimizes the cost. Defining L_r to be the level set $\{\pi \in \Pi : R(\pi) = r\}$, then the goal is to estimate a regime in the set arg $\min_{\pi \in L\{R(\pi^{\mathrm{opt}})\}} N(\pi)$. Define $\hat{L}(r) = \{\pi \in \Pi : \hat{R}(\pi) = r\}$. Let $\tilde{\pi}$ be an estimator of an element in the set arg $\max_{\pi \in \Pi} \hat{R}(\pi)$. In the following we provide an algorithm that ensures our estimator, $\hat{\pi}$, belongs to the set arg $\min_{\pi \in \hat{L}\{\hat{R}(\tilde{\pi})\}} \hat{N}(\pi)$, where $\hat{N}(\pi)$ is defined by replacing the probabilities in $N(\pi)$ with sample proportions.

2.4 Computation

Estimation proceeds in two steps: (i) approximate an element $\tilde{\pi} \in$ arg $\max_{\pi \in \Pi} \hat{R}(\pi)$, where $\hat{R}(\pi)$ is constructed using (2); and (ii) find an element $\hat{\pi} \in$ arg $\min_{\pi \in \hat{L}\{\hat{R}(\tilde{\pi})\}} \hat{N}(\pi)$.

2.4.1 Approximation of arg $\max_{\pi \in \Pi} \hat{R}(\pi)$

Maximizing $\hat{R}(\pi)$ over $\pi \in \Pi$ is computationally burdensome in problems with more than a handful of covariates because of the indicator functions in (2) and the discreteness of the decision list. However, the tree structure of decision lists suggests a greedy algorithm in the spirit of classification and regression trees (CART, Breiman et al., 1984). Assume that for the jth covariate, there is a candidate set of finitely many possible cutoff values χ_j. These cutoffs might be dictated by clinical guidelines, e.g., if the covariate is a comorbid condition then the thresholds might reflect low, moderate, and high levels of impairment; alternatively, these cutoffs could be chosen to equal empirical or theoretical percentiles of that covariate. There is no restriction imposed on these cutoffs. Let C denote the set of all conditions that induce regions of the form in (4) with the cutoffs $\tau_{jk} \in \chi_{jk}$ for $k = 1, 2, j_k \in \{1, \ldots, p\}$.

Before giving the details of the algorithm, we provide a conceptual overview. The algorithm first uses exhaustive search to find a decision list with exactly one clause, of the form $\pi = \{(c_1, a_1), a_1'\}$, which maximizes $\hat{R}(\pi)$. Let $\{(\tilde{c}_1, \tilde{a}_1), \tilde{a}_1'\}$ denote this decision list. The

algorithm then uses exhaustive search to find the decision list that maximizes $R(\hat{\pi})$ over decision lists with exactly two clauses, the first of which must be either $(\tilde{c_1}, \tilde{a}_1)$ or $(\tilde{c}_1', \tilde{a}_1')$, where $\tilde{c}_1'$ is the negation of $\tilde{c_1}$ such that $T(\tilde{c}_1') = T(\tilde{c}_1)^c$; e.g., if $\tilde{c}_1$ has the form $x_{j1} \leq \tau_1$ and $x_{j2} \leq \tau_2$, then $\tilde{c}_1'$ would be $x_{j1} > \tau_1$ or $x_{j2} > \tau_2$. Although the decision lists $\{(\tilde{c_1}, \tilde{a}_1), \tilde{a}_1'\}$ and $\{(\tilde{c}_1', \tilde{a}_1'), \tilde{a}_1\}$ yield identical treatment recommendations and have the same value, their first clauses are distinct, and may lead to substantially different final decision lists. Hence it is necessary to consider both possibilities for the first clause. The algorithm proceeds recursively by adding one clause at a time until some stopping criterion (described below) is met.

Hereafter, for a decision list π described by $\{(c_1, a_1), \ldots, (c_L, a_L), a_0\}$ for some $L \geq 0$, write $R[\hat{\{}(c_1, a_1), \ldots, (c_L, a_L), a_0\}]$ to denote $R(\hat{\pi})$; e.g., for $L = 0$, $R[\hat{\{}a_0\}]$ is the estimated value of the regime that assigns treatment a_0 to all patients. For any decision list with a vacuous condition, e.g., $\{\cap_{\ell < j} T(c_\ell)^c\} \cap T(c_j) = \varnothing$ for some j, define $R[\hat{\{}(c_1, a_1), \ldots, (c_L, a_L), a_0\}] = -\infty$. Let z_ρ be the 100ρ percentile of the standard normal distribution. Let Π_{temp} denote the set of regimes to which additional clauses can be added, and let Π_{final} denote the set of regimes that have met one of the stopping criteria. The algorithm is as follows, and an illustrative example with a step-by-step run of the algorithm is given in the Web Appendix.

Step 1 Choose a maximum list length $L_{\max}$ and a critical level $\alpha \in (0, 1)$. Compute $\tilde{a}_0 = \arg\max_{a_0 \in A} R[\hat{\{}a_0\}]$. Set $\Pi_{\text{temp}} = \varnothing$, and $\Pi_{\text{final}} = \varnothing$.

Step 2 Compute $(\tilde{c}_1, \tilde{a}_1, \tilde{a}_1') = \arg\max_{(c_1, a_1, a_1') \in C \times A \times A} \hat{R}[\{(c_1, a_1), a_1'\}]$ and

$\hat{\Delta}_1 = \hat{R}[\{(\tilde{c}_1, \tilde{a}_1), \tilde{a}_1'\}] - \hat{R}[\{(\tilde{a}_0)\}]$. If $\hat{\Delta}_1 < z_{1-\alpha}\{\widehat{Var}(\hat{\Delta}_1)\}^{1/2}$ then let $\pi = \{\tilde{a}_0\}$, set $\Pi_{\text{final}} = \{\pi\}$, and go to Step 5; otherwise let $\pi = \{(\tilde{c}_1, \tilde{a}_1), \tilde{a}_1'\}$, $\pi' = \{(\tilde{c}_1', \tilde{a}_1'), \tilde{a}_1\}$ set $\Pi_{\text{temp}} = \{\pi, \pi'\}$, and proceed to Step 3, where $\tilde{c}_1'$ is the negation of $\tilde{c}_1$.

Step 3 Pick an element $\overline{\pi} \in \Pi_{\text{temp}}$, say, $\overline{\pi} = \{(\overline{c}_1, \overline{a}_1), \ldots, (\overline{c}_{j-1}, \overline{a}_{j-1}), \overline{a}_{j-1}'\}$ where $j - 1$ is the length of $\overline{\pi}$. Remove $\overline{\pi}$ from Π_{temp}. With the clauses $(\overline{c}_1, \overline{a}_1), \ldots, (\overline{c}_{j-1}, \overline{a}_{j-1})$ held fixed, compute

$(\tilde{c}_j, \tilde{a}_j, \tilde{a}_j') = \arg\max_{(c_j, a_j, a_j') \in C \times A \times A} \hat{R}[\{(\overline{c}_1, \overline{a}_1), \ldots, (\overline{c}_{j-1}, \overline{a}_{j-1}), (c_j, a_j), a_j'\}]$

and

$\hat{\Delta}_j = \hat{R}[\{(\overline{c}_1, \overline{a}_1), \ldots, (\overline{c}_{j-1}, \overline{a}_{j-1}), (\tilde{c}_j, \tilde{a}_j), \tilde{a}_j'\}] - \hat{R}[\{(\overline{c}_1, \overline{a}_1), \ldots, (\overline{c}_{j-1}, \overline{a}_{j-1}), \overline{a}_{j-1}'\}]$. If $\hat{\Delta}_j < z_{1-\alpha}\{\widehat{Var}(\hat{\Delta}_j)\}^{1/2}$, then let $\pi = \{(\overline{c}_1, \overline{a}_1), \ldots, (\overline{c}_{j-1}, \overline{a}_{j-1}), \overline{a}_{j-1}'\}$, and set $\Pi_{\text{final}} = \Pi_{\text{final}} \cup \{\pi\}$, otherwise if $j = L_{\max}$, let

$\pi = \{(\overline{c}_1, \overline{a}_1), \ldots, (\overline{c}_{j-1}, \overline{a}_{j-1}), (\tilde{c}_j, \tilde{a}_j), \tilde{a}_j'\}$, and set $\Pi_{\text{final}} = \Pi_{\text{final}} \cup \{\pi\}$;

otherwise set $\Pi_{\text{temp}} = \Pi_{\text{temp}} \cup \{\pi, \pi'\}$, where $\tilde{c}_j'$ is the negation of $\tilde{c}_j$,

$\pi = \{(\overline{c}_1, \overline{a}_1), \ldots, (\overline{c}_{j-1}, \overline{a}_{j-1}), (\tilde{c}_j, \tilde{a}_j), \tilde{a}_j'\}$ and

$\pi' = \{(\overline{c}_1, \overline{a}_1), \ldots, (\overline{c}_{j-1}, \overline{a}_{j-1}), (\tilde{c}_j', \tilde{a}_j'), \tilde{a}_j\}$.

Step 4 Repeat Step 3 until Π_{temp} becomes empty.

Step 5 Compute $\tilde{\pi} = \arg\max_{\pi \in \Pi_{\text{final}}} R(\hat{\pi})$. Then $\tilde{\pi}$ is the estimated optimal decision list.

The above description is simplified to illustrate the main ideas. The actual implementation of this algorithm avoids exhaustive searches by pruning the search space $C \times A \times A$. It also avoids explicit construction of Π_{temp} and Π_{final}. Complete implementation details are provided in the Web Appendix. In the algorithm, the decision list stops growing if either the estimated increment in the value, $\hat{\Delta}_j$, is not sufficiently large compared to an estimate of its variation, $\{\widehat{Var}(\hat{\Delta}_j)\}^{1/2}$, or if it reaches the pre-specified maximal length $L_{\max}$. We estimate $Var(\hat{\Delta}_j)$ using large sample theory; the expression is given in the Web Appendix. This variance estimator is a crude approximation, as it ignores uncertainty introduced by the estimation of the decision lists; however, it can be computed quickly, and in simulated experiments it appears sufficient for use in a stopping criterion. The significance level a is a user-chosen tuning parameter. In our simulation experiments, we chose $a = 0.05$; results were not sensitive to this choice (see Web Appendix). To avoid lengthy lists, we set $L_{\max} = 10$. Nevertheless, in our simulations and applications the estimated lists never reach this limit. Finally, it may be desirable in practice to restrict the set of candidate clauses so that, for each j, the number of subjects in $R_j = \{\cap_{\ell < j} T(\hat{c}_\ell)^c\} \cap T(\hat{c}_j)$ exceeds some minimal threshold. This can be readily incorporated into the above algorithm by simply discarding candidate clauses that induce partitions that contain an insufficient number of observations.

The time complexity of the proposed algorithm is $O[2^{L_{\max}}mp^2\{n + (\max_j \#\chi_j)^2\}]$ (see Web Appendix), where $\#\chi_j$ is the number of cutoff values in χ_j. Because $2^{L_{\max}}$ and m are constants that are typically small relative to $p^2\{n + (\max_j \#\chi_j)^2\}$, the time complexity is essentially $O(np^2)$ provided that $\max_j \#\chi_j$ is either fixed or diverges more slowly than $n^{1/2}$. Hence, the time complexity is the same as a single least squares fit, indicating that the proposed algorithm runs very fast and scales well in both dimension p and sample size n.

2.4.2 Finding an element of $\arg\min_{\pi \in \hat{L}\{\hat{R}(\tilde{\pi})\}} \hat{N}(\pi)$—To find an element within the set $\min_{\pi \in L\{\hat{R}(\hat{\pi})\}} N(\tilde{\pi})$, we enumerate all regimes in $L\{R(\hat{\pi})\}$ with length no larger than $L_{\max}$ and select among them the list with the minimal cost. The enumeration algorithm is recursive and requires a substantial amount of bookkeeping; therefore, we describe the basic idea here and defer implementation details to the Web Appendix. Suppose $\tilde{\pi}$ is described by $\{(c_1, \tilde{a}_1), \ldots, (c_L, \tilde{a}_L), \tilde{a}_0\}$. Call a condition of the form $x_j \leq \tau_j$ an atom. There exist $K \leq 2L$ atoms, say $d_1, \ldots, d_K$, such that each clause $c_\ell, \ell = 1, \ldots, L$, can be expressed using the union, intersection, and/or negation of at most two of these atoms. The algorithm proceeds by generating all lists with clauses representable using the foregoing combinations of at most two atoms. To reduce computation time, we use a branch-and-bound scheme (Brusco and Stahl, 2006) that avoids constructing lists with vacuous conditions or those that are provably worse than an upper bound on $\min_{\pi \in L\{\hat{R}(\hat{\pi})\}} N(\pi)$ In the simulation experiments in the next section, the average runtime for the enumeration algorithm was less than one second running on a single core of a 2.3GHz AMD Opteron™ processor and 1GB of DDR3 RAM.

3. Simulation Experiments

We use a series of simulated experiments to examine the finite sample performance of the proposed method. The average value $\mathrm{E}\{R(\hat{\pi})\}$ and the average cost $\mathrm{E}\{N(\hat{\pi})\}$ are the primary performance measures. We consider generative models with (i) binary and continuous outcomes; (ii) binary and trinary treatments; (iii) correctly and incorrectly specified models; and (iv) low- and high-dimensional covariates. The class of data-generating models that we consider is as follows. Covariates are drawn from a p-dimensional Gaussian distribution with mean zero and autoregressive covariance matrix such that $\mathrm{cov}(X_k, X_\ell) = 4(1/5)^{|k-\ell|}$, and the treatments are sampled uniformly so that $P(A = a|X = x) = 1/m$ for all $x \in \mathbb{R}^p$ and $a \in A$. Let $\varphi(x, a)$ be a real-valued function of x and a; given $X = x$ and $A = a$, continuous outcomes are normally distributed with mean $2 + x_1 + x_3 + x_5 + x_7 + \varphi(x, a)$ and variance 1, whereas binary outcomes follows a Bernoulli distribution with success probability expit $\{2 + x_1 + x_3 + x_5 + x_7 + \varphi(x, a)\}$, where expit$(u) = \exp(u)/\{1 + \exp(u)\}$. Table 1 lists the expressions of φ used in our generative models and the number of treatments, m, in A. Under these outcome models, the optimal regime is $\pi^{\mathrm{opt}}(x) = \arg \max_a \varphi(x, a)$.

For comparison, we estimate π^{opt} by parametric Q-learning, nonparametric Q-learning, outcome weighted learning (OWL, Zhao et al., 2012) and modified covariate approach (MCA, Tian et al., 2014). For parametric Q-learning, we use linear regression when Y is continuous and logistic regression when Y is binary. The linear component in the regression model has the form $\sum_{a=1}^{m} I(A{=}a)(1, X^{\mathrm{T}})\beta_a$, where $\beta_1, \ldots, \beta_m$ are unknown coefficient vectors. A LASSO penalty (Tibshirani, 1996) is used to reduce overfitting; the amount of penalization is chosen by minimizing 10-fold cross-validated prediction error. For nonparametric Q-learning, we use support vector regression when Y is continuous and support vector machines when Y is binary (Zhao et al., 2011), both are implemented using a Gaussian kernel. Tuning parameters for non-parametric Q-learning are selected by minimizing 10-fold cross-validated prediction error. For OWL, both linear and Gaussian kernels are used and we follow the same tuning strategy as in Zhao et al. (2012). For MCA, we incorporate the efficiency augmentation term described in Tian et al. (2014). Both OWL and MCA are limited to two treatment options.

To implement our method, the mean model, $\mu(x, a)$, in (1), is estimated as in parametric Q-learning. The propensity score $\omega(x, a)$ is estimated by $n^{-1}\sum_{i=1}^{n} I(A_i{=}a)$. All the comparison methods result in treatment regimes that are more difficult to interpret than a decision list; thus, our intent is to show that decision lists are competitive in terms of the achieved value of the estimated regime, $\mathrm{E}\{R(\hat{\pi})\}$, while being significantly more interpretable and less costly.

Results in Table 2 are based on the average over 1000 Monte Carlo replications with data sets of size $n = 500$ if $m = 2$ and Y is continuous; $n = 750$ if $m = 3$ and Y is continuous; $n = 1000$ if $m = 2$ and Y is binary; and $n = 1500$ if $m = 3$ and Y is binary. The value $R(\hat{\pi})$ and cost $N(\hat{\pi})$ were computed using an independent test set of size 10^6.

Table 2 shows that the decision list is competitive in terms of the value obtained across the entire suite of simulation experiments. If π^{opt} can be represented as a decision list, the proposed method produces the best value. However, even in settings in which the optimal regime is not a decision list, the estimated decision list appears to perform well. Recall that the proposed algorithm attempts to find the best approximation of the optimal regime within the class of regimes that are representable as a decision list. Figure 3 shows the average estimated decision list in misspecified settings II and III with continuous outcome and $p =$ 10. In these settings, the estimated decision list provides a reasonable approximation of the true optimal regime. In addition, the cost of the decision list is notably smaller than the cost of the parametric Q-learning estimator or the MCA estimator. Nonparametric Q-learning OWL always use all covariates, so their costs are always equal to p.

In the Web Appendix, we derive point estimates and prediction intervals for $R(\hat{\pi})$. We also present simulation results to illustrate the accuracy of variable selection for the decision list.

4. Applications

4.1 Breast Cancer Data

Gail and Simon (1985) compared the treatment effects of chemotherapy alone and chemotherapy with tamoxifen using data collected from the NSABP trial. Their regime recommended chemotherapy alone to patients with age 50 and PR 10 and chemotherapy plus tamoxifen to all others. Because the variables involved in the treatment regime constructed by Gail and Simon were chosen using clinical judgment, it is of interest to see what regime emerges from a more data-driven procedure. Thus, we use the proposed method to estimate an optimal treatment regime in the form of a decision list using data from the NSABP trial.

As in Gail and Simon (1985), we take three-year disease-free survival as the outcome, so that $Y = 1$ if the subject survived disease-free for three years after treatment, and $Y = 0$ otherwise. Patient covariates are age (years), PR (fmol), estrogen receptor level (ER, fmol), tumor size (centimeters), and number of histologically positive nodes (number of nodes, integer). We estimated the optimal treatment regime representable as a decision list using data from the 1164 subjects with complete observations on these variables. Because treatment assignment was randomized in NSABP, we estimated $\omega(x, a)$ by the sample proportion of subjects receiving treatment a. Based on exploratory analyses, we estimated $\mu(x, a)$ using a logistic regression model with transformed predictors $z = z(x) = \{$age, log(1 + PR), log(1 + ER), tumor-size, log(1 + number-of-nodes)$\}^{\text{T}}$.

The estimated optimal treatment regime representable as a decision list is given in the top panel of Figure 4; the regime estimated by Gail and Simon is given in the bottom panel of this figure. The structure of the two treatment regimes is markedly similar. The treatment recommendations from the two regimes agree for 92% of the patients in the NSABP data. In this data set, 33% of the patients have a PR value less than 3; 13% of the patients have a PR values between 3 and 10; and 54% of the patients have a PR value greater than 10.

In a previous analysis of the NSABP data, Zhang et al. (2012) recommended that patients with age + 7.98 log(1 + PR) ≥ 60 receive chemotherapy alone and all others receive chemotherapy plus tamoxifen. However, this regime was built using only age and PR as potential predictors with no data-driven variable selection. In contrast, the proposed method selects age and PR from the list of potential predictors. For completeness, we also implemented parametric Q-learning using a logistic regression model with covariate vector z. The estimated regime recommends chemotherapy alone if 1.674 – 0.021 age – 0.076 log(1+PR) – 0.116 log(1 + ER) – 0.024 tumor-size – 0.274 log(1 + number-of-nodes) ≥ 0 and chemotherapy with tamoxifen otherwise. The treatment recommendation dictated by parametric Q-learning agrees with that dictated by decision list for 86% of the subjects in the data set.

To estimate the survival probability under each estimated regime, we use cross-validation. The data set was randomly divided into a training set containing 80% of the subjects and a test set containing 20% of the subjects. The optimal regime was estimated using both approaches on the training set, and its value was computed using (2) (with $\hat{\mu} \equiv 0$) on the test set. To reduce variability, this process was repeated 100 times. The estimated survival probability is 0.65 for the regime representable as decision list and 0.66 for the regime obtained from parametric Q-learning. Thus, the proposed method greatly improves interpretability while preserving quality.

4.2 Chronic Depression Data

Keller et al. (2000) compared nefazodone, psychotherapy, and combination of nefazodone and psychotherapy for treating patients with chronic depression in a three-arm randomized clinical trial. Among the three treatments considered, combination therapy was shown to be the most beneficial in terms of efficacy as measured by the Hamilton Rating Scale for Depression score (HRSD). However, the combination treatment is significantly more expensive and burdensome than monotherapy. Therefore, it is of interest to construct a treatment regime that recommends combination therapy only to subjects for whom there is a significant benefit over monotherapy.

Because lower HRSD indicates less severe symptoms, we define outcome $Y = -$HRSD to be consistent with our paradigm of maximizing the mean outcome. Patient covariates comprise 50 pretreatment variables, including personal habits and difficulties, medication history and various scores from several psychological questionnaires; a list of these variables is given in the Web Appendix. We estimate an optimal regime using data from the $n = 647$ (of 680 enrolled) subjects in the clinical trial with complete data. Because treatments were randomly assigned, we estimated $\omega(x, a)$ by the sample proportion of subjects receiving treatment a. We estimated $\mu(x, a)$ using a penalized linear regression model with all patient covariates and treatment by covariate interactions. Penalization was implemented with a LASSO penalty tuned using 10-fold cross-validated prediction error.

The estimated optimal treatment regime representable as a decision list is displayed in Figure 1. One explanation for this rule is as follows. Those with strong physical anxiety symptoms (somatic) and significant cognitive impairment (retardation) may be unlikely to benefit from psychotherapy alone or in combination with nefazodone and are therefore

recommended to nefazodone alone. Otherwise, because psychotherapy is a primary tool for treating anxiety (HAM-A) and nefazodone is associated with sleep disturbance (sleep), it may be best to assign subjects with moderate to severe anxiety and severe sleep disturbance to psychotherapy alone. All others are assigned to the combination therapy.

The estimated regime contains only four covariates. In contrast, the regime estimated by parametric *Q*-learning using linear regression and LASSO penalty involves a linear combination of twenty-four covariates, making it difficult to explain and expensive to implement. To compare the quality of these two regimes, we use random-split cross-validation as in Section 4.1. The estimated HRSD score under the regime representable as decision list is 12.9, while that under the regime estimated by parametric *Q*-learning is 11.8. Therefore, by using decision lists we are able to obtain a remarkably more parsimonious regime with high quality, which facilitates easier interpretation.

5. Discussion

Data-driven treatment regimes have the potential to improve patient outcomes and generate new clinical hypotheses. Estimation of an optimal treatment regime is typically conducted as a secondary, exploratory analysis aimed at building knowledge and informing future clinical research. Thus, it is important that methodological developments are designed to fit this exploratory role. Decision lists are a simple yet powerful tool for estimation of interpretable treatment regimes from observational or experimental data. Because decision lists can be immediately interpreted, clinical scientists can focus on the scientific validity of the estimated treatment regime. This allows the communications between the statistician and clinical collaborators to focus on the science rather than the technical details of a statistical model.

Due to the "if-then" format and the conditions given in (4), the estimated regime, as a function of the data, is discrete. Thus, a theoretical proof of the consistency of the treatment recommendations using decision lists is heavily technical and will be presented elsewhere. We provide some empirical evidence in the Web Appendix that the estimated regime gives consistent treatment decisions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Baker SG, Cook NR, Vickers A, Kramer BS. Using relative utility curves to evaluate risk prediction. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2009; 172:729–748.

Breiman, L.; Friedman, JH.; Olshen, RA.; Stone, CJ. Classification and Regression Trees. New York: CRC Press; 1984.

Brusco, MJ.; Stahl, S. Branch-and-Bound Applications in Combinatorial Data Analysis. New York: Springer; 2006.

Clark P, Niblett T. The CN2 induction algorithm. Machine Learning. 1989; 3:261–283.

Freitas AA. Comprehensible classification models: a position paper. ACM SIGKDD Explorations Newsletter. 2014; 15:1–10.

Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. Biometrics. 1985; 41:361–372. [PubMed: 4027319]

Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. Journal of the National Cancer Institute. 2009; 101:959–963. [PubMed: 19535781]

Gail MH, Costantino JP, Bryant J, Croyle R, Freedman L, Helzlsouer K, Vogel V. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. Journal of the National Cancer Institute. 1999; 91:1829–1846. [PubMed: 10547390]

Huang Y, Laber EB, Janes H. Characterizing expected benefits of biomarkers in treatment selection. Biostatistics. 2015; 16:383–399. [PubMed: 25190512]

Keller MB, McCullough JP, Klein DN, Arnow B, Dunner DL, Gelenberg AJ, et al. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. New England Journal of Medicine. 2000; 342:1462–1470. [PubMed: 10816183]

Laber E, Zhao Y. Tree-based methods for personalized treatment regimes. Biometrika. 2015 to appear.

Laber EB, Linn KA, Stefanski LA. Interactive model building for Q-learning. Biometrika. 2014; 101:831–847. [PubMed: 25541562]

Letham, B.; Rudin, C.; McCormick, TH.; Madigan, D. Building interpretable classifiers with rules using Bayesian analysis. Department of Statistics; University of Washington: 2012. Technical Report TR609

Marchand M, Sokolova M. Learning with decision lists of data-dependent features. Journal of Machine Learning Research. 2005; 6:427–451.

Marlowe DB, Festinger DS, Dugosh KL, Benasutti KM, Fox G, Croft JR. Adaptive programming improves outcomes in drug court an experimental trial. Criminal justice and behavior. 2012; 39:514–532. [PubMed: 22923854]

Moodie EE, Chakraborty B, Kramer MS. Q-learning for estimating optimal dynamic treatment rules from observational data. Canadian Journal of Statistics. 2012; 40:629–645. [PubMed: 23355757]

Moodie EEM, Dean N, Sun YR. Q-learning: Flexible learning about useful utilities. Statistics in Biosciences. 2013; 6:1–21.

Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: Main content. The International Journal of Biostatistics. 2010; 6

Qian M, Murphy SA. Performance guarantees for individualized treatment rules. Annals of Statistics. 2011; 39:1180–1210. [PubMed: 21666835]

Rivest RL. Learning decision lists. Machine Learning. 1987; 2:229–246.

Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. Communications in Statistics-Theory and methods. 1994; 23:2379–2412.

Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000; 11:550–560. [PubMed: 10955408]

Robins JM, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. Statistics in Medicine. 2008; 27:4678–4721. [PubMed: 18646286]

Shiffman RN. Representation of clinical practice guidelines in conventional and augmented decision tables. Journal of the American Medical Informatics Association. 1997; 4:382–393. [PubMed: 9292844]

Shortreed SM, Laber EB, Lizotte DJ, Stroup TS, Pineau J, Murphy SA. Informing sequential clinical decision-making through reinforcement learning: An empirical study. Machine Learning. 2011; 84:109–136. [PubMed: 21799585]

Shortreed SM, Laber EB, Stroup TS, Pineau J, Murphy SA. A multiple imputation strategy for sequential multiple assignment randomized trials. Statistics in Medicine. 2014; 33:4202–4214. [PubMed: 24919867]

Su X, Tsai C-L, Wang H, Nickerson DM, Li B. Subgroup analysis via recursive partitioning. Journal of Machine Learning Research. 2009; 10:141–158.

Taylor JMG, Cheng W, Foster JC. Reader reaction to "A Robust Method for Estimating Optimal Treatment Regimes" by Zhang et al (2012). Biometrics. 2014 In press.

Tian L, Alizadeh AA, Gentles AJ, Tibshirani R. A simple method for estimating interactions between a treatment and a large number of covariates. Journal of the American Statistical Association. 2014; 109:1517–1532. [PubMed: 25729117]

Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B. 1996; 58:267–288.

Tsiatis, AA. Semiparametric Theory and Missing Data. New York: Springer; 2006.

Vansteelandt S, Joffe M, et al. Structural nested models and g-estimation: The partially realized promise. Statistical Science. 2014; 29:707–731.

Wang F, Rudin C. Falling rule lists. arXiv preprint arXiv:1411.5899. 2014

Zhang B, Tsiatis AA, Davidian M, Zhang M, Laber E. Estimating optimal treatment regimes from a classification perspective. Stat. 2012; 1:103–114. [PubMed: 23645940]

Zhang B, Tsiatis AA, Laber EB, Davidian M. A robust method for estimating optimal treatment regimes. Biometrics. 2012; 68:1010–1018. [PubMed: 22550953]

Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. Statistics in Medicine. 2009; 28:3294–3315. [PubMed: 19750510]

Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. Journal of the American Statistical Association. 2012; 107:1106–1118. [PubMed: 23630406]

Zhao Y, Zeng D, Socinski MA, Kosorok MR. Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. Biometrics. 2011; 67:1422–1433. [PubMed: 21385164]

Zhao YQ, Zeng D, Laber EB, Song R, Yuan M, Kosorok MR. Doubly robust learning for estimating individualized treatment with censored data. Biometrika. 2015; 102:151–168. [PubMed: 25937641]
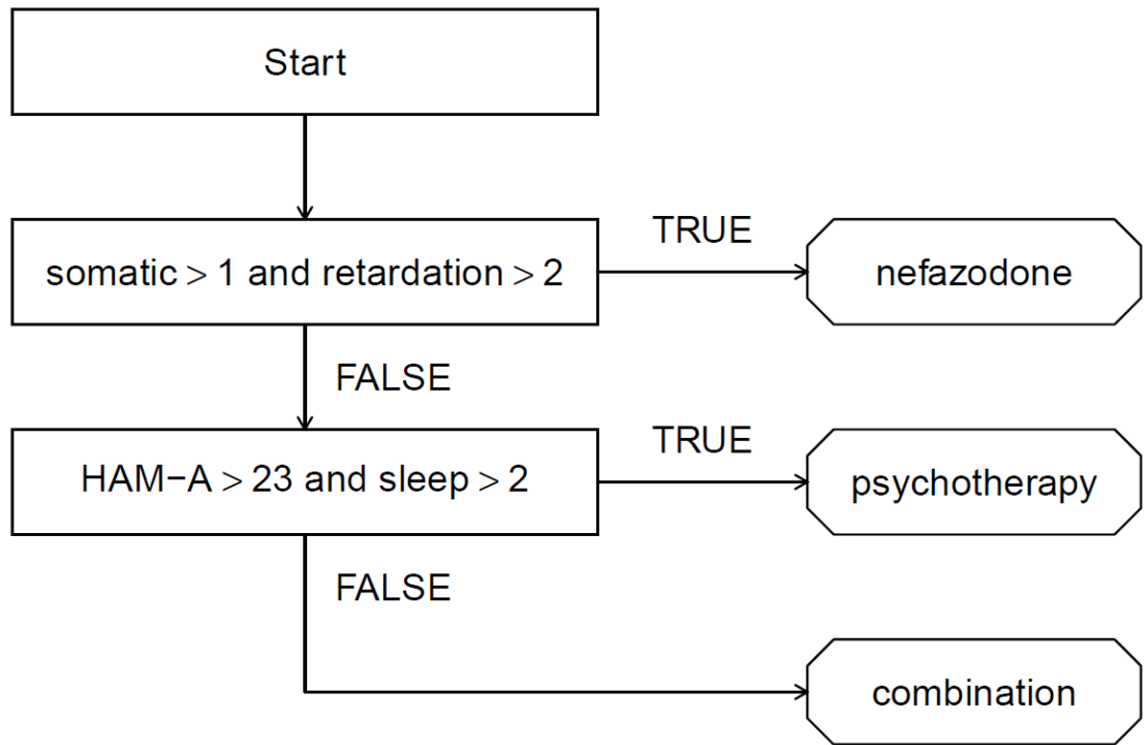
Figure 1.
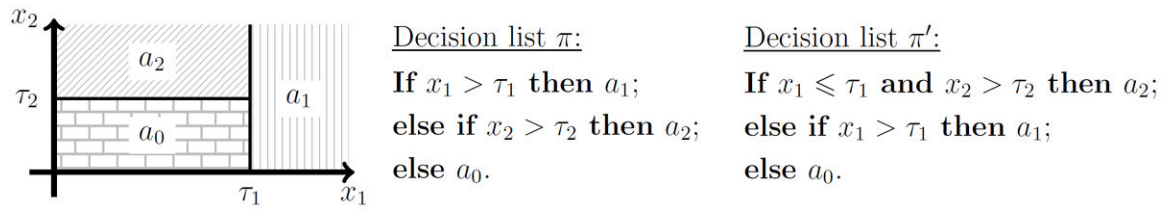Estimated decision list for treating patients with chronic depression.

Figure 2.

Left: diagram of a decision list dictated by regions $R_1 = \{x \in \mathbb{R}^2 : x_1 > \tau_1\}$, $R_2 = \{x \in \mathbb{R}^2 : x_1 \leqslant \tau_1, x_2 > \tau_2\}$, and $R_0 = \{x \in \mathbb{R}^2 : x_1 \leqslant \tau_1, x_2 \leqslant \tau_2\}$, and treatment recommendations a_1, a_2, and a_0. **Middle:** representation of the decision list that requires only x_1 in the first clause. **Right:** alternative representation of the same decision list that requires both x_1 and x_2 in the first clause.
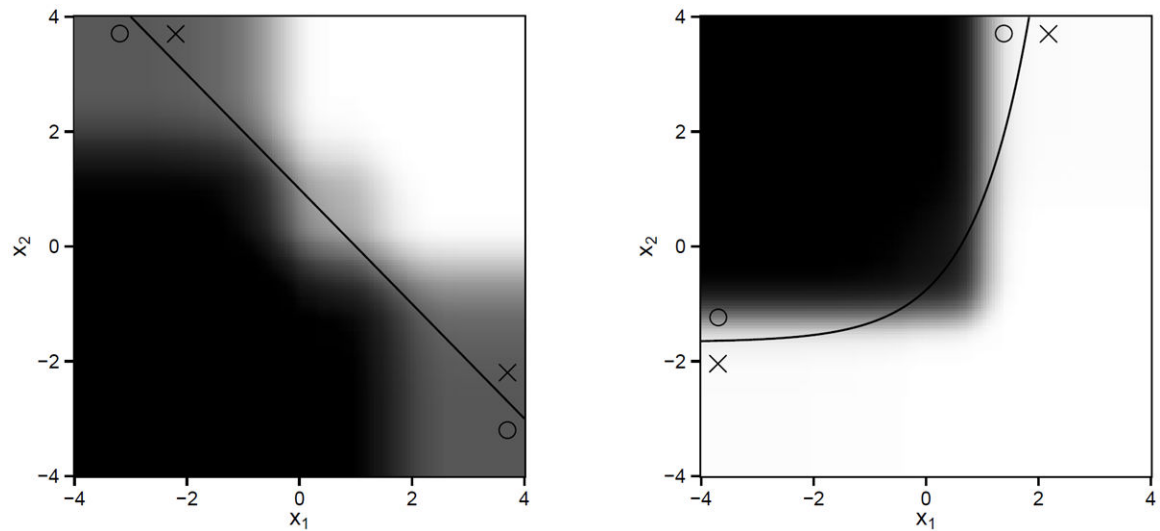
Figure 3.
Left: average estimated regimes under setting II. **Right:** average estimated regimes under setting III. In both settings π^{opt} cannot be represented as decision list. The solid line is the treatment decision boundary under π^{opt}. The region where treatment 1 is better than treatment 2 is marked by circles, while the region where treatment 2 is better than treatment 1 is marked by crosses. For every point $(x_1, x_2)^{\mathrm{T}}$, we compute the proportion of 1000 replications that the estimated regime recommends treatment 1 to a patient with covariate $(x_1, x_2, 0, \ldots, 0) \in \mathbb{R}^{10}$. The larger the proportion, the darker the shade.
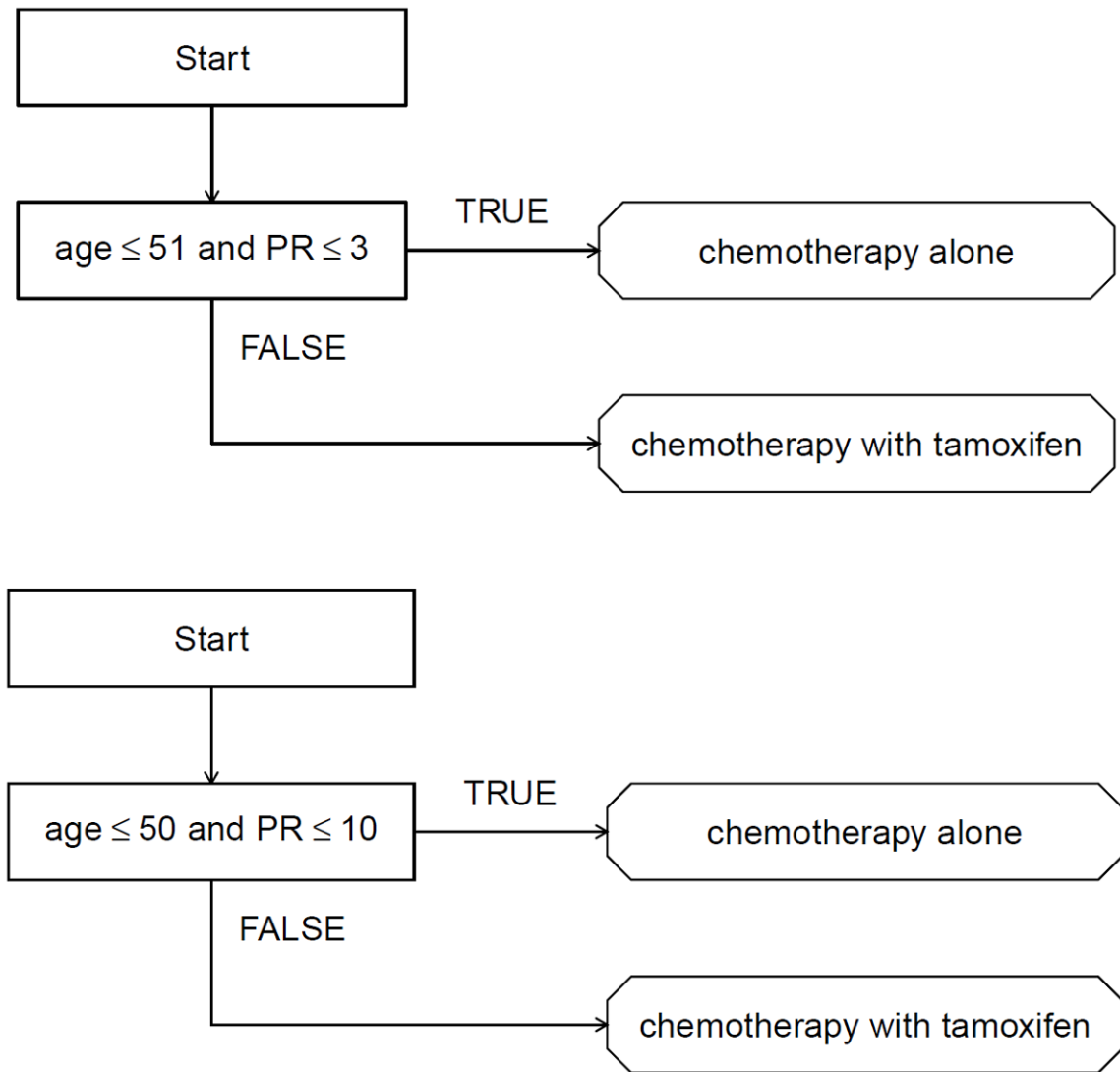
Figure 4.
Top: estimated optimal treatment regime representable as a decision list. **Bottom:** treatment regime proposed by Gail and Simon (1985).

Table 1

The second column gives the number of treatment options m. The third column gives the set of φ functions used in the outcome models. The fourth column specifies the form of the optimal regime $\pi^{opt}(x) = arg\ max_a$ $\varphi(x, a)$ where: "linear" indicates that $\pi^{opt}(x) = arg\ max_a\{(I, x^T)\beta_a\}$ for some coefficient vectors $\beta_a \in \mathbb{R}^{p+1}$, $a \in$ A, "decision list" indicates that π^{opt} is representable as a decision list; and "nonlinear" indicates that $\pi^{opt}(x)$ is neither linear nor representable as a decision list.

Setting	m	Expression of φ	Form of π^{opt}
I	2	$\varphi_1(x, a) = I(a = 2)\{3I(x_1 \quad 1, x_2 > -0.6) -1\}$	decision list
II	2	$\varphi_2(x, a) = I(a = 2)(x_1 + x_2 - 1)$	linear
III	2	$\varphi_3(x, a) = I(a = 2)\ \arctan(\exp(1+x_1) -3x_2 -5)$	nonlinear
IV	2	$\varphi_4(x, a) = I(a = 2)(x_1 - x_2 + x_3 - x_4)$	linear
V	3	$\varphi_5(x, a) = I(a = 2)\{4I(x_1 > 1) -2\} + I(a = 3)I(x_1 \quad 1)\{2I(x_2 \quad -0.3) -1\}$	decision list
VI	3	$\varphi_6(x, a) = I(a = 2)(2x_1) + I(a = 3)(-x_1 x_2)$	nonlinear
VII	3	$\varphi_7(x, a) = I(a = 2)(x_1 - x_2)+ I(a = 3)(x_3 - x_4)$	linear

Table 2

The average value and the average cost of estimated regimes in simulated experiments. In the header, p is the dimension of patient covariates; DL refers to the proposed method using decision list; Q_1 refers to parametric Q-learning; Q_2 refers to nonparametric Q-learning; OWL_1 and OWL_2 refer to outcome weighted learning with linear kernel and Gaussian kernel, respectively; MCA refers to modified covariate approach with efficiency augmentation. OWL and MCA are not applicable under Setting V, VI and VII.

p	Setting	Value						Cost		
		DL	Q_1	Q_2	OWL_1	OWL_2	MCA	DL	Q_1	MCA
Continuous response										
	I	2.78	2.53	2.53	2.33	2.29	2.54	1.64	9.0	5.1
	II	2.70	2.80	2.79	2.61	2.54	2.80	1.64	9.0	5.1
	III	2.59	2.54	2.53	2.29	2.24	2.55	1.68	9.1	4.9
10	IV	2.89	3.37	3.35	3.16	3.09	3.37	2.50	9.5	7.4
	V	2.90	2.67	2.59	–	–	–	1.90	9.5	–
	VI	3.98	3.46	3.95	–	–	–	1.61	9.2	–
	VII	3.22	3.75	3.73	–	–	–	2.56	9.7	–
	I	2.76	2.51	2.36	2.21	2.19	2.53	1.80	21.3	9.2
	II	2.70	2.79	2.73	2.26	2.27	2.79	1.64	21.4	9.3
	III	2.59	2.52	2.35	2.16	2.12	2.54	1.71	23.1	9.0
50	IV	2.89	3.36	3.27	2.76	2.70	3.36	2.53	25.4	14.9
	V	2.87	2.63	2.33	–	–	–	2.14	28.5	–
	VI	3.95	3.43	3.47	–	–	–	1.69	26.6	–
	VII	3.21	3.74	3.61	–	–	–	2.55	30.8	–
Binary response										
	I	0.77	0.74	0.69	0.73	0.73	0.74	1.94	8.9	4.1
	II	0.71	0.72	0.60	0.71	0.71	0.72	1.69	9.2	5.3
	III	0.73	0.73	0.68	0.72	0.72	0.73	2.10	9.2	4.7
10	IV	0.71	0.76	0.66	0.75	0.74	0.75	2.40	9.6	8.4
	V	0.75	0.73	0.62	–	–	–	2.52	9.6	–
	VI	0.79	0.75	0.64	–	–	–	2.09	9.5	–

p	Setting	Value						Cost		
		DL	Q_1	Q_2	OWL$_1$	OWL$_2$	MCA	DL	Q_1	MCA
	VII	0.77	0.81	0.69	–	–	–	2.83	9.9	–
	I	0.76	0.73	0.69	0.71	0.70	0.73	2.64	21.9	8.3
	II	0.71	0.72	0.60	0.70	0.69	0.71	1.87	26.2	6.4
	III	0.73	0.72	0.67	0.70	0.69	0.72	2.53	25.0	7.3
50	IV	0.71	0.76	0.66	0.73	0.72	0.74	2.55	31.0	13.8
	V	0.74	0.72	0.61	–	–	–	3.15	30.4	–
	VI	0.78	0.75	0.63	–	–	–	2.41	29.6	–
	VII	0.76	0.81	0.68	–	–	–	2.97	35.7	–