HHS Public Access

Author manuscript

Biometrika. Author manuscript; available in PMC 2016 February 16.

Published in final edited form as:

Biometrika. 2015 ; 102(3): 501–514. doi:10.1093/biomet/asv028.

Tree-based methods for individualized treatment regimes

E. B. Laber and

Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, North Carolina 27695, U.S.A.

Y. Q. Zhao

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin 53792, U.S.A.

E. B. Laber: laber@stat.ncsu.edu; Y. Q. Zhao: yqzhao@biostat.wisc.edu

Summary

Individualized treatment rules recommend treatments on the basis of individual patient characteristics. A high-quality treatment rule can produce better patient outcomes, lower costs and less treatment burden. If a treatment rule learned from data is to be used to inform clinical practice or provide scientific insight, it is crucial that it be interpretable; clinicians may be unwilling to implement models they do not understand, and black-box models may not be useful for guiding future research. The canonical example of an interpretable prediction model is a decision tree. We propose a method for estimating an optimal individualized treatment rule within the class of rules that are representable as decision trees. The class of rules we consider is interpretable but expressive. A novel feature of this problem is that the learning task is unsupervised, as the optimal treatment for each patient is unknown and must be estimated. The proposed method applies to both categorical and continuous treatments and produces favourable marginal mean outcomes in simulation experiments. We illustrate it using data from a study of major depressive disorder.

Some key words

Continuous treatment; Exploratory analysis; Personalized medicine; Treatment regime; Tree-based method

1. Introduction

Individualized treatment rules are increasingly being used by clinical and intervention scientists to account for patient response heterogeneity (e.g., Ludwig & Weinstein, 2005; Hayes et al., 2007; Allegra et al., 2009; Cummings et al., 2010). These treatment rules belong to the new era of personalized medicine (Piquette-Miller & Grant, 2007; Hamburg & Collins, 2010). There is a vast literature on estimation of treatment rules that maximize the mean of a desirable clinical outcome using data from randomized or observational studies.

Supplementary material

Supplementary material available at *Biometrika* online includes: a list of potential predictors for the case study in § 4; proofs of the technical results in § 2; a detailed description of the tree-growing algorithm, referenced in § 2; and computer code implementing the minimum impurity decision assignments algorithm.

Regression-based methods model the response as a function of patient characteristics and treatment, and select treatments that maximize the predicted mean outcome (Brinkley et al., 2010; Qian &Murphy, 2011). However, because such methods indirectly infer the optimal treatment rule through a regression model, an interpretable, parsimonious treatment rule can be obtained only via a simple regression model which is subject to misspecification; on the other hand, a complex regression model may mitigate the risk of misspecification but at the cost of producing an unintelligible treatment rule (Zhang et al., 2012b). Policy-search or direct-maximization methods offer an alternative to regression-based methods that attempt to search for the best treatment rule in a large class of potential rules (Zhang et al., 2012a, 2012b; Zhao et al., 2012; Zhang et al., 2013), thereby separating the class of decision rules from an underlying regression model. However, without an interpretable representation of the model behind these approaches, clinical investigators may be hesitant to use the estimated treatment rule to inform clinical practice or future research.

Since Breiman et al. (1984) introduced the classification and regression tree algorithm, tree-based methods have enjoyed a great deal of popularity in statistical and machine learning research, largely due to the interpretability and communicability of decision trees. Indeed, trees have been advocated as a tool for representing more complex prediction models to laymen (Craven & Shavlik, 1996). The classification and regression tree algorithm recursively partitions the covariate space into rectangular sets and then fits a simple model within each partition to the response (Breiman et al., 1984; Hastie et al., 2009, ch. 9.2); thus, the classification and regression tree algorithm provides a flexible nonparametric procedure to explore the underlying model structure (Ripley, 1996; Sutton, 2005).

Tree-based methods have been used in personalized medicine primarily for the purpose of identifying subgroups of subjects with outlying, large or small, treatment effects or strong adverse side-effects relative to subjects in some reference population of interest. Such methods include interaction trees (Su et al., 2008, 2009), virtual twins (Foster et al., 2011), and subgroup identification based on differential effect search (Lipkovich et al., 2011). Existing subgroup identification methods aim primarily to find interactions between treatment and covariates. Zhang et al. (2012a) recast treatment selection with binary treatments as a classification problem and used the classification and regression tree algorithm as an illustrative example.

We present a general purpose approach to estimating optimal personalized treatment rules representable as decision trees. The proposed method can be used with high-dimensional covariates, discrete or continuous treatments, and data from observational or randomized studies. In the case of continuous treatments, standard methods for estimating the mean outcome under a specified treatment rule, such as inverse probability weighting, cannot be applied because a required absolute continuity condition does not hold. We derive a novel kernel smoother for estimating the mean outcome in the case of continuous treatments by approximating a deterministic treatment rule with a stochastic one. The proposed estimator relies on a bandwidth parameter, and we derive a plug-in estimator of the optimal bandwidth. The proposed algorithms are available as part of the R package MIDAs (R Development Core Team, 2015) and are provided in the Supplementary Material.

2. Minimum impurity decision assignments decision trees

2·1. Optimal individualized treatment rules

We observe $\mathcal{D}=\{(X_i, A_i, Y_i)\}_{i=1}^n$, comprising n independent identically distributed triples (X, A, Y) where $X \in \mathbb{R}^p$ denotes the baseline subject characteristics, $A \in \mathcal{A}$ represents the treatment received, and can be discrete or continuous, and $Y \in \mathbb{R}$ is an outcome coded so that higher values are more desirable. An individualized treatment rule is a map $\pi : \mathbb{R}^p \to \mathcal{A}$ such that a patient presenting with $X = x$ is assigned treatment $\pi(x)$. Let $Y^*(a)$ denote the potential outcome under treatment $a \in \mathcal{A}$ (Rubin, 1978), and define $Y^*(\pi) = Y^*\{\pi(X)\}$ to be the potential outcome under π. The performance measure of π is the marginal mean outcome $E\{Y^*(\pi)\}$, and the optimal rule, π^{opt}, satisfies $E\{Y^*(\pi^{\text{opt}})\} \geq E\{Y^*(\pi)\}$ for all π. Let $p(a \mid X)$ denote the conditional density of A given X, with respect to an appropriate dominating measure. We make the following assumptions.

Assumption 1 (Positivity)—There exists $\varepsilon > 0$ such that $p(a \mid X) \geq \varepsilon$ with probability 1 for all $a \in \mathcal{A}$.

Assumption 2 (Strong ignorability)—The potential outcomes $\{Y^*(a) : a \in \mathcal{A}\}$ are conditionally independent of A given X.

Assumption 3 (Consistency)—We have $Y = Y^*(A)$.

These assumptions are standard and will allow us to connect the potential outcomes with the observed data. Let E^{π} denote the expectation with respect to (X, A, Y) under the restriction that $A = \pi(X)$, i.e., that all patients are assigned treatments according to π; then, under Assumptions 1–3, it can be shown (Zhang et al., 2012b) that the marginal mean outcome under π is equal to $E^{\pi}(Y)$. We use this representation to construct an estimator of π^{opt} that applies to either observational or randomized study data.

Unlike traditional decision tree problems, the target of estimation, $\pi^{\text{opt}}(x)$, is not directly observed for the associated patient characteristics $X = x$. For example, in a classification problem a correct label $Y = y$ is observed for each observed $X = x$; similarly, in a regression problem an outcome $Y = y$ is observed for each $X = x$. In the treatment selection problem, information about $\pi^{\text{opt}}(x)$ is available only indirectly through the outcome $Y = y$. Thus, any purity measure used to construct splits in a decision tree must make use of this indirect information. We develop purity measures for discrete and continuous treatments and then use these purity measures in a recursive algorithm to estimate an optimal tree-based individualized treatment rule.

2·2. Purity measures for treatment allocation

We first consider the binary treatment setting, where $\mathcal{A} = \{0, 1\}$; generalizations are given below. In the binary treatment case, any decision rule π partitions the domain of X, $\mathbb{R}^p$, into two regions: $\mathcal{R}_0 = \{x \in \mathbb{R}^p : \pi(x) = 0\}$ and $\mathcal{R}_1 = \mathbb{R}^p \setminus \mathcal{R}_0 = \{x \in \mathbb{R}^p : \pi(x) = 1\}$. Let $\mathcal{R}_0^{\text{opt}}$ and $\mathcal{R}_1^{\text{opt}}$ be the partition of $\mathbb{R}^p$ induced by the optimal decision rule π^{opt}. Then identification of π^{opt} is equivalent to identifying $\mathcal{R}_0^{\text{opt}}$ and $\mathcal{R}_1^{\text{opt}}$. For a set of triples $\mathcal{O} = \{(j,$

$\tau, b)\}$ where $j \in \{1, \ldots, p\}$, $\tau \in \mathbb{R}$ and $b \in \{-1, 1\}$, we say that r is the rectangle defined by $\mathscr{O}$ if $x \in r$ if and only if $b(x_j - \tau) \geq 0$. The rectangle defined by $\emptyset$ is taken to be $\mathbb{R}^p$. We define a rectangular region as any finite combination of intersections and unions of rectangles. Both rectangles and their complements are rectangular regions. For simplicity of notation, we assume that X is a continuous random vector; this avoids having to distinguish between closed and open rectangles. A tree-based approach estimates the sets $\mathscr{R}_0^{\text{opt}}$ and $\mathscr{R}_1^{\text{opt}}$ using rectangular regions in $\mathbb{R}^p$. Figure 1 shows an example of a decision rule composed of rectangles $r_1 = \{(1, 1, 1)\}$ and $r_2 = \{(2,-3,-1)\}$ with rectangular regions $\mathscr{R}_0 = r_1^c \cap r_2$ and $\mathscr{R}_1 = (r_1^c \cap r_2) \cup r_1$. The tree in Fig. 1 assigns treatment 0 to subjects presenting covariates $X = x$ which satisfy $x_1 < 1$ and $x_2 \leq -3$ and assigns treatment 1 otherwise. While this tree resembles a classification tree, with labels 0 and 1, it is fundamentally different in that the decision rule does not attempt to describe the rule by which the observed treatments were assigned but rather the rule by which treatments should be assigned to future patients.

To form a tree-based estimator of π^{opt}, we need a measure of node purity that will facilitate a recursive splitting procedure. In determining how to create two child nodes from a parent node determined by, say, the rectangular region $\mathscr{R}$ during the process of recursive partitioning, the general goal is to make the data corresponding to each of the child nodes more pure than the data in the parent node (Sutton, 2005). Intuitively, the first split in the tree is found by determining the rectangle r such that r and r^c best approximate $\mathscr{R}_0^{\text{opt}}$ and $\mathscr{R}_1^{\text{opt}}$, respectively. Recursively, for a given terminal node $\mathscr{R}$, we seek the rectangle r such that splitting $\mathscr{R}$ to form two new terminal nodes $\mathscr{R} \cap r$ and $\mathscr{R} \cap r^c$ will most dramatically improve our current estimates of $\mathscr{R}_0^{\text{opt}}$ and $\mathscr{R}_1^{\text{opt}}$. A node purity measure provides a criterion to formalize the foregoing search procedure. We first describe a measure of node purity when there are a finite number of treatments, and then extend this measure to the continuous treatment case.

2·3. Purity measures for discrete treatments

In the discrete treatment case, the set of treatments is finite and coded so that $\mathscr{A} = \{0, 1, \ldots, K\}$. For $a \in \mathscr{A}$ and $x \in \mathbb{R}^p$, let $p(a \mid x)$ denote $\text{pr}(A = a \mid X = x)$. We assume that the function $p(a \mid x)$ is known and that $p(a \mid X)$ is bounded away from 0 and 1 with probability 1 for each $a \in \mathscr{A}$. If these probabilities are not known, as is the case with observational data, they may be estimated from the data, for example by using a multinomial logistic regression. Recall that the performance measure of a rule π is the expected outcome when patients are assigned treatments according to π. Under the foregoing assumptions, we can apply a change of measure to express the performance of π in terms of the observed data. For any function $g : \mathbb{R}^p \to \mathbb{R}$ and policy π, define the random variable $L_g(\pi) = L_g(\pi, X, A, Y) = \{Y - g(X)\}1_{\pi(X)=A}/p\{\pi(X) \mid X\}$ and let $C_g(\pi) = E\{L_g(\pi)\}$; then it can be shown that $E^{\pi}(Y) = C_0(\pi)$, where 0 denotes the function $g(x) \equiv 0$ (Zhang et al., 2012a, 2012b; Zhao et al., 2012). However, for any fixed function g, $\pi^{\text{opt}} = \arg\max_{\pi} C_g(\pi)$, because

$$C_g(\pi) = E\left[\frac{Y1_{\pi(X)=A}}{p\{\pi(X)|X\}}\right] - E\left(E\left[\frac{g(X)1_{\pi(X)=A}}{p\{\pi(X)|X\}}\middle|X\right]\right) = C_0(\pi) - E\{g(X)\},$$

so that the arg max over π depends only on $C_0(\pi)$. One choice for g is $g(x) = E\{Y \mid X = x, A = \tilde{\pi}(x)\}$ for some reference rule $\tilde{\pi}$; it turns out that this choice minimizes the variance of $L_g(\pi)$ when $\pi = \tilde{\pi}$. Hereafter, we write $E\{Y \mid X, \pi(X)\}$ as shorthand for $E\{Y \mid X = x, A = \pi(x)\}|_{x=X}$. The following result is proved in the Supplementary Material.

Lemma 1—*Assume that* $p(a \mid X) \in [\varepsilon, 1 - \varepsilon]$ *with probability* 1 *for some* $\varepsilon > 0$ *and all* $a \in \mathscr{A}$. *Then, for any* $g : \mathbb{R}^p \to \mathbb{R}$ *and any rule* π, $\text{var}\{L_g(\pi)\} \geq \text{var}\{L_{E\{Y \mid X, A=\pi(X)\}}(\pi)\}$.

Because our goal is to estimate π^{opt}, it is natural to use a crude estimator of π^{opt} as a reference rule. The reference rule is used to reduce variance and does not affect consistency. Thus, in practice, a simple, convenient estimator of π^{opt} can be used as the reference rule. Hereafter, we write $C(\pi)$ for $C_{E\{Y \mid X, \pi^{\text{opt}}(X)\}}(\pi)$. Let $m(x)$ denote an estimator of $E\{Y \mid X = x, A = \pi^{\text{opt}}(x)\}$; we obtain an estimator of m by first constructing an estimator of $E(Y \mid X = x, A = a)$, say $Q(x, a)$, by regressing Y on X and A using a flexible model, and then defining $\hat{m}(x) = \max_a \hat{Q}(x, a)$. In our simulations, we estimated $E(Y \mid X = x, A = a)$ using random forests (Breiman, 2001), although other methods are possible. For an arbitrary rule π, the plug-in estimator of $C(\pi)$ is $\hat{C}(\pi) = E_n[\{Y - \hat{m}(X)\}1_{\pi(X)=A}/p\{\pi(X) \mid X\}]$, where E_n denotes the empirical expectation operator. We use $\hat{C}(\pi)$ as the basis for our purity measure. For any rectangle r, let $\pi_{r,a,a'}$ denote the rule that assigns treatment a to all subjects in r and treatment a' to all subjects in r^c. For any rectangular region $\mathscr{R}$ and rectangle r, define the purity of the partitioning of $\mathscr{R}$ into $\mathscr{R} \cap r$ and $\mathscr{R} \cap r^c$ as

$$\mathscr{P}(\mathscr{R}, r) = \max_{a,a' \in \mathscr{A}} E_n \left[\frac{\{Y - \hat{m}(X)\}1_{X \in \mathscr{R}} 1_{A=\pi_{r,a,a'}}(X)}{p\{\pi_{r,a,a'}(X) | X\}} \right] \left(E_n \left[\frac{1_{X \in \mathscr{R}} 1_{A=\pi_{r,a,a'}}(X)}{p\{\pi_{r,a,a'}(X) | X\}} \right] \right)^{-1}. \quad (1)$$

The above purity measure estimates the performance of the best decision rule that assigns a single treatment to all subjects in $\mathscr{R} \cap r$ and a second treatment to all subjects in $\mathscr{R} \cap r^c$.

2·4. Purity measures for continuous treatments

In the continuous treatment setting we assume that $\mathscr{A} = (0, 1)$. For $a \in \mathscr{A}$, let $p(a \mid x)$ denote the density of A given $X = x$. We assume that $p(a \mid x)$ is known and that $\varepsilon \leq p(a \mid X)$ holds for all $a \in \mathscr{A}$ with probability 1 for some fixed $\varepsilon > 0$. If this density is not known, it can be estimated, for example using mean-variance models (Carroll & Ruppert, 1988). The estimation of personalized treatment rules with continuous treatments has not been well studied. A major difficulty with regression-based methods is that one must first model $Q(x, a) = E(Y \mid X = x, A = a)$ and then invert it to find $\pi^{\text{opt}}(x) = \arg\sup_{a \in (0,1)} Q(x, a)$. Correctly specifying a functional form for $Q(x, a)$ that is interpretable yet sufficiently expressive and easily inverted within a continuous range of treatments is nontrivial when x is moderate- or high-dimensional. Our tree-based approach uses an estimator of $m(x) = E\{Y \mid X = x, A = \pi^{\text{opt}}(x)\}$ to reduce variance in the purity measure, but this need not be interpretable, nor does it require easy invertibility, so a flexible model for $m(x)$ can be used. Furthermore, correct specification of the model for $m(x)$ is not needed for consistency, because the optimal decision rule π^{opt} is invariant with respect to $g(x)$ in $C_g(\pi)$.

In order to define a purity measure, we require an estimator of the quality of an arbitrary rule $\pi : \mathbb{R}^p \to \mathscr{A}$. The change of measure applied in the discrete treatment case is not meaningful for continuous treatments because $E[Y\, 1_{A=\pi(X)}/p\{\pi(X) \mid X\}] \equiv 0$. Instead, we propose a smoothed version of the discrete purity measure (1) that replaces the nonsmooth indicator function with a kernel smoother $\nu_{\pi,h}(a \mid x) = h^{-1}\kappa([f(a) - f\{\pi(x)\}]/h)\, f'(a)$, where κ is a symmetric density function, h is the kernel bandwidth, and f is a one-to-one function mapping the treatment space (0, 1) to $\mathbb{R}$, with derivative f'. For example, one simple approximation to π is $\nu_{\pi,h}(a \mid x) = (2h)^{-1} f'(a) 1|_{f(a)-f\{\pi(x)\}|\le h}$. Thus, we approximate a rule π with a class of distributions over (0, 1) indexed by $\mathbb{R}^p$, so that $\nu_{\pi,h}(a \mid x)$ has mass around $\pi(x)$. Indeed, $\nu_{\pi,h}(a \mid x)$ defines a distribution over treatments for each value of $x \in \mathbb{R}^p$ and is therefore called a stochastic rule (Sutton &Barto, 1998). Using a stochastic rule to approximate a nonstochastic or deterministic rule effectively smooths over treatments. The precision of the approximation $\nu_{\pi,h}(a \mid x)$ to $\pi(x)$ depends on how peaked the function $f'(a)\kappa([f(a) - f\{\pi(x)\}]/h)/h$ is at $\pi(x)$. However, we will show below that making the function too peaked will lead to unstable results due to inflated variance.

For any fixed function $g : \mathbb{R}^p \to \mathbb{R}$, define $L_g(\nu_{\pi,h}) = L_g(\nu_{\pi,h}, X, A, Y) = \{Y - g(X)\} \times \nu_{\pi,h}(A \mid X)/p(A \mid X)$ and $C_g(\nu_{\pi,h}) = E\{L_g(\nu_{\pi,h})\}$. Then $C_0(\nu_{\pi,h})$ is the importance sampling representation of the expected outcome if all patients are assigned treatment according to the stochastic rule $\nu_{\pi,h}$. The plug-in estimator of $C_g(\nu_{\pi,h})$ is $\hat{C}_g(\nu_{\pi,h}) = E_n[\{Y - g(X)\}\nu_{\pi,h}(A \mid X)/p(A \mid X)]$. The following lemma characterizes how the bias and variance of $\hat{C}_g(\nu_{\pi,h})$ depend on the bandwidth h.

Lemma 2—*Assume that* $p(a \mid X) \ge \varepsilon$ *with probability* 1 *for some* $\varepsilon > 0$ *and all* $a \in \mathscr{A}$, *and that* $\kappa(u)$ *is symmetric about* 0 *and satisfies* $\int_{-\infty}^{\infty}\kappa(u)\mathrm{d}u=1$. *Then, for any* $g : \mathbb{R}^p \to \mathbb{R}$ *and any rule* π,

$$E\{\hat{C}_g(\nu_{\pi,h})\}=\int_{-\infty}^{\infty} u^2\kappa(u)\mathrm{d}u\frac{h^2}{2}E\left(E\left[\frac{\partial^2}{\partial a^2}E\{Y-g(X)|X,A=a\}\right|_{a=\pi(X)}\right)+O(h^4)$$

and $\mathrm{var}\{\hat{C}_g(\nu_{\pi,h})\} = O\{1/(nh)\}$.

The mean squared error of $\hat{C}_g(\nu_{\pi,h})$ is approximately

$$\frac{h^4}{4}\left\{\int_{-\infty}^{\infty} u^2\kappa(u)\mathrm{d}uE\left(E\left[\frac{\partial^2}{\partial a^2}E\{Y-g(X)|X,A=a\}\right|_{a=\pi(X)}\right]\right)\right\}^2 + \frac{1}{nh}E\left(E\left[\frac{\{Y-g(X)\}^2}{p\{\pi(X)|X\}}f'\{\pi(X)\}\middle|X,A=\pi(X)\right]\right)\int_{-\infty}^{\infty}\kappa^2(u)\mathrm{d}u,$$

which is a function of the sample size, the bandwidth and the kernel function. This expression shows that a requirement for the mean squared error to decrease to zero as n increases is that $h \to 0$ and $nh \to \infty$. Under additional assumptions, we derive a plug-in estimator of the optimal bandwidth.

The expectation $E\{\nu_{\pi,h}(A \mid X)/p(A \mid X) \mid X\}$ is unity; hence, for an arbitrary function g, it follows from the same argument as in the discrete case that $\arg\max_\pi C_g(\nu_{\pi,h}) = \arg\max_\pi C_0(\nu_{\pi,h})$. Similarly, for a fixed reference rule $\tilde{\pi}$ and kernel κ, the choice of $g(x) = E\{Y \mid X = x, A = \tilde{\pi}(x)\}$ minimizes the variance of $L_g(\nu_{\pi,h})$ at $\pi = \tilde{\pi}$. Thus, as in the discrete case, we use a crude estimator of π^{opt} as our reference rule. To derive a plug-in bandwidth estimator, we assume that

$$Y = h(x) + a\ell(x) + \frac{a^2}{2}\psi(x) + \varepsilon, \quad (2)$$

where h, ℓ and ψ are arbitrary functions from $\mathbb{R}^p$ into $\mathbb{R}$, $\in \mathbb{R}^p$, and ε is an independent additive error with mean zero and variance σ_ε^2. The form of the working model in (2) is a generalization of that used by Rich et al. (2014) for adaptively modelling warfarin dose response. Here $g(x) = h(x) + \pi^{\text{opt}}(x)\ell(x) + \pi^{\text{opt}}(x)^2\psi(x)/2$ and $E\{Y - g(X) \mid X, A = \pi(X)\}^2 = \varepsilon^2$. Under the assumed model we have $(\partial^2/\partial a^2)E\{Y - g(X) \mid X, A = a\}|_{a=\pi^{\text{opt}}(X)} = \psi(X)$, which is independent of the optimal rule π^{opt}. Furthermore, we choose $f(u) = u - 1/2$ so that $f'(u) \equiv 1$. We assume a uniform treatment randomization so that $p(u \mid x) \equiv 1$. Ignoring higher-order error terms, the bandwidth that minimizes the mean squared error is

$$h* = \frac{\sigma_\varepsilon^{2/5}\int_{-\infty}^{\infty}\kappa^2(u)\mathrm{d}u}{(4n)^{1/5}\left[E\{\psi(X)\int_{-\infty}^{\infty}u^2\kappa(u)\mathrm{d}u\}\right]^2},$$

from which we obtain the plug-in estimator

$$\hat{h} = \frac{\hat{\sigma}_\varepsilon^{2/5}\int_{-\infty}^{\infty}\kappa^2(u)\mathrm{d}u}{(4n)^{1/5}\left[E_n\{\hat{\psi}(X)\int_{-\infty}^{\infty}u^2\kappa(u)\mathrm{d}u\}\right]^2}, \quad (3)$$

where $\hat{\sigma}_\varepsilon$ and $\hat{\psi}$ are obtained by regressing Y on X and A using (2) with working models for $h(x)$, $\ell(x)$ and $\psi(x)$. In our simulations we used $E(Y \mid X, A) = X^{\mathrm{T}}\rho + X^{\mathrm{T}}\beta(a - X^{\mathrm{T}}\gamma)^2$, which corresponds to $h(X) = X^{\mathrm{T}}\rho + (X^{\mathrm{T}}\beta)(X^{\mathrm{T}}\gamma)^2$, $\ell(X) = 2X^{\mathrm{T}}\beta X^{\mathrm{T}}\gamma$ and $\psi(X) = 2X^{\mathrm{T}}\beta$; the parameters indexing this model were estimated using nonlinear least squares with a ridge penalty added for stability.

Write $C(\nu_{\pi,h})$ for $C_{E\{Y \mid X, A=\pi^{\text{opt}}(X)\}}(\nu_{\pi,h})$. With $m(\hat{x})$ denoting an estimator of $E\{Y \mid X = x, A = \pi^{\text{opt}}(X)\}$, the plug-in estimator of $C(\nu_{\pi,h})$ is

$$\hat{C}(\nu_{\pi,h}) = E_n\left[\frac{\{Y - \hat{m}(X)\}\nu_{\pi,h}(A|X)}{p(A|X)}\right];$$

we use this estimator as the basis for our purity measure. In our simulated experiments we take $h = \hat{H}$, which we recommend using in practice, although other choices are possible. Write $\nu_{a'h}(a \mid x)$ as shorthand for $\kappa[\{f(a) - f(a')\}/h]\, f'(a)/h$. For any rectangular region $\mathscr{R}$ and rectangle r, define the purity of partitioning $\mathscr{R}$ into $\mathscr{R} \cap r$ and $\mathscr{R} \cap r^{\mathrm{c}}$ with respect to κ as

$$\mathscr{P}_{\kappa}(\mathscr{R}, r) = \sup_{a,a' \in \mathscr{A}} E_n \left[\frac{\{Y - \hat{m}(X)\} 1_{X \in \mathscr{R} \nu_{\pi_{r,a,a'},h}}(A|X)\}}{p\{\pi_{r,a,a'}(X)|X\}} \right] \left(E_n \left[\frac{1_{X \in \mathscr{R} \nu_{\pi_{r,a,a'},h}}(A|X)}{p\{\pi_{r,a,a'}(X)|X\}} \right] \right)^{-1}.$$

This estimates the performance of the best stochastic decision which is concentrated about a single value for subjects in $\mathscr{R} \cap r$ and around a second value for subjects in $\mathscr{R} \cap r^c$. In practice, the supremum is taken over observed values a in the training data.

2·5. Recursive splitting

Having defined the purity measures, we can now describe how to select the split at each stage in the recursive splitting. Generally, it is preferable to choose the split that leads to the greatest increase in node purity, defined in terms of $\mathscr{P}(\mathscr{R}, r)$ for discrete treatments and $\mathscr{P}_\kappa(\mathscr{R}, r)$ for continuous treatments. We use the discrete case as an illustrative example, and apply the same strategy to the continuous case by simply replacing $\mathscr{P}(\mathscr{R}, r)$ with $\mathscr{P}_\kappa(\mathscr{R}, r)$.

To grow the tree at a parent node associated with rectangular region $\mathscr{R}$, say, we split on the rectangle r that maximizes the total purity of its two child nodes $\mathscr{P}(\mathscr{R}, r)$. We then repeat the splitting process on each of the new nodes. Of course it is not possible to split the tree indefinitely, and so stopping measures based on tree complexity and node size are employed. For any rectangular region $\mathscr{R}$, define $\mathscr{P}(\mathscr{R}) = \mathscr{P}(\mathscr{R}, \emptyset)$. Let $\mu \in \mathbb{N}$ denote a minimum node size, sometimes called a bucket size. We employ the following splitting rules.

Rule 1—If $nE_n 1_{X \in \mathscr{R}} < 2\mu$, do not split.

Rule 2—If $nE_n 1_{X \in \mathscr{R}} \geq 2\mu$, compute $r \hat{=} \arg\max_r \{\mathscr{P}(\mathscr{R}, r) : \min(nE_n 1_{X \in \mathscr{R} \cap r}, nE_n 1_{X \in \mathscr{R} \cap r^c}) \geq \mu\}$. If $\mathscr{P}(\mathscr{R}, r) \hat{\geq} \mathscr{P}(\mathscr{R}) + \lambda$, then split $\mathscr{R}$ into $\mathscr{R} \cap r$ and $\mathscr{R} \cap r^c$; otherwise do not split.

Here $\lambda > 0$ is a small positive constant representing a threshold for practical significance. Typically, μ and λ are dictated by problem-specific considerations and are not treated as tuning parameters. If the current data are representative of the whole subject population, i.e., if the data can be viewed as a random sample from the population from which future patients will be drawn, the splitting strategy outlined above ensures that the treatment recommended in a terminal node is the one maximizing the expected outcomes for subjects in the node. See the Supplementary Material for a more detailed description of the tree-growing algorithm.

2·6. Pruning

Each split in the tree-growing algorithm increases, or at least cannot decrease, the purity measure. Therefore, unless either λ or μ is large, the above splitting strategy will produce a large tree and potentially overfit the data. A standard strategy employed when building decision trees is to first construct a large tree and then prune the tree back by merging sibling nodes together, choosing which nodes to merge by using some global measure of performance; this is generally regarded as a superior strategy to building a smaller tree by

stopping the splitting early (Breiman et al., 1984). We adopt the following simple pruning strategy, which is applied to each pair of siblings until no further merging is possible.

Strategy 1—For siblings $\mathscr{R} \bigcup r$ and $\mathscr{R} \bigcup r^c$ with common parent $\mathscr{R}$, if $\mathscr{P}(\mathscr{R}, r) < \mathscr{P}(\mathscr{R}) + \eta$ then merge; otherwise do not merge.

Here $\eta > 0$ is a tuning parameter which we choose by using a ten-fold crossvalidation estimator of the marginal mean outcome. In particular, let $\hat{\pi}_\eta$ denote the estimated treatment rule using complexity parameter η, and let $\hat{C}^{\mathrm{CV}}(\hat{\pi}_\eta)$ be the crossvalidation estimator of $E^{\hat{\pi}_\eta}(Y)$. Define $\hat{\eta} = \arg\max_\eta \hat{C}^{\mathrm{CV}}(\hat{\pi}_\eta)$. Then the final decision rule is $\hat{\pi}_{\hat{\eta}}$.

3. Experiments

3·1. Preliminaries

In this section we conduct a series of simulation experiments to examine the finite-sample performance of the minimum impurity decision assignments estimator. Here, performance is measured in terms of average marginal mean outcome obtained; that is, for an estimator $\hat{\pi}$ of π^{opt}, we compute $E\{E^{\hat{\pi}}(Y)\}$, where the outer expectation is taken with respect to the data used to estimate $\hat{\pi}$. The average marginal mean outcome obtained was estimated using Monte Carlo methods with a large test set of size 10 000 for the inner expectation and 1000 training sets for the outer expectation. For discrete treatments we use a random forest (Breiman, 2001) to estimate $m(x) = \max_a Q(x, a)$ using the default settings of the R package randomForest; for continuous treatments we set $m(x)$ to be identically zero.

To form a baseline for comparison, we also consider two regression-based estimators (Zhao et al., 2011; Schulte et al., 2014). Regression-based estimators first estimate $Q(x, a) = E(Y \mid X = x, A = a)$ using a regression model, obtaining $\hat{Q}(x, a)$, say, and then estimate the optimal decision rule as $\hat{\pi}(X) = \arg\sup_{a \in \mathscr{A}} \hat{Q}(x, a)$. In the discrete case we consider two estimators of $Q(x, a)$: a parametric estimator that assumes a linear working model of the form $Q_{\mathrm{LM}}(x, a) = x^{\mathrm{T}}\beta + \Sigma_{\mathscr{A}\setminus\{0\}} x^{\mathrm{T}}\psi a$, which we estimate using least squares, and a nonparametric estimator that uses support vector regression with radial basis functions, which we denote by $Q_{\mathrm{SVR}}(x, a)$. The estimator Q_{SVR} uses as features x and all pairwise interactions between x and a; the method is tuned using five-fold crossvalidation with mean squared prediction error as the criterion. Diagnostic plots for the linear model using a draw of the data of size $n = 250$ in the $p = 25$ case are displayed in the Supplementary Material; these plots do not exhibit any major signs for concern, so an analyst might consider a linear decision rule to be adequate. In the continuous treatment case, regression-based estimators were constructed by first discretizing treatment into quartiles and then using the foregoing discrete treatment models.

3·2. Discrete treatments

We consider generative models in which treatments are binary and randomized to take the values±1with equal probability, the covariates X are uniformly distributed on the p-dimensional unit cube $[0, 1]^p$ for $p = 10$, 25 and 50, and $Y = u(X) + Ac(X) + Z$ where Z is an independent standard normal variate and u and c are functions from $[0, 1]^p$ to $\mathbb{R}$. The three

generative models that we consider all use $u(x)=k_p+\tau_p\sum_{j=1}^{p}x_j$, where k_p and τ_p are chosen so that var$\{u(X)\}$ = 5 and $E\{u(X)\} = 10 - E\{|c(X)|\}$. Thus, for all generative models the mean outcome under the optimal regime is $E\{m(X)\} + E\{|c(X)|\} = 10$. The forms of c we consider are:

$$c(x)=\sqrt{5}(2\times 1_{\{x_1\leq 0\cdot 6,x_2\geq 0\cdot 2\}}-1), \quad (4)$$

$$c(x)=\sqrt{5}(2\times 1_{\{x_1\leq 0\cdot 3,x_2\geq 0\cdot 2\}}-1), \quad (5)$$

$$c(x)=\sqrt{5}\{\mathrm{sgn}(x_1+x_2-1)\}. \quad (6)$$

Hence var$\{Ac(X)\}$ = var$\{u(X)\}$ = 5 in all settings. The optimal regime for (4) is to assign treatment $a = 1$ to all patients with $x_1 \leq 0{\cdot}6$ and $x_2 \leq 0{\cdot}2$, and assign treatment $a = -1$ otherwise. Thus, under the optimal treatment rule 48% of patients would receive treatment $a = 1$. Similarly, the optimal regime for (5) is to assign treatment $a = 1$ to all patients with $x_1 \leq 0{\cdot}3$ and $x_2 \geq 0{\cdot}2$, and assign treatment $a = -1$ otherwise. However, in contrast to (4), under the optimal treatment rule for (5) only 24% percent of patients would receive treatment $a = 1$. The optional regime for (6) assigns treatment $a = 1$ to all subjects with $x_1 + x_2 > 1$ and $a = -1$ otherwise. Thus, the optimal treatment regime for (6) assigns treatment $a = 1$ to 50% of the subjects. The optimal regimes for both (4) and (5) are representable as decision trees, whereas that for (6) is not. The performance of minimum impurity decision assignments on (4) and (5) demonstrates the method's ability to correctly identify underlying tree structure when it is actually present, whereas the performance on (6) measures the impact of a simple model misspecification.

The average performance of minimum impurity decision assignments and of the two regression-based methods is summarized in Table 1; we have also included the performance obtained under π^{opt} and a rule that guesses randomly. The reported values are based on 1000 Monte Carlo replications and a training set of size $n = 250$; simulations with larger sample sizes gave qualitatively similar results and are therefore omitted. The minimum impurity decision assignments estimator performs well across all settings, yielding the best performance for models (4) and (5), and it is competitive, despite being misspecified, in model (6). Most striking is that the performance of minimum impurity decision assignments remains stable as the number of noise variables increases; this could be due in part to the automatic variable-selection property of decision trees. In contrast, the performance of the regression-based methods deteriorates as p increases.

3·3. Continuous treatments

In the continuous treatment case, we consider generative models in which treatments are uniformly distributed on (0, 1), the covariates X are uniformly distributed on the p-dimensional unit cube $[0, 1]^p$, and $Y = u(X) + c(X, A) + Z$, where Z is an independent standard normal variate and $u(x)$ and $c(x, a)$ are, respectively, functions from $[0, 1]^p$ and $[0, 1]^p \times (0, 1)$ to $\mathbb{R}$. The three generative models that we consider use the same form for $u(x)$ as

in the discrete case, with constants τ_p and κ_p chosen so that var$\{u(X)\} = 5$ and $E\{u(X)\} = 10 - E\{\sup_a{}^c(X, a)\}$. Let ϕ and Φ denote the density and cumulative distribution of a standard normal random variable, respectively. The three forms of $c(x, a)$ we consider are:

$$c(x,a) \propto 1_{\{x_1 \geq 0{\cdot}7\}}\phi[3\{\Phi^{-1}(a)+\Phi^{-1}(0{\cdot}75)\}]+1_{\{x_1<0{\cdot}7,x_2>0{\cdot}5\}}\phi\{3\Phi^{-1}(a)\}+1_{\{x_1<0{\cdot}7,x_2\leq 0{\cdot}5\}}\phi\{\Phi^{-1}(a)+\Phi^{-1}(0{\cdot}25)\}, \quad (7)$$

$$\begin{aligned} c(x,a) \propto \Bigg\{ &\left(1-\frac{|a-0\cdot 20|}{0\cdot 20}\right)_+ 1_{\{x_1>0{\cdot}5,x_3>0{\cdot}5\}}+\left(1-\frac{|a-0\cdot 40|}{0\cdot 20}\right)_+ 1_{\{x_1>0{\cdot}5,x_3\leq 0{\cdot}5\}} \\ &+\left(1-\frac{|a-0\cdot 60|}{0\cdot 20}\right)_+ 1_{\{x_1\leq 0{\cdot}5,x_2>0{\cdot}25\}}+\left(1-\frac{|a-0\cdot 80|}{0\cdot 20}\right)_+ 1_{\{x_1\leq 0{\cdot}5,x_2\leq 0{\cdot}25\}}\Bigg\}, \end{aligned} \quad (8)$$

$$c(x,a) \propto \frac{1}{1+10(2a-x_1-x_2)^2}, \quad (9)$$

Here $(w)_+ = \max(0, w)$, and in each case the positive proportionality constant is chosen so that var$\{c(X, A)\} = 5$. The optimal regime for (7) treatment $a = 0{\cdot}25$ when $x_1 \geq 0{\cdot}7$, $a = 0{\cdot}5$ when $x_1 < 0{\cdot}7$ and $x_2 > 0{\cdot}5$, and $a = 0{\cdot}75$ otherwise. The optimal regime for (8) assigns treatment $a = 0{\cdot}20$ if $x_1 > 0{\cdot}5$ and $x_3 > 0{\cdot}5$, $a = 0{\cdot}40$ if $x_1 > 0{\cdot}5$ and $x_3 \leq 0{\cdot}5$, $a = 0{\cdot}60$ if $x_1 \leq 0{\cdot}5$ and $x_2 > 0{\cdot}25$, and $a = 0{\cdot}80$ otherwise. Thus, both (7) and (8) have an inherent tree structure. In contrast, the optimal regime for (9) is $\pi^{\text{opt}}(X) = (x_1 + x_2)/2$ and so the tree-based decision rule is misspecified. We used the uniform kernel and the plug-in estimator (3) to choose the bandwidth for the minimum impurity decision assignment.

The average performances of minimum impurity decision assignment and the two regression-based methods are reported in Table 2. Minimum impurity decision assignment has the highest average performance of the methods compared. In addition, the estimator appears somewhat robust with respect to the addition of noise variables, as in the discrete treatment case. To give a sense of the rule estimated by minimum impurity decision assignment, Fig. 2 shows the average learned decision rules for (7) over 150 Monte Carlo replications as a function of the predictors x_1 and x_2 when $p = 25$ and $n = 250$. Corresponding figures for models (4)–(6), (8) and (9) are presented in the Supplementary Material, and show that the minimum impurity decision assignments estimator is roughly unbiased for the true underlying structure. The plug-in bandwidth estimator performed well despite violation of the assumptions used in its derivation. Additional simulations, omitted for brevity, indicated that the bandwidth $\hat{\sigma_\varepsilon}/n^{1/5}$ performed equally well.

4. Nefazodone study

In this section we apply the minimum impurity decision assignments method to data from a randomized trial comparing the drug nefazodone with cognitive behavioural therapy as treatments for chronic depression (Keller et al., 2000). Patients were randomized to receive, with equal probability, nefazodone, cognitive behavioural therapy, or both nefazodone and cognitive behavioural therapy. Cognitive behavioural therapy requires as often as twice-weekly visits to a clinic, and thus imposes significant time and monetary burdens on patients

relative to treatment with nefazodone alone. An important question is whether cognitive behavioural therapy is necessary for all patients in the population of interest, either alone or as an augmentation to nefazodone, or if there is a subgroup of patients for which cognitive behavioural therapy is unnecessary. We perform a complete case analysis.

The data we use in this analysis comprise 215 subjects randomized to nefazodone, 212 randomized to cognitive behavioural therapy, and 220 randomized to both. The primary outcome of the study was a score measured on the Hamilton Rating Scale for Depression, which we use as our response. To match our development, which assumes that higher values are better, we subtract each score on the rating scale from 50. We consider 22 potential covariates for tailoring treatment; these are listed in the Supplementary Material. Figure 3 shows the decision rule estimated by minimum impurity decision assignment. The estimated decision rule assigns nefazodone and cognitive behavioural therapy to patients with a high mood disturbance, high sleep disturbance, or high baseline depression score. So the estimated decision rule recommends intensive treatment, i.e., nefazodone together with cognitive behavioural therapy, to patients presenting with more severe symptoms.

The marginal mean outcome of the learned decision rule, estimated using ten-fold crossvalidation, is 38·8, which turns out to be the marginal mean outcome of assigning all subjects to the more intensive nefazodone and cognitive behavioural therapy. A linear decision rule fit using ridge regression tuned with generalized crossvalidation assigns all subjects to combined nefazodone and cognitive behavioural therapy. Thus, the difference between the learned decision rule using minimum impurity decision assignments and assigning all patients to nefazodone and cognitive behavioural therapy is not significant. Hence, for reasons of cost and patient burden, one should prefer the rule learned by minimum impurity decision assignments, which assigns the drug alone to 18% of patients. Assigning all patients to nefazodone has an estimated marginal mean outcome of only 33·9, suggesting that the minimum impurity decision assignments estimator has effectively identified individuals in the population who are unlikely to benefit from augmenting nefazodone with cognitive behavioural therapy.

5. Discussion

Decision trees are a cornerstone of exploratory analysis and the canonical example of an interpretable predictive model. Trees are particularly suitable for treatment allocation rules because they are easily interpreted and vetted by intervention scientists. Furthermore, unlike other flexible decision rules (e.g., Zhao et al., 2009, 2012; Zhang et al., 2013), they do not require additional computation to determine a recommended treatment for a newly presenting patient; thus, they are easily deployed and disseminated.

An important extension of this work is the development of tree-based treatment rules for multi-stage treatment problems. There is growing interest in evidence-based sequential treatment rules for the treatment of chronic illness. It is increasingly appreciated that nonlinear models are required for sequential decision rules (Laber et al., 2014). One approach is to use flexible models based on machine learning techniques, but for the reasons mentioned above, this may lead to models which are not interpretable or easily

disseminated. We believe that the direct search framework (Zhang et al., 2013) is an avenue by which our work can be extended to the multi-stage setting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Eric Laber acknowledges support from the U.S. National Institutes of Health.

References

Allegra CJ, Jessup JM, Somerfield MR, Hamilton SR, Hammond EH, Hayes DF, McAllister PK, Morton RF, Schilsky RL. American society of clinical oncology provisional clinical opinion: Testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. J. Clin. Oncol. 2009; 27:2091–2096. [PubMed: 19188670]

Breiman L. Random forests. Mach. Learn. 2001; 45:5–32.

Breiman, L.; Friedman, JH.; Olshen, RA.; Stone, CJ. Classification and Regression Trees. Monterey, California: Wadsworth and Brooks; 1984.

Brinkley J, Tsiatis A, Anstrom KJ. A generalized estimator of the attributable benefit of an optimal treatment regime. Biometrics. 2010; 66:512–522. [PubMed: 19508237]

Carroll, RJ.; Ruppert, D. Transformation and Weighting in Regression. New York: Chapman and Hall; 1988.

Craven, MW.; Shavlik, JW. Adv. Neural Info. Proces. Syst. 9 (NIPS 1996). San Francisco: Morgan Kaufmann Publishers; 1996. Extracting tree-structured representations of trained networks; p. 24-30.

Cummings J, Emre M, Aarsland D, Tekin S, Dronamraju N, Lane R. Effects of rivastigmine in Alzheimer's disease patients with and without hallucinations. J. Alzheimer's Dis. 2010; 20:301–311. [PubMed: 20164585]

Foster JC, Taylor JM, Ruberg SJ. Subgroup identification from randomized clinical trial data. Statist. Med. 2011; 30:2867–2880.

Hamburg MA, Collins FS. The path to personalized medicine. New Engl. J. Med. 2010; 363:301–304. [PubMed: 20551152]

Hastie, TJ.; Tibshirani, RJ.; Friedman, J. The Elements of Statistical Learning. 2nd ed.. New York: Springer; 2009.

Hayes DF, Thor AD, Dressler LG, Weaver D, Edgerton S, Cowan D, Broadwater G, Goldstein LJ, Martino S, Ingle JN, et al. HER2 and response to paclitaxel in node-positive breast cancer. New Engl. J. Med. 2007; 357:1496–1506. [PubMed: 17928597]

Keller MB, McCullough JP, Klein DN, Arnow B, Dunner DL, Gelenberg AJ, Markowitz JC, Nemeroff CB, Russell JM, Thase ME, et al. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. New Engl. J. Med. 2000; 342:1462–1470. [PubMed: 10816183]

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

Lipkovich I, Dmitrienko A, Denne J, Enas G. Subgroup identification based on differential effect search: A recursive partitioning method for establishing response to treatment in patient subpopulations. Statist. Med. 2011; 30:2601–2621.

Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nature Rev. Cancer. 2005; 5:845–856. [PubMed: 16239904]

Piquette-Miller M, Grant D. The art and science of personalized medicine. Clin. Pharmacol. Therap. 2007; 81:311–315. [PubMed: 17339856]

Qian M, Murphy SA. Performance guarantees for individualized treatment rules. Ann. Statist. 2011; 39:1180–1210.

R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. ISBN 3-900051-07-0. http://www.R-project.org.

Rich B, Moodie EEM, Stephens DA, Platt RW. Simulating sequential multiple assignment randomized trials to generate optimal personalized warfarin dosing strategies. Clin. Trials. 2014; 11:435–444. [PubMed: 24464036]

Ripley, BD. Pattern Recognition and Neural Networks. Cambridge: Cambridge University Press; 1996.

Rubin D. Bayesian inference for causal effects: The role of randomization. Ann. Statist. 1978; 6:34–58.

Schulte PJ, Tsiatis AA, Laber EB, Davidian M. Q- and A-learning methods for estimating optimal dynamic treatment regimes. Statist. Sci. 2014; 29:640–661.

Su X, Zhou T, Yan X, Fan J, Yang S. Interaction trees with censored survival data. Int. J. Biostatist. 2008; 4:1–26.

Su X, Tsai C-L, Wang H, Nickerson DM, Li B. Subgroup analysis via recursive partitioning. J. Mach. Learn. Res. 2009; 10:141–158.

Sutton, CD. Classification and regression trees, bagging, and boosting. In: Rao, CR.; Wegman, EJ.; Solka, JL., editors. Handbook of Statistics. Vol. 24. Amsterdam: Elsevier; 2005. p. 303-329.

Sutton, RS.; Barto, AG. Reinforcement Learning: An Introduction. Cambridge: Cambridge University Press; 1998.

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

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

Zhang B, Tsiatis AA, Laber EB, Davidian M. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. Biometrika. 2013; 100:681–694.

Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. Statist. Med. 2009; 28:3294–3315.

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 Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. J. Am. Statist. Assoc. 2012; 107:1106–1118.

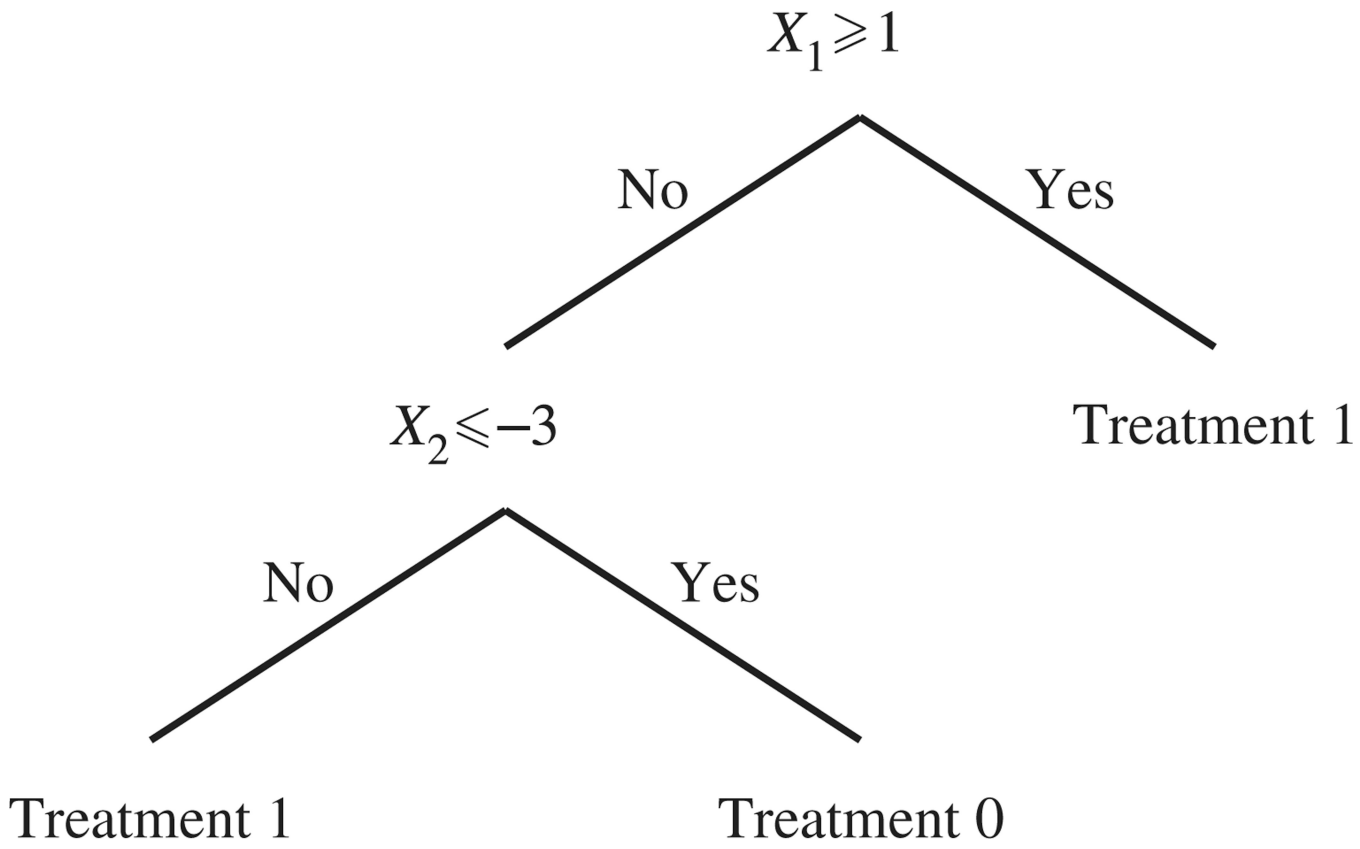

Fig. 1.

A decision rule composed of rectangles $r_1 = \{(1, 1, 1)\}$ and $r_2 = \{(2,-3,-1)\}$ with rectangular regions $\mathscr{R}_0 = r_1^c \cap r_2$ and $\mathscr{R}_1 = (r_1^c \cap r_2) \cup r_1$.

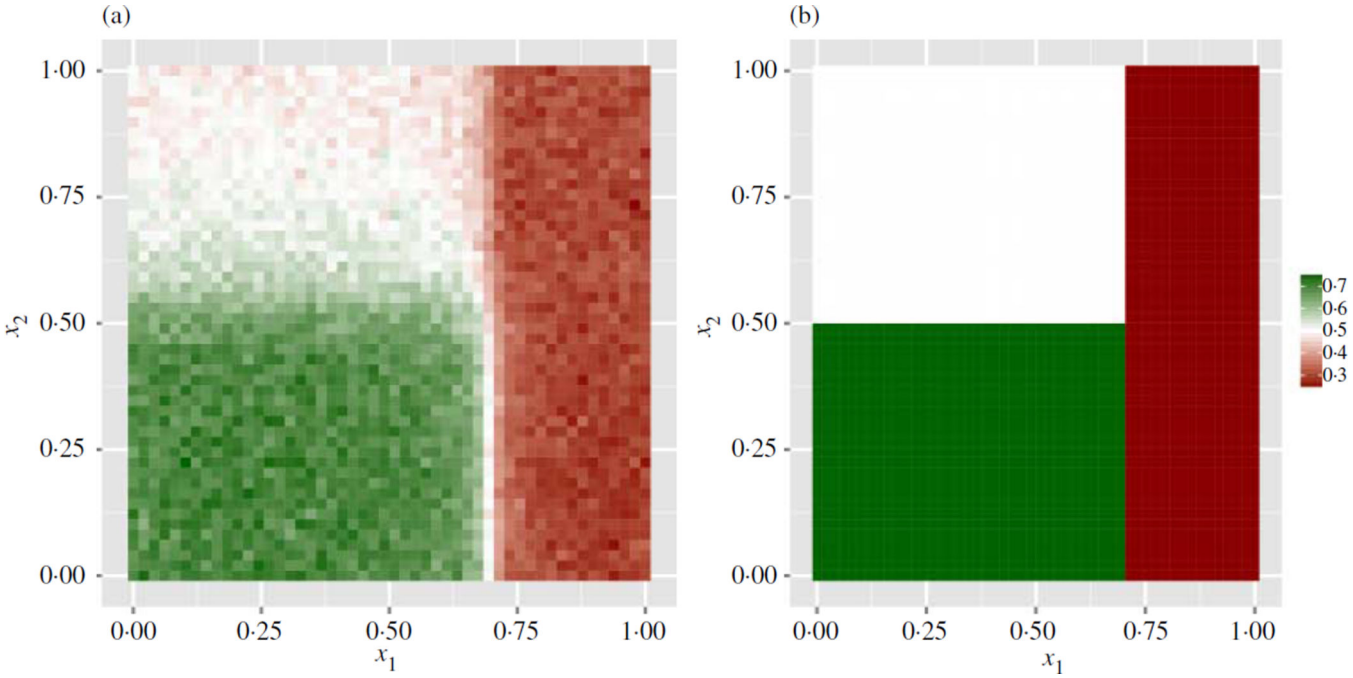

Fig. 2.

Heatmaps of true and estimated optimal treatment rules: (a) average treatment assignment as a function of x_1 and x_2 over 25 learned decision rules for (7), with $p = 25$ and $n = 250$; (b) optimal treatment assignment as a function of x_1 and x_2 for (7); optimal treatment assignment for (7) depends exclusively on x_1 and x_2.

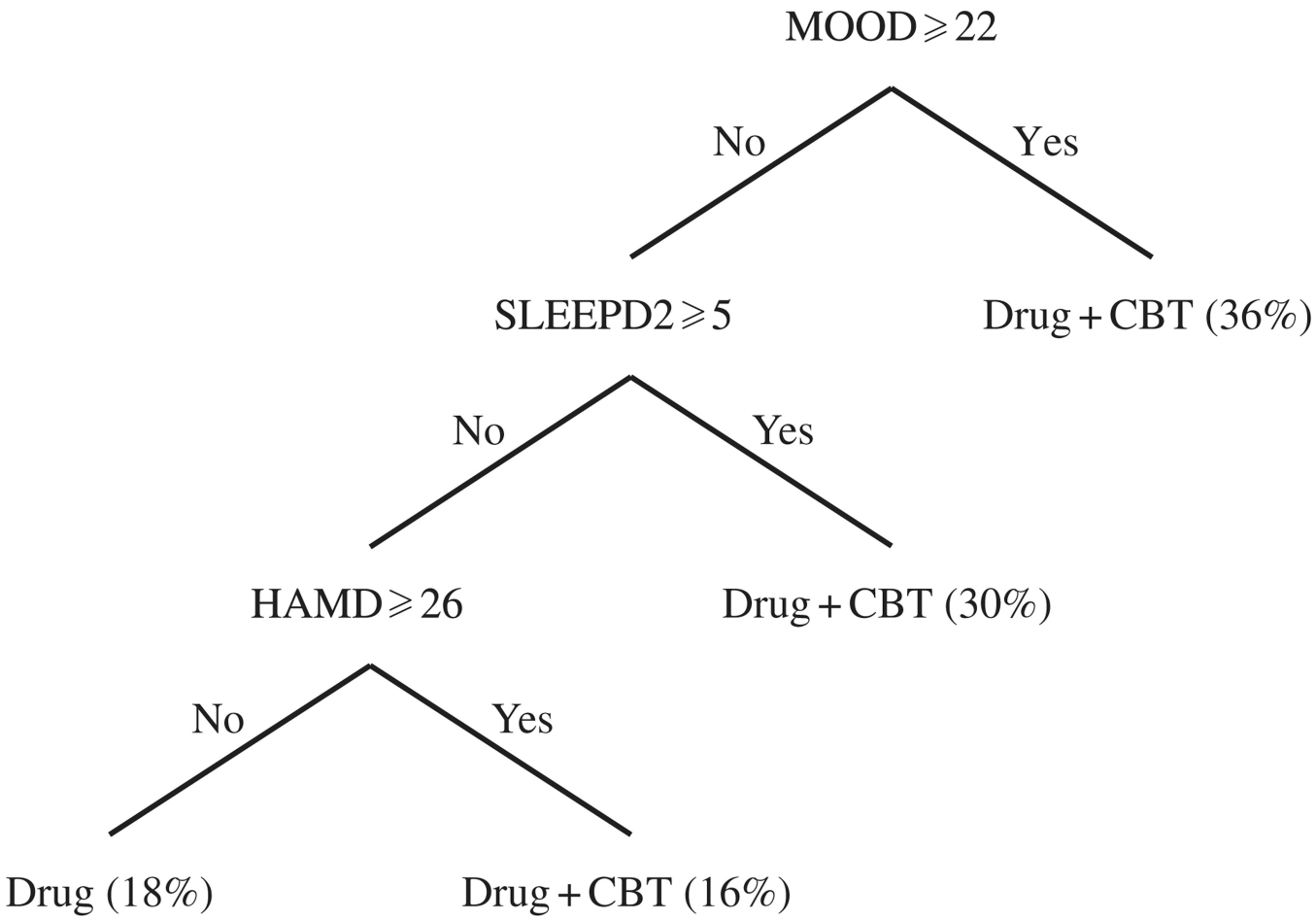

Fig. 3.

Learned decision rule for nafazodone study: patients with high mood disturbance (MOOD), poor sleep (SLEEPD2), or more severe depression symptoms (HAMD) are assigned nefazodone and cognitive behavioural therapy (Drug + CBT); others are assigned nefazodone only.

Table 1

Marginal mean outcomes obtained from minimum impurity decision assignments, two regression-based methods, and random guessing. Data are generated from binary treatment examples with a sample size of n =250; reported values are based on 1000 Monte Carlo replications, using a test set of size 10 000

Model	*p*	MIDAs	QLin	SVR	Random
(4)	10	9·86	9·31	9·39	7·76
(4)	25	9·87	9·20	9·17	7·76
(4)	50	9·88	9·03	8·80	7·76
(5)	10	9·77	9·55	9·53	7·76
(5)	25	9·76	9·44	9·36	7·76
(5)	50	9·77	9·21	8·94	7·76
(6)	10	9·15	9·74	9·60	7·76
(6)	25	9·14	9·51	9·35	7·76
(6)	50	9·15	9·32	8·99	7·76

MIDAs, minimum impurity decision assignments; QLin, Q-learning with a linear model; SVR, Q-learning with support vector regression; Random, random guessing.

Table 2

Marginal mean outcomes obtained from minimum impurity decision assignments, two regression-based methods, and random guessing. Data are generated from continuous treatment examples with a sample size of n =250; reported values are based on 1000 Monte Carlo replications, using a test set of size 10 000

Model	*p*	MIDAs	QLin	SVR	Random
(7)	10	8·81	7·61	7·29	5·94
(7)	25	8·90	7·21	7·21	5·94
(7)	50	8·75	6·62	6·60	5·94
(8)	10	6·70	6·03	5·95	4·61
(8)	20	6·72	5·57	5·61	4·61
(8)	50	6·69	5·04	5·29	4·61
(9)	10	7·44	7·61	7·62	6·12
(9)	20	7·36	7·19	7·19	6·12
(9)	50	7·16	6·70	6·80	6·12