Statistica Sinica **24** (2014), 1461-1485
doi:http://dx.doi.org/10.5705/ss.2012.196

ESTIMATION OF TREATMENT POLICIES BASED ON FUNCTIONAL PREDICTORS

Ian W. McKeague and Min Qian

Columbia University

Abstract: Biosignatures such as brain scans, mass spectrometry, or gene expression profiles might one day be used to guide treatment selection and improve outcomes. This article develops a way of estimating optimal treatment policies based on data from randomized clinical trials by interpreting patient biosignatures as functional predictors. A flexible functional regression model is used to represent the treatment effect and construct the estimated policy. The effectiveness of the estimated policy is assessed by furnishing prediction intervals for the mean outcome when all patients follow the policy. The validity of these prediction intervals is established under mild regularity conditions on the functional regression model. The performance of the proposed approach is evaluated in numerical studies.

Key words and phrases: Empirical processes, functional data analysis, inverse treatment probability weighting, locally efficient estimation.

1. Introduction

Recent advances in biomedical imaging, mass spectrometry, and high-throughput gene expression technology produce massive amounts of data on individual patients. This has the potential to advance the clinical prediction of disease origin, prognosis, and therapeutic response, and opens up the possibility of tailoring treatments to the biosignatures of individual patients. For example, brain imaging studies of how treatments for depression engage various neural mechanisms may one day be used to guide treatment selection and improve outcomes (EMBARC (2013)), and PET studies that compare patients treated with cognitive therapy and patients treated with anti-depressants may be used to predict which treatment is more likely to benefit a given patient (DeRubeis et al. (2008)). In addition, mass spectrometry profiling has successfully detected differences between cancer cases and controls, and may contribute to personalized cancer care (Koomen et al. (2008); Taguchi et al. (2011)). Also, gene expression profiles may be useful for designing individualized therapies for cancer or cardiovascular disease (Heidecker and Hare (2007); van 't Veer and Bernards (2008)). Unfortunately, however, it is very difficult to develop optimal treatment policies based on high-dimensional patient profiles and establish their effectiveness, not only

from the practical point of view of designing clinical trials that can exploit new scientific understanding, but also because statistical methodology has received limited attention in this setting.

In this paper we study treatment policies based on large-scale patient profiles by interpreting them as functional predictors. Given pre-treatment patient profiles, treatment assignments, and outcomes of all the patients in a randomized clinical trial, our aim is to develop a way of estimating and evaluating the effectiveness of the treatment policy that optimizes interactions between the functional predictor and treatment. Our approach to estimating effectiveness is to provide prediction intervals for the mean outcome when all patients follow the estimated policy. These prediction intervals could be used in practice to help clinicians decide whether it is desirable to implement an estimated policy, or to decide between alternative policies.

Early research in the area of individualized treatment concentrated on identifying qualitative interactions between treatment and patient pretreatment clinical variables, with treatments for various subsets of patients selected by hypothesis testing (Byar and Corle (1977); Shuster and Van Eys (1983); Gail and Simon (1985)). In recent years, statistical methods for estimating a patient's risk category based on profile information have been developed (Dettling and Bühlmann (2004); Cai et al. (2010, 2013)), and such estimates can help inform decisions as to the best treatment. This approach is only appropriate in settings where the best treatment (from among competing treatments) in each risk category is known. There is also an abundance of literature focusing on methods for identifying features of high-dimensional patient profiles associated with clinical outcomes (e.g., Gui and Li (2005); Fan and Lv (2008); Engler and Li (2009)). These methods can be used to predict outcome under each competing treatment, so an individualized treatment policy can be formulated by choosing the treatment that achieves the best predicted outcome. Such treatment policies are generally less cost effective, though, since they are likely to involve features that do not interact with treatment.

A more direct approach is to restrict attention to a class of policies indexed by a finite-dimensional parameter (η say), and then optimize a "doubly robust" estimate of the mean outcome (Murphy et al. (2001)) that would be obtained if everyone in the population follows the policy, see Zhang et al. (2012). This approach produces an optimal policy that is robust to model misspecification when the data come from a randomized clinical trial, but it is computationally intractable unless η has very low dimension, because of non-concavity of the objective function. For functional predictors, however, it is desirable to consider more complex policies that are capable of exploiting the wealth of patient information available, so η would need to have high dimension. Furthermore, this

direct approach can result in unstable estimates of the policy, especially in small samples, because the estimated objective function is not a smooth function of η.

We study an alternative approach based on initially fitting a functional regression model for the interaction between the patient profile and treatment, and then optimizing the interaction over competing treatments. This model-based approach is computationally tractable, provides stable estimates of the policy (in the sense of having a low variance compared with the direct approach), and is able to capture complex features of the patient's profile. To assess the effectiveness of the resulting estimated treatment policy, we develop several prediction intervals for the expected outcome under the estimated policy.

The paper is organized as follows. In Section 2 we describe the proposed procedure for estimating treatment policies. Various prediction intervals for evaluating the estimated policy in terms of the mean outcome are provided in Section 3. In Section 4 we discuss several functional regression models that fit into our framework. In Section 5 we study the effectiveness of the estimated policy and the accuracy of the proposed prediction intervals using simulations, and the approach is illustrated using gene expression data from the Cancer Genome Atlas pilot project (TCGA (2008)). Further discussion is presented in Section 6. Proofs of the main results are given in the Appendix.

2. Estimating Optimal Treatment Policies via Functional Regression Models

Suppose we are given pre-treatment profiles, treatment assignments, and outcomes from n patients in a randomized clinical trial. For simplicity we consider only two competing treatments, denoted by $A = \pm 1$, and a scalar outcome Y for which large values are desirable. Let a patient's profile be represented by $X \in \mathcal{X}$, where $\mathcal{X}$ is a measurable space. A treatment policy is a map d from $\mathcal{X}$ into the treatment space $\{-1, 1\}$.

Various methods are available for specifying optimal treatment policies based on low-dimensional profiles (e.g., Song and Pepe (2004); Song and Zhou (2011); Laber et al. (2010)), and also high-dimensional profiles (Qian and Murphy (2011); Lu et al. (2013); Zhao et al. (2012)). However, biosignatures such as brain scans are more naturally treated as functional predictors, $X = \{X(t), t \in \mathcal{T}\}$, where $\mathcal{T}$ is a finite-dimensional index set. In this setting we can expect better results if the treatment policies d adapt to the smooth nature of $X(t)$.

An appropriate measure of the effectiveness of d is the expected outcome, $P^d[Y]$, that would have resulted if d had been used to choose treatment for the entire study population; here P^d denotes the distribution of (X, A, Y) given $A = d(X)$, and $V(d) \triangleq P^d[Y]$ is called the *value* of the policy. We assume throughout that we have i.i.d. data on $(X, A, Y) \sim P$ from a randomized clinical

trial (as when unbiased coin tosses are used to assign subjects to treatment groups), and X is observed without measurement error. Our approach could readily be extended to observational studies in which the treatment assignment probability is unknown, but for simplicity we restrict attention to the case in which it is known.

A treatment policy that maximizes the function $d \mapsto V(d)$ over all possible d is called *optimal*. It is easy to verify that any $d_0(x) \in \arg\max_{a\in\{-1,1\}} E(Y|X = x, A = a)$, $x \in \mathcal{X}$, is optimal, where E denotes expectation under P. Since A is independent of X and has mean zero,

$$E(Y|X, A) = E(Y|X) + T(X)A, \tag{2.1}$$

where

$$T(X) = \frac{E(Y|X, A = 1) - E(Y|X, A = -1)}{2}$$

is the *treatment effect*, so we can write $d_0(x) = \operatorname{sign}(T(x))$ where $\operatorname{sign}(0) \equiv 1$. Then it is natural to estimate d_0 by $\hat{d}_n(x) = \operatorname{sign}(\hat{f}_n(x))$, where $\hat{f}_n$ is an estimator of T based on the data from the randomized clinical trial.

To furnish such an estimate, we need a model for both the treatment effect and the main effect $m(X) = E(Y|X)$ of the pre-treatment profile. As we are only interested in T, there is no loss of generality in assuming that m is known (because of the independence of A and X); see Remark 1 after Theorem 1 for further discussion of misspecification of the model for $m(\cdot)$. We model the treatment effect T by a family of functions of the general form $f_\theta \colon \mathcal{X} \to \mathbb{R}$, with θ ranging over a subset Θ of a finite-dimensional Euclidean space, and we use $\hat{f}_n = f_{\hat{\theta}_n}$, where $\hat{\theta}_n$ is the penalized least squares estimator

$$\hat{\theta}_n \triangleq \arg\min_{\theta\in\Theta} \left\{ \mathbb{P}_n[Y - m(X) - f_\theta(X)A]^2 + p_\lambda(\theta) \right\}. \tag{2.2}$$

Here $\mathbb{P}_n$ is the empirical distribution for a sample of size n from the randomized clinical trial, $p_\lambda(\cdot)$ is a continuous penalty function and $\lambda = \lambda_n$ is the tuning parameter. In Section 4 we discuss examples of the model f_θ. The penalty term is included to avoid over-fitting, but in some of the examples penalization is not necessary.

We will see in the next section that $\hat{d}_n$ is asymptotically optimal in the sense that $V(\hat{d}_n)$ converges in probability to $V(d_0)$ as $n \to \infty$, under suitable regularity conditions on the model $\{f_\theta, \theta \in \Theta\}$ and provided that T is in the model. Moreover, this result holds even if the main effect of X is replaced by an estimator (possibly in an inconsistent way); this robustness property is a consequence of $E(A|X) = 0$ a.s. Note that "asymptotically optimal" in this setting is different from its usage in the theory of efficient estimation.

3. Assessing the Estimated Policy

In this section we study how to assess the effectiveness of the estimated policy $\hat{d}_n$ using an estimator of the form $\hat{V}(\hat{d}_n)$, for various choices of estimators $\hat{V}(d)$ of $V(d)$. The estimators $\hat{V}$ are based on exploiting different aspects of the model for the treatment effect T. In each case, under suitable conditions, we show that the error involved in this assessment, $\hat{V}(\hat{d}_n) - V(\hat{d}_n)$, is asymptotically normal with mean zero and a variance that can be consistently estimated. This leads to asymptotically valid prediction intervals for $V(\hat{d}_n)$ in the sense that the probability of $V(\hat{d}_n)$ falling in the interval tends to the nominal coverage level. These intervals can then provide an attractive way of assessing the potential clinical effectiveness of $\hat{d}_n$ in the study population.

Our first result shows that the value of the estimated policy converges in probability to the value of the policy $d_0'(x) = \text{sign}(f_{\theta_0}(x))$ corresponding to the best fit f_{θ_0} for the treatment effect (in terms of mean squared error). In particular, if the model f_θ for the treatment effect is correctly specified, then $V(\hat{d}_n)$ converges in probability to the value of the optimal policy, since $d_0' = d_0$ in this case; if the model f_θ is misspecified, then d_0' may not be the best policy within the class of policies defined by the model. This issue also arises when a smooth surrogate of the empirical value function is maximized (Zhao et al. (2012)).

We need some mild regularity conditions on the model $f_\theta(X)$ and the noise component $\epsilon = Y - E(Y|X, A)$; they are Assumptions (A1)−(A3) in the Appendix.

Theorem 1. *Suppose* (A1)−(A3) *hold and the penalty* $p_{\lambda_n}(\cdot) \to 0$ *uniformly over compact subsets of* Θ *as* $n \to \infty$. *If* $\theta_0 = \arg\min_{\theta\in\Theta} E\big[T(X) - f_\theta(X)\big]^2$ *is unique and either* $f_{\theta_0}(X) \neq 0$ *or* $T(X) = 0$ *a.s., then* $V(\hat{d}_n) \xrightarrow{p} V(d_0')$.

Remark 1. For simplicity, we have assumed that the main effect $m(X)$ is known. Nevertheless, the above result still holds when $m(\cdot)$ is unknown and is modeled appropriately, for example by a parametric model m_γ that is linear in a Euclidean parameter $\gamma \in \Gamma$, and having a positive-definite Gram matrix. Using similar arguments to the proof of Theorem 1, it can be shown that penalized least squares consistently estimates

$$(\gamma_0, \theta_0) = \arg \min_{\gamma\in\Gamma, \theta\in\Theta} E[Y - m_\gamma(X) - f_\theta(X)A]^2$$

under its conditions. Moreover, if $T(X) = f_{\theta_0}(X)$, then $V(\hat{d}_n)$ converges to the value $V(d_0)$ of the optimal policy. This holds even if m_γ is misspecified, provided A and X are independent, since then θ_0 does not depend on m_γ:

$$(\gamma_0, \theta_0) = \left(\arg\min_{\gamma\in\Gamma} E[m(X) - m_\gamma(X)]^2, \arg\min_{\theta\in\Theta} E[T(X) - f_\theta(X)]^2 \right).$$

Remark 2. The uniqueness condition on θ_0 can be relaxed; all we need is that f_{θ_0} is unique. The uniqueness condition makes the proof easier to present because it can be shown that $\hat{\theta}_n \xrightarrow{p} \theta_0$ (see Lemma A.1 in the Appendix) provided θ_0 is a well-separated minimizer, using general M-estimation theory, see van der Vaart (1998, Theorem 5.7).

Remark 3. Note that $f_{\theta_0}(x) \neq 0$ or $T(x) = 0$ means $d_0'(x)$ is "decisive" or the optimal policy $d_0(x)$ is "indecisive." The method of proof of the theorem is based on the convergence of $\hat{f}_n$ to f_{θ_0}, using a Lipschitz condition (A2) on $f_\theta(x)$ as a function of θ. The estimated policy, however, is the sign of $\hat{f}_n(x)$ and the Continuous Mapping Theorem is not applicable if $f_{\theta_0}(x) = 0$. The condition $f_{\theta_0}(x) \neq 0$ guarantees that $\hat{d}_n(x)$ will eventually agree with $d_0'(x)$, so the expected outcomes at $X = x$ will agree. On the other hand, the condition $T(x) = 0$ implies that the expected outcome at $X = x$ does not depend on the treatment assignment, so again the expected outcomes of the two policies agree.

We list several estimators of $V(d)$ for a given policy d, noting that $V(d)$ is identifiable in terms of the randomization probability $p(a|x)$ (earlier assumed to be $1/2$ for simplicity) on the basis of the identity

$$V(d) = P^d[Y] = \int Y dP^d = \int Y \frac{dP^d}{dP} dP = E\Big[W(X, A; d)Y\Big], \tag{3.1}$$

where the weight $W(X, A; d) = 1_{A=d(X)}/p(A|X)$ is a version of the Radon–Nikodym derivative dP^d/dP (since P^d is absolutely continuous with respect to P under the assumption that $p(a|X) > 0$ almost surely for each value of a). Although $EW(X, A; d) = 1$, in an empirical version of (3.1), it is preferable to normalize the observed weights by their sample mean. This leads to the inverse probability of treatment weighted (IPTW) estimator (Robins (2000))

$$\hat{V}_I(d) = \frac{\mathbb{P}_n[W(X, A; d)Y]}{\mathbb{P}_n[W(X, A; d)]}.$$

In our randomized trial setting where $p(a|x)$ is known, the IPTW estimator is consistent by the Law of Large Numbers. Alternatively, the randomization probability can be eliminated by re-expressing (3.1) as

$$V(d) = E\big[1_{d(X)=1}E(Y|X, A = 1) + 1_{d(X)=-1}E(Y|X, A = -1)\big]. \tag{3.2}$$

This can be estimated by replacing $E(Y|X, A)$ in the above expression by $\hat{E}(Y|X, A) = m(X) + \hat{f}_n(X)A$, resulting in the G-computation estimator (Robins (1986))

$$\hat{V}_G(d) = \mathbb{P}_n\big[m(X) + (21_{d(X)=1} - 1)\hat{f}_n(X)\big].$$

A correct model f_θ is needed for this estimator to be consistent, along with more complex conditions. Instead, we consider a locally efficient estimator derived by Murphy et al. (2001) utilizing the randomization probability and plugging in $\hat{E}(Y|X,A)$:

$$\hat{V}_L(d) = \hat{V}_G(d) + \mathbb{P}_n[W(X,A;d)\hat{R}_n],$$

where $\hat{R}_n = Y - \hat{E}(Y|X,A)$ is the residual. This estimator is derived using semiparametric efficiency theory (see, e.g., Tsiatis (2006)) by projecting the score function $W(X,A;d)(Y-V(d))$ of the IPTW estimator onto the orthogonal complement of the tangent space for the nuisance parameter (the treatment assignment probability). It is consistent (given knowledge of $p(a|x)$) even if f_θ is misspecified, and it is locally efficient under the model $E(Y|X,A) = m(X) + f_\theta(X)A$.

As $\hat{V}_I$ and $\hat{V}_L$ are (pointwise) consistent estimates of V, we study the performance of $\hat{V}_I(\hat{d}_n)$ and $\hat{V}_L(\hat{d}_n)$ as estimates of $V(\hat{d}_n)$ in the setting of Theorem 1.

Theorem 2. *Suppose* (A1)–(A4) *hold and* $p_{\lambda_n}(\cdot) \to 0$ *uniformly over compact subsets of* Θ *as* $n \to \infty$. *If* $\theta_0 = \arg\min_{\theta\in\Theta} E\big[T(X) - f_\theta(X)\big]^2$ *is unique and* $P(f_{\theta_0}(X) = 0) = 0$, *then*

(a) $\sqrt{n}\big[\hat{V}_I(\hat{d}_n) - V(\hat{d}_n)\big] \xrightarrow{d} N(0,\sigma_I^2)$, *where* $\sigma_I^2 = \mathrm{Var}\,[W(X,A;d_0')(Y - V(d_0'))]$ *is consistently estimated by* $\hat{\sigma}_I^2 = \mathbb{P}_n[W(X,A;\hat{d}_n)(Y - \hat{V}_I(\hat{d}_n))]^2$.

(b) $\sqrt{n}\big[\hat{V}_L(\hat{d}_n) - V(\hat{d}_n)\big] \xrightarrow{d} N(0,\sigma_L^2)$, *where* $\sigma_L^2 = \mathrm{Var}\big[W(X,A;d_0')R_0 + |f_{\theta_0}(X)| + m(X)\big]$, *with* $R_0 = Y - m(X) - f_{\theta_0}(X)A$, *is consistently estimated by its sample variance* $\hat{\sigma}_L^2$ *in which* θ_0 *is replaced by* $\hat{\theta}_n$.

Remark 4. To estimate σ_L^2 when $m(X)$ is unknown, $m(X)$ can be replaced by $m_{\hat{\gamma}_n}(X)$ in $\hat{\sigma}_L^2$, where $\hat{\gamma}_n$ is the consistent estimator of γ_0 defined after Theorem 1.

Remark 5. It is natural to ask whether our results can be extended to the setting of observational studies in which the treatment assignment distribution, $p(a|x)$, is unknown. Implementing $\hat{V}_I$ and $\hat{V}_L$ requires estimation of $p(a|x)$, which can be done by fitting a parametric model, as suggested in Murphy et al. (2001). Both $\hat{V}_I$ and $\hat{V}_L$ are consistent if $p(a|x)$ can be consistently estimated; $\hat{V}_L$ has the double robustness property (Murphy et al. (2001)) in the sense that it is also consistent if model $E(Y|X,A) = m(X) + f_\theta(X)A$ is correct and its parameters can be consistently estimated. The asymptotic variances of the assessment errors $\hat{V}(\hat{d}_n) - V(\hat{d}_n)$ based on $\hat{V}_I$ and $\hat{V}_L$ are inflated due to the estimation of $p(a|x)$.

Prediction intervals.

Our asymptotic results lead to the $100(1-\alpha)\%$ prediction interval $\hat{I}_n$ for $V(\hat{d}_n)$,

$$\hat{V}(\hat{d}_n) \pm z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$-quantile of the standard normal distribution, $\hat{V}$ and $\hat{\sigma}^2$, the IPTW or the locally efficient estimator and its associated consistent variance estimator, respectively. The interval $\hat{I}_n$ satisfies $\mathrm{P}(V(\hat{d}_n) \in \hat{I}_n) \approx 1-\alpha$, where P is the probability measure for randomness in the sample.

If we are interested in comparing two different treatment policies, $\hat{d}_n$ and $\hat{d}_n^c$, obtained from fitting separate models for $T(X)$, for example to examine whether it is worthwhile to individualize treatment, we would compare $\hat{d}_n$ with the non-individualized policy $\hat{d}_n^c(x) = \mathrm{sign}(\hat{c}_n)$, where $\hat{c}_n$ is obtained by fitting the model $E(Y|X,A) = m(X) + cA$ with c a real-valued parameter. This can be done by constructing a prediction interval for $V(\hat{d}_n) - V(\hat{d}_n^c)$ based on a routine extension of Theorem 2 to the joint asymptotic distribution of the error terms $\hat{V}_n(\hat{d}_n) - V(\hat{d}_n)$ and $\hat{V}_n(\hat{d}_n^c) - V(\hat{d}_n^c)$, along with a consistent estimator of their asymptotic covariance matrix. The result is given in Theorem A.1 in the Appendix.

4. Examples

There is a rich literature on functional regression models involving a predictor $X = \{X(t), t \in \mathcal{T}\}$ indexed by a Euclidean parameter t. One especially flexible formulation is the FAME model of James and Silverman (2005):

$$f_\theta(X) = \alpha + \sum_{k=1}^{r} g_k\left(\int_{\mathcal{T}} X(t)\eta_k(t)\,dt\right), \tag{4.1}$$

where the functions g_k and η_k are represented by a finite linear combination of smooth basis functions. Here θ contains α along with all the coefficients used to represent g_k and η_k. James and Silverman (2005) showed that penalized least squares consistently estimates θ_0 assuming that the penalty term converges to zero uniformly over compact sets. Our results apply to this model provided the basis functions for g_k are Lipschitz. Although the model is flexible, the resulting policy is hard to interpret, and costly to implement because observation of $X(t)$ at all $t \in \mathcal{T}$ is needed.

In some applications it is desirable to restrict attention to treatment policies that are easy to interpret and inexpensive to implement, for example basing the policy on a single component of X. A simple example has the threshold type policies $d_{\tau,c}(x) = 1$ if $X(\tau) > c$, and -1 otherwise, where τ represents the

component of X to be used in the policy, and c is the threshold. These policies amount to using the three-parameter marginal model

$$f_\theta(X) = \alpha + \beta X(\tau), \tag{4.2}$$

where $\theta = (\alpha, \beta, \tau) \in \mathbb{R}^2 \times \mathcal{T}$, and the threshold is given by $c = -\alpha/\beta$ provided $\beta \neq 0$. Here τ could represent a location in time or space, for example a time point in a chromatograph, a genetic locus in a gene expression profile, or a voxel in an fMRI image. Our results hold under conditions placed on the functional predictor $X = \{X(t), t \in \mathcal{T}\}$, i.e., (A2′), (A3′) and (A4′) in the Appendix.

A practical difficulty with threshold policies derived from the marginal model (4.2) is that their performance is sensitive to small registration (or time-warping) errors in X, especially when the sample paths of X are rough or contaminated with measurement error. In that case it is preferable to use a smoother form of treatment selection, say depending on X in a localized *region*, rather than a point location τ, in which case both the location and the size (area) of the region are parameters in the model. One possibility is to use an "area impact" model based on a kernel function K:

$$f_\theta(X) = \alpha + \beta \int_{\mathcal{T}} X(t) K_{\tau,b}(t)\, dt, \tag{4.3}$$

where $K_{\tau,b}(t) = K((t-\tau)/b)/b$, the bandwidth b restricted to be bounded away from zero and infinity, and the kernel K to vanish at infinity. Our results continue to apply. As the bandwidth tends to zero, the area impact model reduces to the point impact model (4.2). One normally observes X at a set $\mathcal{T}_0$ of discrete points in $\mathcal{T}$, $\mathcal{T}_0$ considered to be dense in $\mathcal{T}$, The integrals in the above examples are then well-approximated by their corresponding Riemann sums.

5. Numerical Studies

We evaluated the numerical performance of our proposed method and compared our approach with the Lasso, and with the fused Lasso that takes into account the time ordering. In contrast to the Lasso, our functional data approach takes advantage of the smooth nature of $X(t)$, rather than focusing on isolated components. Yet our approach is flexible enough to allow treatment policies based on localized features of X, as with the area impact model (4.3) or the point impact model (4.2).

5.1. Simulations

Let the functional predictor $X = \{X(t), t \in [0,1]\}$ be Gaussian with mean zero and covariance kernel $\mathrm{Cov}(X(s), X(t)) = 2^{-|100(s-t)|}$, $s, t \in [0,1]$. We took

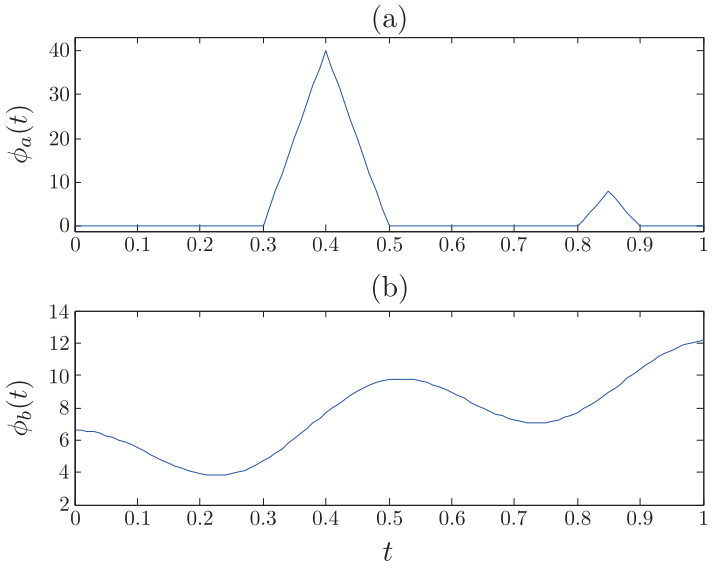

Figure 1. The regression functions in Scenarios (a) and (b).

n i.i.d. trajectories of X observed over a uniform grid $\mathcal{T}_0$ of p "time" points in $\mathcal{T} = [0, 1]$, and considered $p = 100$ and 400. The treatment assignment was $A = \pm 1$ with probability 1/2 each, independently of X. The conditional distribution of the outcome Y given (X, A) was normal with mean

$$E(Y|X, A) = 1 + 15 \int_0^1 tX(t)dt + T(X)A$$

and unit variance.

Scenario (a) $T(X) = 0.61 + \int_0^1 \phi_a(t)X(t)\,dt$, where $\phi_a(t)$ is the piecewise linear function in Figure 1 (a).

Scenario (b) $T(X) = 0.55 + \int_0^1 \phi_b(t)X(t)\,dt$, where $\phi_b(t) = 2[\log(150t^2+10) + \cos(4\pi t)]$, see Figure 1 (b).

In Scenario (a), there is a region of high impact around $t = 0.4$, and region of small impact around $t = 0.85$. Scenario (b) is adapted from a numerical example studied in Cardot et al. (2003), and represents the impact of the predictor as a trend combined with a periodic signal. The coefficients in each scenario were chosen to produce a *medium* effect size in terms of Cohen's d index (Cohen (1988)), the standardized difference in mean responses between the two treatment groups:

$$\mathrm{d} = \frac{E(Y|A = 1) - E(Y|A = -1)}{([\mathrm{Var}(Y|A = 1) + \mathrm{Var}(Y|A = -1)]/2)^{1/2}},$$

with the effect size considered to be "small" if d = 0.2, "medium" if d = 0.5 and "large" if d = 0.8.

We considered four ways of estimating $T(X)$ based on a model for f_θ. The first two exploited the functional nature of X, the other two based on treating X as a high-dimensional vector (with p components): the Lasso (Tibshirani (1996)), and the fused Lasso (Tibshirani et al. (2005)) that extends the Lasso by taking into account the ordering of the predictors.

B-spline: $f_\theta(X) = \mathbf{U}^T\theta$, where $\theta \in \mathbb{R}^{14}$, $\mathbf{U} = (U_j)$, $U_1 = 1$, $U_j = \int_0^1 \psi_j(t)X(t)\,dt$ for $j = 2, \ldots, 14$, with ψ_j the cubic B-spline basis functions having nine interior knots equally spaced from 0.1 to 0.9.

Area impact: $f_\theta(X) = \alpha + \beta \int_0^1 K_{\tau,b}(t)X(t)dt$, where $\theta = (\alpha, \beta, \tau, b)$ and $K(t) = 0.75\big(1 - t^2\big)1_{[-1,1]}$ is the Epanechnikov quadratic kernel.

The B-spline model is a special case of the FAME model in (4.1) with $r = 1$, $g_1(x) = x$, and η_1 represented using the ψ_j as basis functions. The estimated policy $\hat{d}_n$ is based on the estimator $\hat{\theta}_n$ of Remark 1, with $m_\gamma(X) = \mathbf{U}^T\gamma$, $\mathbf{U}$ as before. We do not include a penalty term when fitting these models because the dimension of θ is not large, and we wish to evaluate the potential effect of over-fitting on the value of the resulting policy.

Lasso: $f_\theta(X) = \mathbf{X}^T\theta$, $m_\gamma(X) = \mathbf{X}^T\gamma$, where $\mathbf{X} = (1, X^T)^T$ and $\theta, \gamma \in \mathbb{R}^{p+1}$, with 10-fold cross validation to select the tuning parameter.

Fused Lasso: Same as the Lasso except there are two tuning parameters, that we equal to make the cross validation computationally tractable.

Our first set of simulation results provide the value of the optimal policy d_0 along with summary statistics (mean ± SD) for the values of the estimated policies $\hat{d}_n$ based on the B-spline, area impact, and Lasso methods, see Figure 2. The sample size ranges from $n = 50$ to 800, in increments of 50, and 1,000 replicated samples were used for each given n; the plots on the right are based on the finer grid with $p = 400$. The "true" values were computed from simulated data sets of size $n = 10^4$.

The results under Scenario (a) (top row) show that the area impact method (dotted line) produced policies with the highest value, as expected, because the bulk of the treatment effect resembles an area-impact around $t = 0.4$. Under Scenario (b) (bottom row), the B-spline method (solid line) was preferable for sample sizes above 100, as expected due to its flexibility; when the sample size is small, the area impact method was better, being more parsimonious, though it falls well short of being optimal due to model misspecification. The Lasso method (dash-dot line) was uncompetitive in both scenarios unless the sample size was extremely large, and its performance deteriorated as the resolution of the

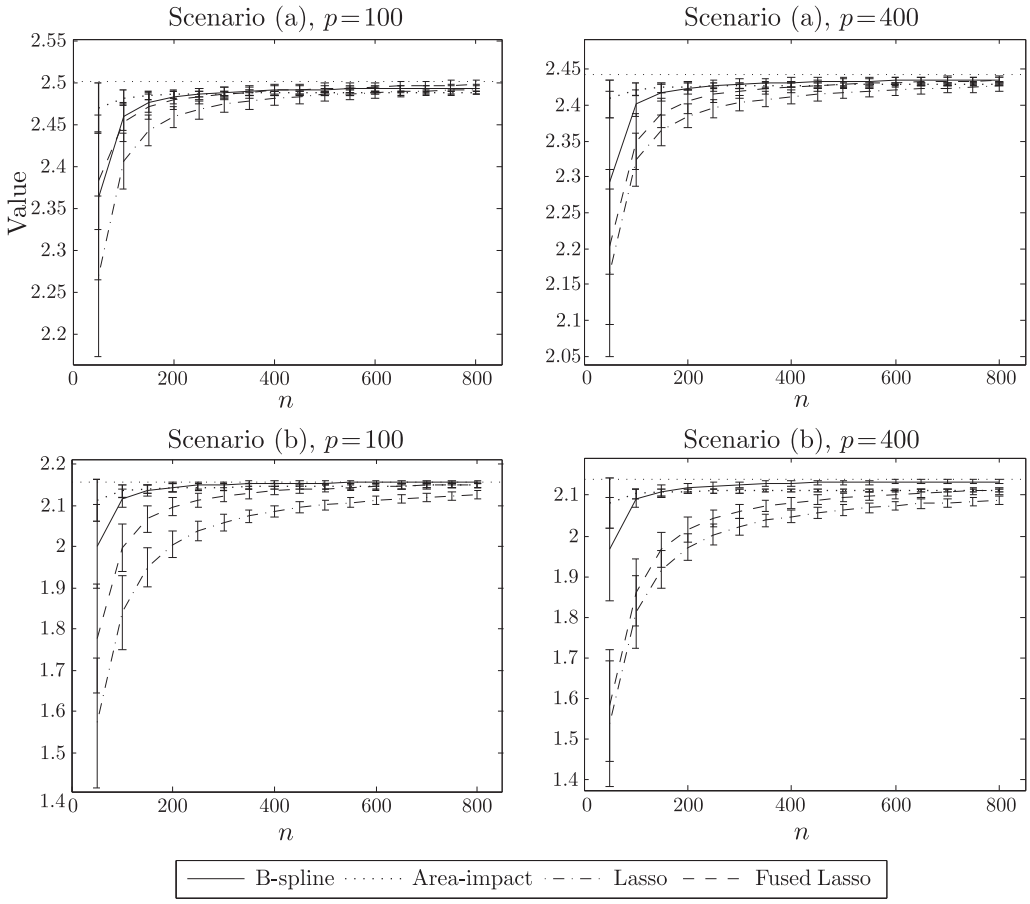

Figure 2. Mean ± SD of values of policies derived from the four methods based on 1,000 replicated samples with n ranging from 50 to 800 in increments of 50. The horizontal dotted line represents the value of the optimal policy.

grid increased. The fused Lasso method (dashed line) was competitive with the B-spline method under Scenario (a) based on the coarse grid, but otherwise not. However, the fused Lasso method always did better than the Lasso, as expected, because it takes into account the ordering of features.

Our second set of simulation results assess the accuracy and the width of the proposed prediction intervals for $V(\hat{d}_n)$ under each method, in the case of Scenario (b) and for the fine grid, see Figure 3. The results for Scenario (a) and the coarse grid were similar, and are included in the supplementary material. The prediction intervals based on the IPTW estimator were always wider than those based on the locally efficient estimator, yet the accuracy was no better for IPTW unless the sample size was small. Undercoverage of the prediction

interval based on the locally efficient estimator can be explained as follows. In small samples ($n \leq 100$), the estimated variance of $\hat{V}_L(\hat{d}_n)$ is downwardly biased under the B-spline method because of overfitting: the contribution of the residual R_0 (in Theorem 2 (b)) is under-estimated. This problem is almost absent with the area impact method, because it is relatively parsimonious. Both the locally efficient and IPTW estimators performed poorly under the Lasso and fused Lasso methods, especially as the resolution of the grid increased, although there was naturally some improvement with increasing sample size. We also examined the performance of each method for finding prediction intervals for the difference in the values of the individualized and non-individualized policies, under Scenario (b), and the results (not shown) were similar to those presented in Figure 3.

5.2. Gene expression example

We considered gene expression data from the tumors of 156 patients diagnosed with a common type of adult brain cancer (glioblastoma), collected in the Cancer Genome Atlas pilot project (TCGA (2008)). Our analysis was based on log gene expression profiles $X = \{X(t), t \in \mathcal{T}\}$, with 181 uniform grid points on $\mathcal{T} = [0, 1]$ representing the 181 loci along chromosome 1; Figure 4 shows the profile for one of the patients. We (artificially) assigned treatment $A = -1$ to the first 78 patients, and $A = 1$ to the remaining patients. The primary outcome Y was taken to be the observed log-survival time, shifted by $A[0.34 + \int_0^1 \phi(t)X(t)\,dt]$, where $\phi(t) = \phi_a(t)/8$ and $\phi_a(t)$ is the regression function in Scenario (a) in the simulation study. The coefficients were chosen to provide a "medium" effect size (based on the sample Cohen's d).

We viewed the constructed triples (X, A, Y) as data from a randomized clinical trial. A simple two-sample t-test (ignoring X) found that treatment $A = 1$ results in significantly higher mean outcome Y than treatment $A = -1$ (P-value $= 0.002$). To estimate the optimal treatment policy, we considered the same four methods as in the simulation study except, to address the overfitting problem with the B-spline method, we used *penalized* B-splines (James and Silverman (2005)) along with 10-fold cross validation to select the tuning parameter.

The estimates of the regression function $\phi(t)$ from the penalized B-spline and area impact methods are given in Figure 5. Comparing with Figure 1 (a), the area impact method captures the region of high impact around $t = 0.4$, and the penalized B-spline method gives considerable weight to the same region, although in a more dispersed fashion. If the estimated treatment policy based on the penalized B-spline or the area impact method were applied to patients in the trial, treatment 1 would be selected for 129 and 122 patients, respectively. The 95% prediction intervals for the value of policies estimated using the penalized B-spline and area impact methods are given in Table 1. The intervals based

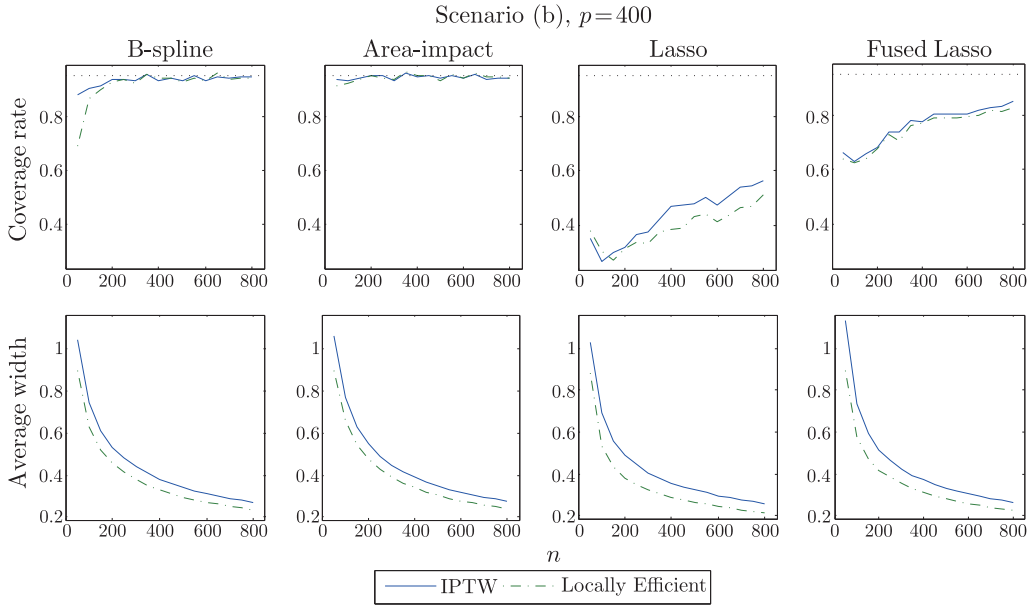

Figure 3. Coverage rate and average width of 95% prediction intervals for the values of estimated policies under Scenario (b), fine grid ($p = 400$).

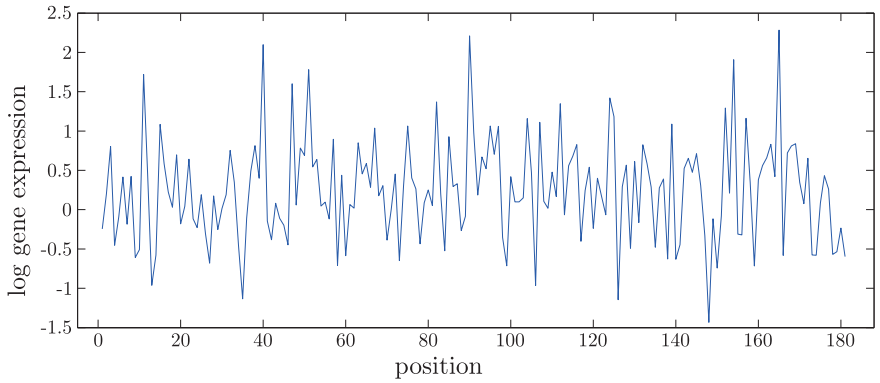

Figure 4. Log gene expression at 181 loci along chromosome 1 in tissue from a brain cancer patient.

on the locally efficient estimator here are slightly narrower than those based on the IPTW estimator, as we found in the simulation study. The treatment policy estimated via the Lasso and the fused Lasso involved gene expression information from 12 and 47 loci, respectively. If the estimated treatment policy based on the Lasso and fused Lasso were applied to patients in the trial, treatment 1 would be selected for 110 and 107 patients, respectively. The four competing methods produced markedly different treatment policies in this example. We did not

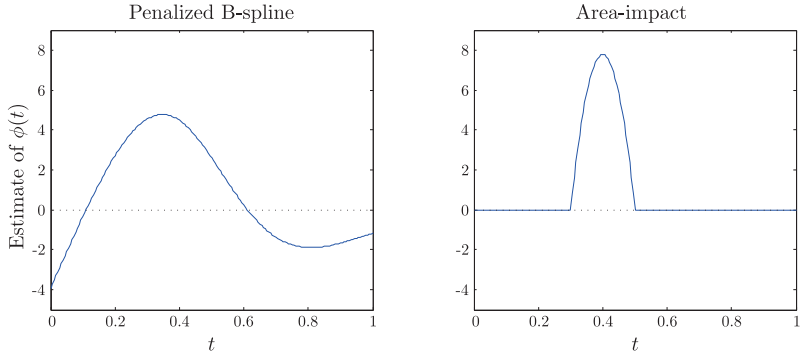

Figure 5. Estimates of the regression function $\phi(t)$ based on the penalized B-spline method (left) and the area impact method (right).

Table 1. 95% prediction intervals for values of policies derived from the penalized B-spline and area impact methods.

	Penalized B-spline	Area-impact
IPTW	[2.67, 3.15]	[2.79, 3.18]
Locally efficient	[2.72, 3.17]	[2.84, 3.21]

Table 2. 95% prediction intervals for pairwise comparisons between the values of policies derived from the penalized B-spline, area impact and non-individualized methods.

	$V(\hat{d}_n^{\mathrm{PBS}}) - V(\hat{d}_n^{\mathrm{AI}})$	$V(\hat{d}_n^{\mathrm{PBS}}) - V(\hat{d}_n^{c})$	$V(\hat{d}_n^{\mathrm{AI}}) - V(\hat{d}_n^{c})$
IPTW	[−0.22, 0.07]	[0.04, 0.36]	[0.07, 0.48]
Locally efficient	[−0.22, 0.04]	[0.02, 0.34]	[0.07, 0.47]

compute the prediction intervals for the Lasso and fused Lasso methods since they are not reliable in such a small sample.

To compare treatment policies derived from the penalized B-spline method (denoted $\hat{d}_n^{\mathrm{PBS}}$) and the area impact method (denoted $\hat{d}_n^{\mathrm{AI}}$), as well as to test whether it is worthwhile to individualize treatment, we constructed 95% prediction intervals for $V(\hat{d}_n^{\mathrm{PBS}}) - V(\hat{d}_n^{\mathrm{AI}})$, $V(\hat{d}_n^{\mathrm{PBS}}) - V(\hat{d}_n^{c})$ and $V(\hat{d}_n^{\mathrm{AI}}) - V(\hat{d}_n^{c})$ based on Theorem A.1, where $\hat{d}_n^{c}$ is the non-individualized policy constructed as discussed at the end of Section 3. In this example $\hat{d}_n^{c}$ selects treatment 1 for all patients. We see from the results in Table 2 that there was no significant difference between the two individualized treatment policies derived from the penalized B-spline and the area impact methods, yet they both produced significantly better expected outcome than the non-individualized treatment policy.

6. Discussion

We have developed a way of estimating optimal treatment policies based

on regression models having functional predictors. Randomized trials involving functional predictors are not yet common due to the high costs involved, but they are the wave of the future; one promising example is the ongoing EMBARC study, aimed at identifying features of biosignatures (including fMRI brain scans) that can explain why some people respond better to different antidepressants than others EMBARC (2013). Our methods show that it is feasible to exploit such data for developing personalized treatments, and this may provide an incentive for further studies of this type in the future.

The large sample accuracy of the proposed prediction intervals for the mean outcome resulting from the policy is established in Theorem 2, assuming that the decision boundary is avoided almost surely. Under this condition, the policy corresponding to the best fit of the model for the treatment effect (namely d_0') is unique and "decisive." When d_0' is indecisive, the proposed prediction intervals fail because the distribution of the error $(\hat{V}(\hat{d}_n) - V(\hat{d}_n))$ involved in assessing the mean outcome is not asymptotically normal. This asymptotic non-regularity occurs due to the non-smoothness of the value $V(d_0')$ at the indecisive points, and is similar to the problem of assessing the test error of a classifier. In the case of finite-dimensional predictors, a bootstrap procedure based on pre-testing has been proposed to deal with non-regularity in classification Laber and Murphy (2011); an interesting direction for future research is to develop a bootstrap procedure for pre-testing that is suitable for functional predictors.

A referee asked us why functional predictors make the problem of treatment selection more challenging, and how our approach engages this issue. With functional predictors we are potentially dealing with infinitely many variables, so the optimal policy is likely complex. This raises challenges for the statistical methodology and for its practical implementation. Computational issues become especially severe because the value function is a non-concave function of the policy (or whatever parameters are used to specify it). We have addressed this problem with an approach that is straightforward to implement, and yet that can still be justified theoretically under some natural conditions on the sample paths of the functional predictors; in addition, we have furnished several examples of functional regression models to which the theory applies.

As mentioned earlier, Zhao et al. (2012) optimize a concave surrogate for the value function, rather than the value function itself, for computationally tractability. They use an SVM approach based on a very flexible Gaussian radial basis to eliminate problems with misspecification (in an asymptotic sense), but in practice if the SVM model is misspecified, then, just as in our case, it is rather unclear what asymptotic optimality their proposed estimator would have. Moreover, their aim was somewhat different than ours — our aim, in addition to estimating the policy, has been to provide a prediction interval for the value of

the estimated policy, which was not considered by Zhao et al.; indeed, within the SVM framework, the model space was treated by Zhao et al. as having a growing dimension, so it would not be easy to provide theoretical justification for such a prediction interval.

We have restricted attention to a single-stage decision problem. However, time-varying treatments are common, and are needed, e.g., for individuals with a chronic disease who experience a waxing and waning course of illness. The goal then is to construct a policy that tailors the type and dosage of treatment through time according to the individual's changing health status. There is an abundance of statistical literature in this area (Thall et al. (2000, 2002); Murphy (2003, 2005); Robins (2004); Lunceford et al. (2002); van der Laan et al. (2005); Wahed and Tsiatis (2006)) but, to our knowledge, functional predictors such as protein expression or fMRI have not been considered. It would be interesting to extend our approach to this multi-stage setting.

Acknowledgements

The authors thank the reviewers for their helpful comments. Ian McKeague was supported by NSF Grant DMS-1307838 and NIH Grant R01GM095722-01, and Min Qian was supported by NSF Grant DMS-1307838 and NIH grants R01MH080015 and P50DA10075 (PI: Susan Murphy).

Appendix: Assumptions and Proofs

Assumptions.

(A1) The noise $\epsilon = Y - E(Y|X, A)$ is independent of (X, A). In addition, $E(Y|X, A)$ and ϵ are square integrable.

(A2) $\hat{\theta}_n$ is bounded in probability.

(A3) There exist $\upsilon > 0$ and a square integrable random variable ξ such that $|f_t(X) - f_s(X)| \leq \xi \|t - s\|^{\upsilon}$ for all $s, t \in \Theta$ a.s., where $\|\cdot\|$ is the Euclidean norm. In addition, $f_{\theta_0}(X)$ is square integrable.

(A4) The function class $\{1_{f_\theta(X)\geq 0} : \theta \in K\}$ is P-Donsker for any compact $K \subset \Theta$, and this property is preserved under products with either Y or $m(X)$.

The last three conditions hold for all the functional regression models considered in Section 4 under explicit conditions on X. In particular, for the point impact model (4.2) and the area impact model (4.3), with $X = \{X(t), t \in \mathcal{T}\}$ and $\mathcal{T}$ a subset of a finite-dimensional Euclidean space, the following three assumptions imply (A2), (A3), and (A4) needed for Theorems 1 and 2.

(A2′) $\mathcal{T}$ is compact.

(A3′) There exist $\upsilon > 0$ and a square integrable random variable ξ such that $|X(t) - X(s)| \leq \xi \|t - s\|^{\upsilon}$ for all $s, t \in \mathcal{T}$ a.s. In addition, $X(0)$ is square integrable.

(A4′) There exist $C > 0$, $r > 0$ such that $P(X(t) \in [x_1, x_2]) \leq C|x_2 - x_1|^r$ for all $t \in \mathcal{T}$, $x_1 \leq x_2$.

Remark A.1. (A2) is a mild tightness assumption needed to establish the consistency of $\hat{\theta}_n$, and is a standard condition in parametric M-estimation (see van der Vaart and Wellner, (1996, p. 308)); it is not needed if the parameter space Θ is bounded.

Remark A.2. Assumption (A3) is a Lipschitz continuity condition for the function class $\{f_\theta : \theta \in \Theta\}$ and holds for typical models, for example if f_θ is linear in $\theta \in \Theta$. Similarly, Assumption (A3′) is a Lipschitz continuity condition for the sample paths of X, and can be checked for specific processes using Kolmogorov's continuity theorem Revuz and Yor (2006). More specifically, if there exist some $c, \upsilon > 0$ and $\gamma > 0$ such that $E[|X(t) - X(s)|^\gamma] \leq c\|t - s\|^{1+\upsilon\gamma}$ for all $s, t \in \mathcal{T}$, then (A3′) holds with $\xi = \sup_{s \neq t}(|X(t) - X(s)|/\|t - s\|^{\upsilon})$. If X is a Gaussian process, then it suffices that $E[X(t) - X(s)]^2 \leq c\|t - s\|^{\upsilon}$ for some $c, \upsilon > 0$.

Remark A.3. Assumption (A4′) is needed to show the Donsker property (using bracketing entropy) for classes of weighted indicator functions involving events of the form $\{\alpha + \beta X(t) \geq 0\}$, see Lemma A.3. If X is a Gaussian process, then it suffices that $\mathrm{Var}(X(t))$ is bounded away from zero. Assumption (A4′) is used in Theorem 5 of Kuelbs et al. (2013) to prove that a class of (unweighted) indicator functions is Donsker, but it is not clear whether their approach could be extended to our setting.

Proof of Theorem 1. For any policy d, $V(d) = E[m(X) + (21_{d(X)=1} - 1)T(X)]$. Thus

$$|V(d_0') - V(\hat{d}_n)| \leq 2E\left|\left(1_{d_0'(X)=1} - 1_{\hat{d}_n(X)=1}\right) T(X)\right|.$$

Under the condition that either $f_{\theta_0}(X) \neq 0$ or $T(X) = 0$ a.s., we have

$$\begin{aligned} |V(d_0') - V(\hat{d}_n)| &\leq 2E\left|1_{f_{\theta_0}(X)\neq 0}\left(1_{d_0'(X)=1} - 1_{\hat{d}_n(X)=1}\right) T(X)\right| \\ &\leq [g(\hat{\theta}_n)]^{1/2}\left[ET^2(X)\right]^{1/2}, \end{aligned}$$

where $g(\theta) = P\big(f_\theta(X) < 0 < f_{\theta_0}(X)\big) + P\big(f_{\theta_0}(X) < 0 \leq f_\theta(X)\big)$. We show that $g(\theta)$ is continuous at $\theta = \theta_0$. This, together with $\hat{\theta}_n \xrightarrow{p} \theta_0$, Lemma A.1, guarantees $g(\hat{\theta}_n) \xrightarrow{p} g(\theta_0) = 0$ by the Continuous Mapping Theorem. Note that (A1) implies that $ET^2(X)$ is finite, and the result follows.

For any $\delta > 0$, the first term in $g(\theta)$ is upper bounded by $P(0 < f_{\theta_0}(X) \leq \delta) + P(f_{\theta_0}(X) - f_\theta(X) > \delta)$. In particular,

$$P(f_{\theta_0}(X) - f_\theta(X) > \delta) \leq \frac{E|f_{\theta_0}(X) - f_\theta(X)|}{\delta} \leq \frac{(E\xi)\|\theta - \theta_0\|^\upsilon}{\delta},$$

where the first inequality follows from Chebyshev's inequality, and the second inequality follows from (A3). Choosing $\delta = \|\theta - \theta_0\|^{\upsilon/2}$ yields

$$P(f_\theta(X) < 0 \leq f_{\theta_0}(X)) \leq P(0 < f_{\theta_0}(X) \leq \|\theta - \theta_0\|^{\upsilon/2}) + (E\xi)\|\theta - \theta_0\|^{\upsilon/2},$$

which converges to zero as $\theta \to \theta_0$. Similarly, $P(f_{\theta_0}(X) < 0 \leq f_\theta(X))$ tends to zero as $\theta \to \theta_0$. Thus $g(\theta)$ is continuous at $\theta = \theta_0$. This completes the proof.

Lemma A.1. *Suppose* $\theta_0 = \arg\min_{\theta\in\Theta} E\big[T(X) - f_\theta(X)\big]^2$ *is unique,* (A1)–(A3) *hold, and the penalty* $p_{\lambda_n}(\cdot) \to 0$ *uniformly over compact subsets of* Θ. *Then* $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Proof. Let $l_\theta(X, A, Y) = \big[Y - m(X) - f_\theta(X)A\big]^2$, $\mathbb{L}(\theta) = Pl_\theta$ and $\mathbb{L}_n(\theta) = \mathbb{P}_n l_\theta + p_{\lambda_n}(\theta)$. Then θ_0 minimizes $\mathbb{L}(\theta)$ and $\hat{\theta}_n$ minimizes $\mathbb{L}_n(\theta)$ over Θ. We show that a) $\sup_{\theta\in K} |\mathbb{L}_n - \mathbb{L}| \xrightarrow{p} 0$ for any compact set $K \subset \Theta$, and b) $\mathbb{L}(\cdot)$ is continuous. The Consistency then follows from the argmax continuous mapping theorem under (A2) and the fact that θ_0 is unique.

For a), we can ignore the penalty term because it is assumed to be asymptotically negligible uniformly over compact sets. Let $\mathcal{F} = \{f_\theta(X) : \theta \in K\}$ for any compact subset $K \subset \Theta$. Under (A3), the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\varepsilon > 0$, by Theorem 2.7.11 of van der Vaart and Wellner (1996), so $\mathcal{F}$ is P-GC, and a) follows from GC preservation properties and Assumption (A).

For b), for any θ and θ',

$$\begin{aligned}
|\mathbb{L}(\theta) - \mathbb{L}(\theta')| =& |P[f_\theta(X) - T(X)]^2 - P[f_{\theta'}(X) - T(X)]^2| \\
=& |P[f_\theta(X) + f_{\theta'}(X) - 2T(X)][f_\theta(X) - f_{\theta'}(X)]| \\
\leq& P\big|[f_\theta(X) + f_{\theta'}(X) - 2T(X)]\xi\big|\|\theta - \theta'\|^\upsilon,
\end{aligned}$$

which converges to zero as $\theta \to \theta'$, so b) holds.

In the sequel we use the notation $W_d = W(X, A, d)$, $W_n = W(X, A, \hat{d}_n)$, $W_0 = W(X, A, d_0')$, and

$$\mathcal{D}_K = \{d(X) = \operatorname{sign}(f_\theta(X)) : \theta \in K\} \tag{A.1}$$

for any compact subset $K \subset \Theta$. We also use the (empirical process) notation $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$, and the convention that P and $\mathbb{P}_n$ only act on (X, A, Y).

Proof of Theorem 2. For (a), the definition of $\hat{V}_I(d)$ implies $\mathbb{P}_n[W_n(Y - \hat{V}_I(\hat{d}_n))] = 0$ and $P[W_n(Y - V(\hat{d}_n))] = 0$, so

$$\begin{aligned}\sqrt{n}[\hat{V}_I(\hat{d}_n) - V(\hat{d}_n)] &= \mathbb{G}_n[W_n(Y - \hat{V}_I(\hat{d}_n))] \\ &= \mathbb{G}_n[W_n(Y - V(d_0'))] + [V(d_0') - \hat{V}_I(\hat{d}_n)]\mathbb{G}_n W_n. \quad \text{(A.2)}\end{aligned}$$

We show that the second term above is $o_p(1)$. By Lemma A.2 below, the class of functions $\{W_d Y : d \in \mathcal{D}_K\}$ is P-Donsker for any compact $K \subset \Theta$, and thus $(\mathbb{P}_n - P)W_n Y = o_p(1)$ by (A2). This together with Theorem 1 implies that $\mathbb{P}_n[W_n Y] = (\mathbb{P}_n - P)[W_n Y] + V(\hat{d}_n) = V(d_0') + o_p(1)$. Similarly, $\mathbb{P}_n W_n = 1 + o_p(1)$. Thus $\hat{V}_I(\hat{d}_n) = \mathbb{P}_n[W_n Y]/\mathbb{P}_n W_n \xrightarrow{p} V(d_0')$ by Slutsky's lemma. In addition, $\{W_d : d \in \mathcal{D}_K\}$ is P-Donsker, and

$$P(W_n - W_0)^2 = 4E\left[1_{\hat{d}_n(X) \neq d_0'(X)}\right] = 4g(\hat{\theta}_n) \xrightarrow{p} 0, \quad \text{(A.3)}$$

where the condition that $P(f_{\theta_0}(X) = 0) = 0$ is used in the second equality, and $g(\hat{\theta}_n)$ is defined in the proof of Theorem 1 where it was shown to be $o_p(1)$. Thus $\mathbb{G}_n W_n \xrightarrow{d} N(0, \mathrm{Var}(W_0))$ by (A2) and Lemma 19.24 of van der Vaart (1998), so by Slutsky's lemma the second term in (A.2) is $o_p(1)$. Similar arguments show that the first term in (A.2) converges in distribution to $N\left(0, \sigma_I^2\right)$, which completes the proof of the first part of (a).

The proof that $\hat{\sigma}_I^2 \xrightarrow{p} \sigma_I^2$ is based on expanding the quadratic terms in the difference between $\mathbb{P}_n[W_n^2(Y - \hat{V}_I(\hat{d}_n))^2]$ and $P[W_0^2(Y - V(d_0'))^2]$. The result then follows using similar arguments to (A.3), and P-GC properties of the classes of functions $\{W_d^2 : d \in \mathcal{D}_K\}$, $\{W_d^2 Y : d \in \mathcal{D}_K\}$ and $\{W_d^2 Y^2 : d \in \mathcal{D}_K\}$ that hold since their bracketing numbers are finite by the proof of Lemma A.2.

For (b),

$$\sqrt{n}[\hat{V}_L(\hat{d}_n) - V(\hat{d}_n)] = \mathbb{G}_n h(X, A, Y; \hat{\theta}_n)$$

where

$$h(X, A, Y; \theta) = W_d Y + (1 - W_d)m(X) + |f_\theta(X)| - W_d f_\theta(X) A,$$

and $d(X) = \mathrm{sign}(f_\theta(X))$. Here we used $V(\hat{d}_n) = EW_n Y = Eh(X, A, Y; \hat{\theta}_n)$, which holds since

$$E[W_n(m(X) + \hat{f}_n(X)A)|X] = m(X) + |\hat{f}_n(X)|.$$

We apply Lemma 19.24 of van der Vaart (1998), for which it suffices to check that the class of functions $\{h(X, A, Y; \theta) : \theta \in K\}$ is P-Donsker for any compact $K \subset \Theta$, and $P[h(X, A, Y; \hat{\theta}_n) - h(X, A, Y; \theta_0)]^2 \xrightarrow{p} 0$. The first of these follows from Lemma A.2 and Donsker preservation properties (see, e.g., Kosorok

(2008, Section 9.4)), and the second follows from Lemma A.1 and the Continuous Mapping Theorem, where the continuity condition is checked using a similar argument to (A.3).

The proof of $\hat{\sigma}_L^2 \xrightarrow{p} \sigma_L^2$ is similar to the proof of $\hat{\sigma}_I^2 \xrightarrow{p} \sigma_I^2$, and is omitted.

Lemma A.2. *Under* (A1), (A3), *and* (A4), *the classes of functions* $\mathcal{G}_1 = \{W_d : d \in \mathcal{D}_K\}$, $\mathcal{G}_2 = \{W_d Y : d \in \mathcal{D}_K\}$, $\mathcal{G}_3 = \{W_d m(X) : d \in \mathcal{D}_K\}$, $\mathcal{G}_4 = \{|f_\theta(X)| : \theta \in K\}$, *and* $\mathcal{G}_5 = \{W_d f_\theta(X) : d(X) = \operatorname{sign}(f_\theta(X)), \theta \in K\}$ *are* P*-Donsker for any compact* $K \subset \Theta$, *where* $\mathcal{D}_K$ *is the class of functions defined in* (A.1).

Proof. For any policy $d \in \mathcal{D}_K$ we have $1_{A=d(X)} = 1_{f_\theta(X)\geq 0} 1_{A=1} + 1_{f_\theta(X)<0} 1_{A=-1}$, and the randomization probability that appears in W_d is simply $p(a|x) = 1/2$. Using (A4) it then follows that $\mathcal{G}_1$ is P-Donsker, by Donsker preservation properties. Similarly, we can show that $\mathcal{G}_2$ and $\mathcal{G}_3$ are P-Donsker.

Using Theorem 2.7.11 of van der Vaart and Wellner (1996) and (A3), it can be checked that $\mathcal{G}_4$ satisfies the bracketing entropy condition, and thus $\mathcal{G}_4$ is P-Donsker.

For any function in $\mathcal{G}_5$,

$$\begin{aligned} W_d f_\theta(X) =& \big[1_{f_\theta(X)\geq 0} 1_{A=1} + 1_{f_\theta(X)<0} 1_{A=-1}\big] \frac{f_\theta(X)}{p(A|X)} \\ =& 2[1_{A=1} \max\{f_\theta(X), 0\} + 1_{A=-1} \min\{f_\theta(X), 0\}]. \end{aligned}$$

By Theorem 2.7.11 of van der Vaart and Wellner (1996) and (A3), it can be checked that $\{f_\theta(X) : \theta \in K\}$ satisfies the bracketing entropy condition, and thus is P-Donsker. Hence $\mathcal{G}_5$ is P-Donsker using Donsker preservation properties.

We show that the class of weighted indicator functions $\mathcal{G}_1$ for the point impact model (4.2) is P-Donsker, so the main part of (A4) holds; this can be checked in the same way for the area impact model (4.3).

Lemma A.3. *Suppose* $f_\theta(X) = \alpha + \beta X(\tau)$ *as in model* (4.2). *Under* (A2′), (A3′) *and* (A4′), *the class of functions* $\mathcal{G}_1 = \{1_{A=d(X)} : d \in \mathcal{D}\}$ *is* P*-Donsker, where* $\mathcal{D} = \{d(X) = \operatorname{sign}(\alpha + \beta X(\tau)) : (\alpha, \beta, \tau) = \theta \in \mathbb{R}^2 \times \mathcal{T}\}$.

Proof. For simplicity we assume $\mathcal{T} = [0, 1]$. We derive an upper bound for the bracketing number of $\mathcal{G}_1$ relative to the $L_2(P)$-norm. Note that for any policy $d(X) = \operatorname{sign}(\alpha + \beta X(\tau))$,

$$\begin{aligned} 1_{A=d(X)} =& 1_{\alpha\geq 0, \beta=0} 1_{A=1} + 1_{\alpha<0, \beta=0} 1_{A=-1} + 1_{\beta>0}[1_{A=1} 1_{X(\tau)\geq -\alpha/\beta} \\ &+ 1_{A=-1} 1_{X(\tau)<-\alpha/\beta}] + 1_{\beta<0}[1_{A=-1} 1_{X(\tau)>-\alpha/\beta} + 1_{A=1} 1_{X(\tau)\leq -\alpha/\beta}]. \end{aligned}$$

If both $1_{X(\tau)<-\alpha/\beta}$ and $1_{X(\tau)\leq -\alpha/\beta}$ fall in some bracket $[g_1, g_2]$ for $0 \leq g_1 \leq g_2$, when $\beta \neq 0$, we see that $1_{A=d(X)}$ lies in one of the four brackets:

$$[1_{A=-1}, 1_{A=-1}],\ [1_{A=1}, 1_{A=1}],\ [1_{A=-1} g_1 + 1_{A=1}(1 - g_2), 1_{A=-1} g_2 + 1_{A=1}(1 - g_1)],$$

$$[1_{A=1}g_1 + 1_{A=-1}(1-g_2), 1_{A=1}g_2 + 1_{A=-1}(1-g_1)].$$

These brackets have $L_2(P)$-size no larger than the $L_2(P)$-size of the bracket $[g_1, g_2]$.

Next we construct brackets for the class of functions $\mathcal{G}_1' = \{1_{X(\tau)<\rho}, 1_{X(\tau)\le\rho} : (\rho,\tau) \in \mathbb{R}\times[0,1]\}$. For any $0<\varepsilon<1$, choose a grid of time points $0 = t_0 < t_1 < \cdots < t_B = 1$ such that $t_b - t_{b-1} \le \varepsilon^{2/\upsilon}$ for $b = 1,\dots,B$. Clearly the total number of intervals B can be chosen smaller than $2\varepsilon^{-2/\upsilon}$. By (A3′) and Markov's inequality,

$$P\Big(\sup_{\tau\in[0,1]}|X(\tau)| > x\Big) \le \frac{E[\xi + |X(0)|]}{x}$$

for any $x > 0$. Thus, there exists a grid of points $\rho_0 < \cdots < \rho_H$ on $\mathbb{R}$ with $\rho_h - \rho_{h-1} = \varepsilon^{1/r}$ and $H < 4E[\xi + |X(0)|]\varepsilon^{-(1+1/r)}$ such that $P(X(\cdot)$ goes outside $[\rho_0,\rho_H]) < \varepsilon$. Functions in $\mathcal{G}_1'$ with $\rho > \rho_H$ are in the bracket $[1_{\sup_{\tau\in[0,1]}X(\tau)\le\rho_H}, 1]$, which has $L_2(P)$-size $[P(\sup_{\tau\in[0,1]}X(\tau) > \rho_H)]^{1/2} < \varepsilon^{1/2}$. Similarly, functions in $\mathcal{G}_1'$ with $\rho < \rho_0$ are in the bracket $[0, 1_{\inf_{\tau\in[0,1]}X(\tau)\le\rho_0}]$, which has $L_2(P)$-size smaller than $\varepsilon^{1/2}$. The remaining functions in $\mathcal{G}_1'$ are covered by the brackets $[1_{\sup_{\tau\in[t_{b-1},t_b]}X(\tau)\le\rho_{h-1}}, 1_{\inf_{\tau\in[t_{b-1},t_b]}X(\tau)\le\rho_h}]$ for $b = 1,\dots,B, h = 1,\dots,H$. Note that $1_{X(\tau)<\rho}$ and $1_{X(\tau)\le\rho}$ are in the same bracket. The $L_2(P)$-size of such a bracket is $[P(Q)]^{1/2}$, where

$$Q = \Big\{\inf_{\tau\in[t_{b-1},t_b]}X(\tau) \le \rho_h, \sup_{\tau\in[t_{b-1},t_b]}X(\tau) > \rho_{h-1}\Big\}.$$

By (A3′) and (A4′), we have

$$\begin{aligned}
P(Q\cap\{X(t_{b-1})\notin[\rho_0,\rho_H]\}) &\le P(\{X(t_{b-1})\notin[\rho_0,\rho_H]\}) \le \epsilon,\\
P(Q\cap\{X(t_{b-1})\in[\rho_{h-2},\rho_{h+1}]\}) &\le P(\{X(t_{b-1})\in[\rho_{h-2},\rho_{h+1}]\}) \le 3^rC\varepsilon,\\
P(Q\cap\{X(t_{b-1})\in[\rho_0,\rho_{h-2}]\cup[\rho_{h+1},\rho_H]\}) &\le P\Big(\sup_{\tau\in[t_{b-1},t_b]}|X(\tau)-X(t_{b-1})|>\varepsilon\Big)\\
&\le E\Big(\sup_{\tau\in[0,\varepsilon^{2/\upsilon}]}|X(t_{b-1}+\tau)-X(t_{b-1})|\Big)\varepsilon^{-1}\\
&\le \varepsilon E\xi.
\end{aligned}$$

Thus $P(Q) \le (3^rC + E\xi + 1)\varepsilon$. This implies that $\mathcal{G}_1'$ can be covered by $8E[\xi + |X(0)|]\varepsilon^{-(1+1/r+2/\upsilon)}$ brackets of $L_2(P)$-size less than $[(3^rC+E\xi+1)\varepsilon]^{1/2}$, and $\mathcal{G}_1$ can be covered by $32E[\xi+|X(0)|]\varepsilon^{-(1+1/r+2/\upsilon)}$ brackets of the same $L_2(P)$-size. Replacing $[(3^rC + E\xi + 1)\varepsilon]^{1/2}$ with ε, we obtain

$$N_{[]}(\varepsilon,\mathcal{G}_1,L_2(P)) \le 32E[\xi+|X(0)|](3^rC+E\xi+1)^{1+1/r+2/\upsilon}\varepsilon^{-2(1+1/r+2/\upsilon)}. \quad \text{(A.4)}$$

It follows that $\mathcal{G}_1$ satisfies the bracketing entropy condition, and is thus P-Donsker.

In addition to f_θ, suppose we have second model of the same form, say g_c, and let c_0, $\hat{d}_n^c$, and $\hat{d}_0^c$ correspond to θ_0, $\hat{d}_n$, and $\hat{d}_0'$, respectively.

Theorem A.1. *If the conditions of Theorem* 2 *hold for both* $f_\theta(X)$ *and* $g_c(X)$ *as models for the treatment effect* $T(X)$*, then*

(a) $\sqrt{n}\{\hat{V}_I(\hat{d}_n)-\hat{V}_I(\hat{d}_n^c)-[V(\hat{d}_n)-V(\hat{d}_n^c)]\} \xrightarrow{d} N(0,\sigma_{I,c}^2)$*, where* $\sigma_{I,c}^2 = \mathrm{Var}\big[W(X, A; d_0')(Y-V(d_0'))-W(X,A;d_0^c)(Y-V(d_0^c))\big]$ *can be consistently estimated by its sample variance with* V *replaced by* $\hat{V}_I$*,* d_0' *replaced by* $\hat{d}_n$*, and* d_0^c *replaced by* $\hat{d}_n^c$*.*

(b) $\sqrt{n}\{\hat{V}_L(\hat{d}_n)-\hat{V}_L(\hat{d}_n^c)-[V(\hat{d}_n)-V(\hat{d}_n^c)]\} \xrightarrow{d} N(0,\sigma_{L,c}^2)$*, where* $\sigma_{L,c}^2 = \mathrm{Var}\big[W(X, A; d_0')R_0 + |f_{\theta_0}(X)| - W(X,A;d_0^c)R_0^c - |g_{c_0}(X)|\big]$ *with* $R_0 = Y - m(X) - f_{\theta_0}(X)A$ *and* $R_0^c = Y - m(X) - g_{c_0}(X)A$*, can be consistently estimated by its sample variance.*

The proof is similar to that of Theorem 2 and is omitted.

Contents of supplementary material: Further results from the simulation study discussed in Section 5.1 (pdf file).

References

Byar D. P. and Corle, D. K. (1977). Selecting optimal treatment in clinical trials using covariate information. *J. Chronic Diseases* **30**, 445-459.

Cai, T., Tian, L., Lloyd-Jones, D. and Wei, L.J. (2013). Evaluating subject-level incremental values of new markers for risk classification rule. *Lifetime Data Anal.* **19**, 547-567

Cai, T., Tian, L., Uno, H., Solomon, S. D. and Wei, L. J. (2010). Calibrating parametric subject-specific risk estimation. *Biometrika* **92**, 389-404.

Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13**, 571-591.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* 2nd edition. Lawrence Erlbaum Associates, Hillsdale, NJ.

Dettling, M. and Bühlmann, P. (2004). Finding predictive gene groups from microarray data. *J. Multivariate Anal.* **90**, 106-131.

DeRubeis, R. J., Siegle, G. J. and Hollon, S. D. (2008). Cognitive therapy vs. medications for depression: treatment outcomes and neural mechanisms. *Nature Reviews Neuroscience* **9**, 788-796.

Engler, D. and Li, Y. (2009). Survival analysis with high-dimensional covariates: an application in microarray studies. *Statist. Appl. Genet. Mol. Biol.* **1**, Article 14.

Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care (EMBARC Study). `http://embarc.utsouthwestern.edu`

Fan, J. and Lv, J. (2008). Sure independence screening for ultra-high dimensional feature space (with discussion). *J. Roy. Statist. Soc. Ser. B* **70,** 849-911.

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

Gui, J. and Li, H. (2005). Penalized Cox regression analysis in the high-dimensional and low sample size settings, with applications to microarray gene expression data. *Bioinformatics* **21**, 3001-3008.

Heidecker, B. and Hare, J. M. (2007). The use of transcriptomic biomarkers for personalized medicine. *Heart Failure Rev.* **12**, 1-11.

James, G. M. and Silverman, B. W. (2005). Functional adaptive model estimation. *J. Amer. Statist. Assoc.* **100**, 565-576.

Koomen, J. M., Haura, E. B., Bepler, G., Sutphen, R., Remily-Wood, E. R., Benson, K., Hussein, M., Hazlehurst, L. A., Yeatman, T. J., Hildreth, L. T., Sellers, T. A., Jacobsen, P. B., Fenstermacher, D. A. and Dalton, W. S. (2008). Proteomic contributions to personalized cancer care. *Molecular & Cellular Proteomics* **7**, 1780-1794.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

Kuelbs, J., Kurtz, T. and Zinn, J. (2013). A CLT for empirical processes involving time-dependent data. *Ann. Probab.* **41**, 785-816.

Laber, E. and Murphy, S. A. (2011). Adaptive confidence intervals for the test error in classification. *J. Amer. Statist. Assoc.* **106**, 904-913.

Laber, E., Lizotte, D., Qian, M. and Murphy, S. A. (2010). Statistical inference in dynamic treatment regimes. Preprint, arXiv:1006.5831v2 [stat.ME].

Lu, W, Zhang, H. H. and Zeng, D. (2013). Variable selection for optimal treatment decision. *Statist. Meth. Medical Res.* **22**, 493-504.

Lunceford, J. K., Davidian, M. and Tsiatis, A. A. (2002). Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics* **58**, 48-57.

Murphy, S. A., van der Laan, M. J., Robins, J. M. and CPPRG (2001). Marginal mean models for dynamic regimes. *J. Amer. Statist. Assoc.* **96**, 1410-1423.

Murphy, S. A. (2003). Optimal dynamic treatment regimes (with discussion). *J. Roy. Statist. Soc. Ser. B* **65**, 331-366.

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

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39**, 1180-1210.

Revuz, D. and Yor, M. (2006). *Continuous Martingales and Brownian Motion.* Springer, New York.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect. *Mathematical Modelling* **7,** 1393-1512.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (Edited by M. E. Halloran and D. Berry), 95-134. Springer, New York.

Robins, J. M. (2004). Optimal regime structural nested models. In *Proceedings of the Second Seattle Symposium on Biostatistics, Lecture Notes in Statistics* (Edited by D.Y. Lin and P. J. Heagerty). Springer, New York.

Shuster, J. and Van Eys, J. (1983). Interaction between prognostic factors and treatment. *Controlled Clinical Trials* **4**, 209-214.

Song, X. and Pepe, M. S. (2004). Evaluating markers for selecting a patient's treatment. *Biometrics* **60**, 874-883.

Song, X. and Zhou, X.-H. (2011). Evaluating markers for treatment selection based on survival time. *Statist. Medicine* **30**, 2251-2264.

Taguchi, A., Politi, K., Pitteri, S. J., Lockwood, W. W., Faca, V. M., Kelly-Spratt, K, Wong, C.-H., Zhang, Q., Chin, A., Park, K.-S., Goodman, G., Gazdar, A. F., Sage, J., Dinulescu, D. M., Kucherlapati, R., DePinho, R. A., Kemp, C. J., Varmus, H. E. and Hanash, S. M. (2011). Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. *Cancer Cell* **20**, 289-299.

Thall, P. F., Millikan, R. E. and Sung, H. G. (2000). Evaluating multiple treatment courses in clinical trials. *Statist. Medicine* **19**, 1011-1028.

Thall, P. F., Sung, H. G. and Estey, E. H. (2002). Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *J. Amer. Statist. Assoc.* **97**, 29-39.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **32**, 135-166.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. Roy. Statist. Soc. Ser. B* **67**, 91-108.

Tsiatis, A. (2006). *Semiparametric Theory and Missing Data*. Springer, New York.

Wahed, A. S. and Tsiatis, A. A. (2006). Semiparametric efficient estimation of survival distribution for treatment policies in two-stage randomization designs in clinical trials with censored data. *Biometrika* **93**, 163-177.

van der Laan, M. J., Petersen, M. L. and Joffe, M.M. (2005). History-adjusted marginal structural models and statically-optimal dynamic treatment regimens. *Internat. J. Biostatist.* **1**, Article 4.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

van 't Veer, L. J. and Bernards, R. (2008). Enabling personalized cancer medicine through analysis of gene expression patterns. *Nature* **452**, 564-570.

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

Zhao, Y., Zeng, D., Rush, A. J. and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107**, 1106-1118.

Department of Biostatistics, Columbia University, 722 West 168th Street, New York, NY 10032, USA.

E-mail: im2131@columbia.edu

Department of Biostatistics, Columbia University, 722 West 168th Street, New York, NY 10032, USA.

E-mail: mq2158@columbia.edu

(Received June 2012; accepted September 2013)