

# Functional Feature Construction for Individualized Treatment Regimes

Eric B. Laber & Ana-Maria Staicu

To cite this article: Eric B. Laber & Ana-Maria Staicu (2018) Functional Feature Construction for Individualized Treatment Regimes, Journal of the American Statistical Association, 113:523, 1219-1227, DOI: [10.1080/01621459.2017.1321545](https://doi.org/10.1080/01621459.2017.1321545)

To link to this article: <https://doi.org/10.1080/01621459.2017.1321545>



View supplementary material [↗](#)



Accepted author version posted online: 26 Jun 2017.  
Published online: 26 Jun 2017.



Submit your article to this journal [↗](#)



Article views: 293



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)



# Functional Feature Construction for Individualized Treatment Regimes

Eric B. Laber and Ana-Maria Staicu

Department of Statistics, North Carolina State University, Raleigh, NC

## ABSTRACT

Evidence-based personalized medicine formalizes treatment selection as an individualized treatment regime that maps up-to-date patient information into the space of possible treatments. Available patient information may include static features such as race, gender, family history, genetic and genomic information, as well as longitudinal information including the emergence of comorbidities, waxing and waning of symptoms, side-effect burden, and adherence. Dynamic information measured at multiple time points before treatment assignment should be included as input to the treatment regime. However, subject longitudinal measurements are typically sparse, irregularly spaced, noisy, and vary in number across subjects. Existing estimators for treatment regimes require equal information be measured on each subject and thus standard practice is to summarize longitudinal subject information into a scalar, ad hoc summary during data preprocessing. This reduction of the longitudinal information to a scalar feature precedes estimation of a treatment regime and is therefore not informed by subject outcomes, treatments, or covariates. Furthermore, we show that this reduction requires more stringent causal assumptions for consistent estimation than are necessary. We propose a data-driven method for constructing maximally prescriptive yet interpretable features that can be used with standard methods for estimating optimal treatment regimes. In our proposed framework, we treat the subject longitudinal information as a realization of a stochastic process observed with error at discrete time points. Functionals of this latent process are then combined with outcome models to estimate an optimal treatment regime. The proposed methodology requires weaker causal assumptions than Q-learning with an ad hoc scalar summary and is consistent for the optimal treatment regime. Supplementary materials for this article are available online.

## ARTICLE HISTORY

Received March 2015  
Revised January 2017

## KEYWORDS

Functional data; Personalized medicine; Regression; Treatment regimes

## 1. Introduction

It is widely recognized by clinical and intervention scientists that the best possible care requires personalized treatment based on individual patient characteristics (Piquette-Miller and Grant 2007; Insel 2009; Hamburg and Collins 2010; Lauer and Collins 2010; Topol 2012). Individualized treatment regimes (ITRs) formalize personalized treatment selection as a function from current patient information to a recommended treatment. An optimal ITR maximizes the expectation of a desirable outcome when applied to an entire population of interest. There has been a recent surge of research on developing estimators for optimal ITRs including regression-based estimators (Lu, Zhang, and Zeng 2011; Qian and Murphy 2011a; McKeague and Qian 2011), and direct-search estimators (Orellana, Rotnitzky, and Robins 2010; Zhao et al. 2012; Zhang et al. 2012a, b). ITRs are the atoms of dynamic treatment regimes (Robins 1986, 1989, 1997, 2004; Murphy 2003), which are currently a subject of intense research including both estimation (Henderson, Ansell, and Alshibani 2010a; Zhao et al. 2011; Goldberg and Kosorok 2012; Zhang et al. 2013; Moodie, Dean, and Sun 2013; Laber, Lizotte, and Ferguson 2014) and inference (Chakraborty, Murphy, and Strecher 2010; Moodie and Richardson 2010; Chakraborty, Laber, and Zhao 2013; Laber et al. 2013b); recent surveys include Schulte et al. (2014); Chakraborty and Moodie (2013); Chakraborty and Murphy (2014).

Despite this influx of research, most existing methods for estimating optimal ITRs from observational or randomized study data assume that each patient is measured at the same, fixed time points and that patient measurements are made without error. This is unrealistic in many clinical settings where noisy patient measurements are taken at clinic visits, which are sparsely observed and irregularly spaced. Thus, to apply existing estimation methods, a standard approach is to reduce patient longitudinal information into an ad hoc scalar summary, which is assumed to be measured perfectly. However, this potentially obfuscates important scientific information contained the patient longitudinal process, ignores measurement error, requires strong causal assumptions, and may lead to an ITR of poor quality.

We model the pretreatment patient longitudinal information as a noisy proxy of a smooth latent process and employ methods from functional data analysis to form a regression-based estimator of the optimal ITR. We assume that the effect of the true latent process is represented either as: (i) an inner product of the latent process and an unknown, nonparametrically modeled, coefficient function; or (ii) a nonlinear parametric functional of the latent process. Under mild regularity conditions, we prove that the proposed estimator is consistent if the true optimal decision rule is linear in the static patient information and depends on the true latent process as described

above. The proposed estimator applies to both randomized and observational studies with sparsely observed, irregularly spaced, and noisy longitudinal measurements. While we focus on a single binary treatment decision, the proposed work provides a foundation for several extensions including multiple treatments and treatment decision points.

There are several recently proposed methods that use functional data to inform construction of optimal ITRs. McKeague and Qian (2014), Ciarleglio et al. (2015), and Ciarleglio et al. (2016) developed estimators for optimal ITRs that allow for pretreatment functional predictors that are densely sampled without noise. In contrast, as mentioned previously, our focus is on sparse, noisy, and irregularly spaced data.

Our approach uses nonparametric estimators of the mean and covariance functions of the latent process. An alternative approach would be to posit parametric models for the mean and/or covariance functions using, for example, mixed-effects models (Laird and Ware 1982; Diggle et al. 2002; Fitzmaurice, Laird, and Ware 2012); this approach is especially appealing in contexts where scientific theory can be used to inform the choice of parametric model. Our proposed approach can also be used with such models, however, in our motivating biomedical applications there is not sufficient theory to inform the choice of these models so we will not discuss this approach further.

In Section 2, we review regression-based estimators of the optimal ITR and conditions needed to identify the optimal ITR using randomized or observational data. In Section 3, we describe our proposed estimator and state conditions under which it consistently estimates the optimal ITR. The finite sample performance of the proposed estimator is illustrated using simulation experiments in Section 4 and application to a study of major depressive disorder in Section 5. Extensions and open problems are discussed in Section 6.

## 2. Regression-Based Methods for Estimating Optimal ITRs

Assume that data available to estimate an ITR are of the form  $\mathcal{D}_n = \{(\mathbf{X}_i, W_i(\mathbf{T}_i), A_i, Y_i)\}_{i=1}^n$ , which comprise  $n$  iid copies of a trajectory  $\{\mathbf{X}, W(\mathbf{T}), A, Y\}$  where:  $\mathbf{X} \in \mathbb{R}^p$  denotes pretreatment subject covariate information;  $W(\mathbf{T}) = \{W(T_1), \dots, W(T_M)\}$  denotes  $M$  pretreatment proxy measurements taken at times  $\mathbf{T} = (T_1, \dots, T_M)$ ;  $A \in \{-1, 1\}$  denotes treatment assigned; and  $Y \in \mathbb{R}$  denotes an outcome coded so that higher values are better. Capital letters, like  $\mathbf{X}$ , denote random variables and lowercase letters, like  $\mathbf{x}$ , denote instances of these random variables. Both  $M$  and  $\mathbf{T}$  are treated as random variables as the number and timing of observations varies across subjects. Observation times  $\mathbf{T}$  are allowed to vary in number, be sparse, and irregularly spaced. For example, in the STAR\*D study of major depressive disorder,  $W(\mathbf{T})$  might denote patient-reported Quick Inventory of Depressive Symptomatology (QIDS-SR, Rush et al. 2003), recorded at each clinic visit.

An ITR  $\pi$  maps available patient information to a recommended treatment so that  $\pi : \text{dom } \mathbf{X} \times \text{dom } W(\mathbf{T}) \rightarrow \text{dom } A$ . Under  $\pi$ , a patient presenting with  $\mathbf{X} = \mathbf{x}$  and  $W(\mathbf{T}) = w(\mathbf{t})$  would be recommended treatment  $\pi\{\mathbf{x}, w(\mathbf{t})\}$ . To define an optimal ITR, we use potential outcomes (Rubin 1978;

Splawa-Neyman Dabrowska, and Speed 1990). Let  $Y^*(a)$  denote the potential outcome under treatment  $a \in \{-1, 1\}$  and define  $Y^*(\pi) = \sum_{a \in \{-1, 1\}} Y^*(a) 1_{\pi\{\mathbf{X}, W(\mathbf{T})\} = a}$  to be the potential outcome under ITR  $\pi$ . The optimal ITR, say  $\pi^{\text{opt}}$ , satisfies  $\mathbb{E}Y^*(\pi^{\text{opt}}) \geq \mathbb{E}Y^*(\pi)$  for all  $\pi$ . To define the optimal ITR in terms of the data-generating model we use assume: (C1) consistency,  $Y = Y^*(A)$ ; (C2) positivity, there exists  $\epsilon > 0$  so that  $\epsilon < P\{A = a | \mathbf{X}, W(\mathbf{T})\}$  for each  $a \in \{-1, 1\}$  with probability one; and (C3) ignorability,  $\{Y^*(-1), Y^*(1)\} \perp\!\!\!\perp A | \mathbf{X}, W(\mathbf{T})$ . Assumption (C3) is true by construction in a randomized study, but unverifiable with data from an observational study (Robins, Hernan, and Brumback 2000). A common justification for (C3) in the analysis of an observational study is that patient covariates, that is,  $\{\mathbf{X}, W(\mathbf{T})\}$ , are sufficiently rich so as to capture all information used in the treatment decision process. However, current analysis methods *implicitly* require the stronger causal assumption (C3')  $\{Y^*(-1), Y^*(1)\} \perp\!\!\!\perp A | \mathbf{X}, s\{W(\mathbf{T})\}$ , where  $s\{w(\mathbf{t})\}$  is a prespecified, ad hoc, summary of  $w(\mathbf{t})$ . The proposed method incorporates  $w(\mathbf{t})$  directly into the model and is thereby more consistent with (C3).

Define  $Q\{\mathbf{x}, w(\mathbf{t}), a\} = \mathbb{E}\{Y | \mathbf{X} = \mathbf{x}, W(\mathbf{T}) = w(\mathbf{t}), A = a\}$ . Under (C1)–(C3), it can be shown that  $\pi^{\text{opt}}\{\mathbf{x}, w(\mathbf{t})\} = \arg \max_{a \in \{-1, 1\}} Q\{\mathbf{x}, w(\mathbf{t}), a\}$ . Regression-based estimators postulate a working model for  $Q\{\mathbf{x}, w(\mathbf{t}), a\}$ , which is fit using the data, say  $\hat{Q}\{\mathbf{x}, w(\mathbf{t}), a\}$ , and then used to infer the optimal ITR through  $\hat{\pi}^Q\{\mathbf{x}, w(\mathbf{t})\} = \arg \max_a \hat{Q}\{\mathbf{x}, w(\mathbf{t}), a\}$ . Postulating a working model for  $Q\{\mathbf{x}, w(\mathbf{t}), a\}$  that fully incorporates  $w(\mathbf{t})$  is difficult because this information is often sparsely observed and irregularly spaced. Thus, a common strategy is to assume  $Q\{\mathbf{x}, w(\mathbf{t}), a\} = Q\{\mathbf{x}, s\{w(\mathbf{t})\}, a\}$  for some scalar summary  $s\{w(\mathbf{t})\}$ . In this case, one can use standard regression methods, for example, linear regression, to estimate  $Q\{\mathbf{x}, w(\mathbf{t}), a\}$ , by regressing  $Y$  on  $\mathbf{X}$ ,  $s\{W(\mathbf{T})\}$ , and  $A$ . For example, Chakraborty, Laber, and Zhao (2013), Chakraborty and Moodie (2013), Laber, Linn, and Stefanski (2013a), and Schulte et al. (2014) used the slope of QIDS-SR,  $s\{w(\mathbf{t})\} = \{w(t_m) - w(t_1)\}/(t_m - t_1)$  in their analyses of STAR\*D; while this quantity is informed by clinical expertise, it is not informed by the response  $Y$ , the treatment  $A$ , nor does it account for potential measurement error in QIDS-SR as a proxy for depression severity. Another common ad hoc summary is the most recent pretreatment measured score  $s\{w(\mathbf{t})\} = t_m$  (e.g., Shortreed et al. 2011; Nahum-Shani et al. 2012; Laber et al. 2013b; Laber, Lizotte, and Ferguson 2014); again, this summary may be informed by expert judgment but is not data-driven. In the context of an observational study, allowing  $w(\mathbf{t})$  to enter regression model only through  $s\{w(\mathbf{T})\}$  is implicitly assuming that the stronger causal assumption (C3') holds.

For clarity, we describe a simple version of regression-based estimation of the optimal ITR using a linear working model for  $Q\{\mathbf{x}, w(\mathbf{t}), a\}$  of the form  $Q\{\mathbf{x}, w(\mathbf{t}), a; \theta\} = \mathbf{x}^\top \alpha + \beta s\{w(\mathbf{t})\} + a[\mathbf{x}^\top \delta + \gamma s\{w(\mathbf{t})\}]$ , where  $\theta = (\alpha^\top, \beta, \delta^\top, \gamma)^\top$  and  $s\{w(\mathbf{t})\}$  is a fixed and known scalar summary of  $w(\mathbf{t})$ . Let  $\mathbb{P}_n$  denote the empirical measure and define  $\hat{\theta}_n = \arg \min_{\theta} \mathbb{P}_n[Y - Q\{\mathbf{X}, W(\mathbf{T}), A; \theta\}]^2$ . A regression-based estimator of the optimal ITR is  $\hat{\pi}^Q\{\mathbf{x}, w(\mathbf{t})\} = \arg \max_a Q\{\mathbf{x}, w(\mathbf{t}), a; \hat{\theta}_n\}$ . Regression-based estimators are convenient, intuitive, and benefit from standard regression diagnostics and model-building techniques. For these reasons,

we use the foregoing regression-based estimator as the basis for the estimators described in Section 3. However, in some settings, direct-search estimators (Zhao et al. 2012; Zhang et al. 2012a, b, 2013) may be more appropriate; see Remark 1 in Section 3 for a description of direct-search estimation.

### 3. Functional Q-Learning for Estimating Optimal ITRs

Treating patient longitudinal information as sparse functional data, we derive estimators of the optimal ITR,  $\pi^{\text{opt}}\{\mathbf{x}, w(\mathbf{t})\}$ , that do not require ad hoc summary of  $w(\mathbf{t})$ . Without loss of generality suppose that  $0 \leq T_1 < \dots < T_M \leq 1$ . We assume that  $W(t) = Z(t) + \epsilon(t)$  for all  $t \in [0, 1]$ , where  $Z(\cdot)$  is a square integrable latent process with smooth mean and covariance functions, and  $\epsilon(\cdot)$  is mean-zero white-noise process. Thus, we assume that the subject longitudinal measurements  $W(\mathbf{T})$  are noisy realizations of the true latent process  $Z(\cdot)$  at time points  $\mathbf{T}$ . We introduce two general models for  $\mathbb{E}\{Y|\mathbf{X}, Z(\cdot), A\}$ : (i) a linear model in  $\mathbf{X}$ ,  $A$ , and  $Z(\mathbf{T})$ ; and (ii) a hybrid model that combines linear effects for  $\mathbf{X}$  and  $A$  and a linear effect of  $f\{Z(\cdot); \eta\}$  for a parametric functional  $f(\cdot; \eta)$  indexed by unknown parameters  $\eta$ . These two classes of models serve as building blocks for more complex classes of models, which could be obtained by adding polynomial terms or other nonlinear basis functions.

#### 3.1. Linear Working Model for $\mathbb{E}\{Y|\mathbf{X}, Z(\cdot), A\}$

We assume a linear model of the form

$$(A1) \quad \mathbb{E}\{Y|\mathbf{X} = \mathbf{x}, Z(\cdot) = z(\cdot), A = a\} \\ = \mathbf{x}^T \alpha^* + \int_0^1 z(t) \beta^*(t) dt + a \left\{ \mathbf{x}^T \delta^* + \int_0^1 z(t) \gamma^*(t) dt \right\},$$

where  $\alpha^*$  and  $\beta^*(\cdot)$  are the main effects associated with  $\mathbf{X}$  and  $Z(\cdot)$  and  $\delta^*$  and  $\gamma^*(\cdot)$  are interaction effects. If the coefficient functions  $\beta^*(\cdot)$  and  $\gamma^*(\cdot)$  were known and the latent process  $Z(\cdot)$  were observed, then (A1) would correspond to a linear model. However, the coefficient functions are unknown and  $Z(\cdot)$  is not observed. Model (A1) was studied by McKeague and Qian (2014) and Ciarleglio et al. (2015) in the setting where  $Z(\cdot)$  is observed on a fine grid of points for each subject.

The idea behind our estimation procedure is to use orthogonal basis function expansions for both the latent process and the unknown coefficient functions, and then reduce the infinite summation by an appropriate finite truncation. Let the spectral decomposition of the covariance function,  $G(t, t') = \text{cov}\{Z(t), Z(t')\}$ , be  $G(t, t') = \sum_{k \geq 1} \lambda_k \phi_k(t) \phi_k(t')$ , where  $\{\lambda_k, \phi_k(\cdot)\}_{k \geq 1}$  are the pairs of eigenvalues/eigenfunctions with  $\lambda_1 > \lambda_2 > \dots \geq 0$ , and  $\{\phi_k(\cdot)\}_{k \geq 1}$  form an orthonormal basis in  $L^2[0, 1]$ . Then  $Z(\cdot)$  can be represented using the Karhunen–Loève expansion as  $Z(t) = \mu(t) + \sum_{k \geq 1} \xi_k \phi_k(t)$ , where  $\xi_k \triangleq \int_0^1 \{Z(t) - \mu(t)\} \phi_k(t) dt$  are functional principal component scores, which have mean zero, variance  $\lambda_k$ , and are mutually uncorrelated. Using the eigenfunctions  $\{\phi_k(\cdot)\}_{k \geq 1}$ , the coefficient functions  $\beta^*(\cdot)$  and  $\gamma^*(\cdot)$  can be represented as  $\beta^*(t) = \sum_{k \geq 1} \beta_k^* \phi_k(t)$  and  $\gamma^*(t) = \sum_{k \geq 1} \gamma_k^* \phi_k(t)$ , where  $\beta_k^* \triangleq \int_0^1 \beta(t) \phi_k(t) dt$  and  $\gamma_k^* \triangleq \int_0^1 \gamma(t) \phi_k(t) dt$  are unknown basis coefficients. We assume that  $\mathbf{X}$  has an intercept term. Thus, the model (A1) can be equivalently written as  $\mathbb{E}\{Y|\mathbf{X} =$

$\mathbf{x}, Z(\cdot) = z(\cdot), A = a\} = \mathbf{x}^T \alpha^* + \sum_{k \geq 1} \xi_k \beta_k^* + a(\mathbf{x}^T \delta^* + \sum_{k \geq 1} \xi_k \gamma_k^*)$ , where we have absorbed  $\int \mu(t) \beta^*(t) dt$  and  $\int \mu(t) \gamma^*(t) dt$  into the intercept and main effect of treatment.

To obtain an expression for  $Q\{\mathbf{x}, w(\mathbf{t}), a\}$  we assume:

- (A2)  $Y$  is conditionally independent of  $W(\cdot)$  given  $Z(\cdot)$ ;
- (A3)  $A$  is independent of  $Z(\cdot)$  given  $\mathbf{X}$  and  $W(\mathbf{T})$ , and  $0 < c_1 \leq P(A = 1|\mathbf{X}, W(\mathbf{T})) \leq c_2 < 1$  with probability one for fixed constants  $c_1$  and  $c_2$ ;
- (A4)  $Z(\cdot)$  is independent of  $\mathbf{X}$  given  $W(\mathbf{T})$ ;
- (A5)  $Z(\cdot), W(\cdot), \beta(\cdot)$ , and  $\gamma(\cdot)$ , belong to  $L^2[0, 1]$ .

The foregoing assumptions are relatively mild: (A2) states that the noise-corrupted process  $W(\cdot)$  provides no additional information about  $Y$  given the true signal  $Z(\cdot)$ ; (A3) is satisfied if treatment is assigned according to observables  $\mathbf{X}$  and  $W(\mathbf{T})$  and all subjects have a positive probability of receiving each treatment; (A4) allows us to estimate  $Z(\cdot)$  using  $W(\mathbf{T})$  without reference to  $\mathbf{X}$ , this simplifies notation and calculations but is not necessary (Jiang and Wang 2010).

Under assumptions (A1)–(A5)

$$Q\{\mathbf{x}, w(\mathbf{t}), a\} = \mathbf{x}^T \alpha^* + \sum_{k \geq 1} \beta_k^* \mathbb{E}\{\xi_k | W(\mathbf{T}) = w(\mathbf{t})\} \\ + a \left[ \mathbf{x}^T \delta^* + \sum_{k \geq 1} \gamma_k^* \mathbb{E}\{\xi_k | W(\mathbf{T}) = w(\mathbf{t})\} \right]. \quad (1)$$

Let  $\ell_k\{w(\mathbf{t})\} \triangleq \mathbb{E}\{\xi_k | W(\mathbf{T}) = w(\mathbf{t})\}$  and let  $K$  be the number of terms in a finite truncation of the infinite summations in the foregoing display. Under the assumption of joint normality of the functional principle components and measurement error,  $\ell_k\{w(\mathbf{t})\} = \lambda_k \phi_k(\mathbf{t}) \{\Phi(\mathbf{t}) \Lambda \Phi(\mathbf{t})^T + \sigma^2 I_m\}^{-1} \Phi(\mathbf{t}) w(\mathbf{t})$ , where  $\phi_k(\mathbf{t}) = \{\phi_k(t_1), \dots, \phi_k(t_m)\}^T$ ,  $\Phi(\mathbf{t})$  is the  $m \times K$  matrix  $\{\phi_1(\mathbf{t}) \dots \phi_K(\mathbf{t})\}$ ,  $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_K\}$  is  $K \times K$  diagonal matrix,  $I_m$  is the  $m \times m$  identity matrix, and  $\text{var}\{\epsilon(t)\} \equiv \sigma^2$  is the error variance.

From (1), if the functions  $\ell_k\{w(\mathbf{t})\}$   $k \geq 1$  were known, then the parameters  $\alpha^*, \delta^*, \beta^*, \gamma^*$  could be estimated by the regression of  $Y$  on  $\mathbf{X}, \ell_k\{W(\mathbf{T})\}$   $k \geq 1, A$ . However, these functions are not known because the process  $Z(\cdot)$  is not directly observable, the basis functions  $\{\phi_k(\cdot)\}_{k \geq 1}$  are not known, or both. We use functional principal component analysis (Silverman and Ramsay 2005; Yao, Müller, and Wang 2005; Di et al. 2009) to construct estimators  $\hat{\ell}_{n,k}\{w(\mathbf{t})\}$  of  $\ell_k\{w(\mathbf{t})\}$  for  $k = 1, \dots, K$ . Subsequently, we regress  $Y$  on  $\mathbf{X}, A$ , and  $\hat{\ell}_{n,k}\{W(\mathbf{T})\}$   $1 \leq k \leq K$ . The complete algorithm, which we term linear functional Q-learning, is as follows.

(FQ1) Use functional principal components to construct estimators  $\hat{\phi}_{n,k}(\cdot)$  of  $\phi_k(\cdot)$ , and  $\hat{\ell}_{n,k}\{w(\mathbf{t})\}$  of  $\ell_k\{w(\mathbf{t})\}$ .

(FQ2) Let  $\theta \triangleq (\alpha^T, \delta^T, \beta_1, \dots, \beta_K, \gamma_1, \dots, \gamma_K)^T$ , define

$$\hat{Q}_n^K\{\mathbf{x}, w(\mathbf{t}), a; \theta\} \triangleq \mathbf{x}^T \alpha + \sum_{k=1}^K \beta_k \hat{\ell}_k(w(\mathbf{t})) \\ + a \left\{ \mathbf{x}^T \delta + \sum_{k=1}^K \gamma_k \hat{\ell}_k(w(\mathbf{t})) \right\},$$

and compute

$$\hat{\theta}_n = \arg \min_{\theta} \mathbb{P}_n[Y - \hat{Q}_n^K\{\mathbf{X}, W(\mathbf{T}), A; \theta\}]^2.$$

(FQ3) The estimated optimal decision rule is  $\hat{\pi}_n^{\text{FQ}}(\mathbf{x}, w(\mathbf{t})) = \arg \max_a \hat{Q}_n^K\{\mathbf{x}, w(\mathbf{t}), a; \hat{\theta}_n\}$ .



The finite truncation  $K$  corresponds to the number of leading functional principal components and can be selected using the percentage variance explained, cross-validation, or information criteria (see Yao, Müller, and Wang 2005). In our simulation experiments and illustrative application, we chose  $K$  using the percentage variance explained.

The following results provide convergence rates for the functional  $Q$ -learning estimator. Subsequently, we establish that  $\hat{\pi}_n^{\text{FQ}}$  is consistent for the optimal regime  $\pi^{\text{opt}}$  and derive rates of convergence for the difference  $\mathbb{E}Y^*(\pi^{\text{opt}}) - \mathbb{E}Y^*(\hat{\pi}_n^{\text{FQ}})$ , commonly known as the *regret* of applying the estimated regime  $\hat{\pi}_n^{\text{FQ}}$  (Lai and Robbins 1985; Henderson, Ansell, and Alshibani 2010b). Recall that  $\mathcal{D}_n$  denotes the observed data. Proofs are given in the supplemental material. We use the following assumptions:

- (A6)  $\lim_{n \rightarrow \infty} \bigcup_{i=1}^n \mathbf{T}_i$  forms a dense set in  $[0, 1]$  with probability one;
- (A7)  $\epsilon(\cdot)$  is a Gaussian white-noise process;
- (A8)  $\sup_{t \in \mathcal{T}} |\hat{\mu}(t) - \mu(t)| = O_p(n^{-\Delta})$  for some positive  $\Delta$ , where  $\hat{\mu}(t)$  is an estimator of  $\mu(t) \triangleq \mathbb{E}\{Z(t)\}$ ;
- (A9)  $\sup_{t, t' \in \mathcal{T}} |\hat{G}(t, t') - G(t, t')| = O_p(n^{-\Delta})$  for some positive  $\Delta$ , where  $\hat{G}(t, t')$  is an estimator of  $G(t, t') \triangleq \text{cov}\{Z(t), Z(t')\}$ ;
- (A10)  $|\hat{\sigma}^2 - \sigma^2| = O_p(n^{-\Delta})$  for some positive  $\Delta$ , where  $\hat{\sigma}^2$  is an estimator of  $\sigma^2 \triangleq \text{var}\{\epsilon(t)\}$ ;
- (A11)  $\mathbb{E}\|\mathbf{X}\|^4 < \infty$  and  $E[\mathbf{X}\mathbf{X}^\top]$  is nonsingular;
- (A12)  $\sum_{k \geq 1} \lambda_k^2 < \infty$ ;
- (A13)  $|\beta_k^*| \leq Ck^{-\varsigma}$  and  $|\gamma_k^*| \leq Ck^{-\varsigma}$  for all  $k \geq 1$ , some  $\varsigma > 1$ , and some  $C > 0$ .

Assumptions (A5)–(A10) are standard in sparse functional data analysis (Yao, Müller, and Wang 2005; Hall and Hosseini-Nasab 2006; Li and Hsing 2010; Staicu et al. 2014); (A11) is a standard assumption in regression; and (A12)–(A13) are common in scalar-on-function regression models (Hall et al. 2007; Lian 2011; Shin and Lee 2012).

**Theorem 1.** Assume (A1)–(A13). Let  $K_n$  be an increasing sequence of integers such that  $K_n \rightarrow \infty$  and  $K_n/n^{2\Delta} \rightarrow 0$  as  $n \rightarrow \infty$ , then

$$\begin{aligned} & \mathbb{E} \left\{ \left| \hat{Q}_{K_n}^{\pi^{\text{opt}}} \{\mathbf{X}, W(\mathbf{T}), A; \hat{\theta}_n\} - Q\{\mathbf{X}, W(\mathbf{T}), A\} \right| \middle| \mathcal{D}_n \right\} \\ &= O_p(K_n n^{-1/2} + K_n^{1/2} n^{-\Delta}). \end{aligned}$$

The sequence  $K_n$  is dictated by the rate of decay of the tail sum  $\sum_{k \geq m} \lambda_k^2$  as  $m \rightarrow \infty$ ; in the extreme case in which  $\lambda_k = 0$  for all sufficiently large  $k$  then one can set  $K_n$  to be a constant (i.e., it need not diverge to infinity, see the supplemental materials for additional details). As indicated by (A8)–(A10), the constant  $\Delta$  is dictated by the slowest rate of convergence among the estimators  $\hat{\mu}$ ,  $\hat{G}$ , and  $\hat{\sigma}$ ; this rate depends on the sampling design as well as the smoothness of the mean and covariance functions of the latent process (Yao, Müller, and Wang 2005; Hall and Hosseini-Nasab 2006; Li and Hsing 2010). For example, with kernel-based estimators of the mean and covariance functions under a sparse sampling design, it can be shown that  $\Delta = 1/4$  is sufficient under mild regularity conditions (see Yao, Müller, and Wang 2005). If the functional predictor is observed on a dense grid without error, then it is possible to obtain  $\Delta = 1/2$  (Zhang and Wang 2016). It can be seen that if  $\Delta = 1/2$  and  $\lambda_k = 0$  for all but finitely many values, so that  $K_n$

can be chosen to be a constant, then the rate in the preceding theorem is parametric; however, in the sparse and noisy design we consider here we expect  $\Delta < 1/2$  (see Zhang and Wang 2016) in which case the resulting rates are sub-parametric.

**Corollary 1.** Assume that the conditions for Theorem 1 hold and furthermore that  $P[Q\{\mathbf{X}, W(\mathbf{T}), 1\} - Q\{\mathbf{X}, W(\mathbf{t}), -1\} = 0] = 0$ , then  $\mathbb{E}\{|\hat{\pi}_n^{\text{FQ}}\{\mathbf{X}, W(\mathbf{T})\} - \pi^{\text{opt}}\{\mathbf{X}, W(\mathbf{T})\}| \middle| \mathcal{D}_n\}$  converges to zero in probability.

**Corollary 2** (Qian and Murphy 2011a). Assume that the conditions for Theorem 1 hold, then

$$\mathbb{E}Y^*(\pi^{\text{opt}}) - \mathbb{E}\{Y^*(\hat{\pi}_n^{\text{FQ}})\} = O_p\{(K_n n^{-1/2} + K_n^{1/2} n^{-\Delta})\}.$$

The estimated coefficients  $\hat{\beta}_{n,k}$ ,  $\hat{\gamma}_{n,k}$ ,  $k = 1, \dots, K$  and the estimated basis functions  $\hat{\phi}_{n,k}(\cdot)$  can be used to form estimates of the coefficient functions  $\beta(t)$  and  $\gamma(t)$  indexing the  $Q$ -function  $Q\{\mathbf{x}, w(\mathbf{t}), a\}$  via  $\hat{\beta}_n(t) = \sum_{k=1}^K \hat{\beta}_{n,k} \hat{\phi}_{n,k}(t)$  and  $\hat{\gamma}_n(t) = \sum_{k=1}^K \hat{\gamma}_{n,k} \hat{\phi}_{n,k}(t)$ . Plotting the estimated coefficient functions shows how the latent process is being weighted over  $[0, 1]$  in the  $Q$ -function and thus may be scientifically informative. For example, a coefficient function that is large in magnitude and of opposite in sign at the endpoints of  $[0, 1]$  but small in magnitude in the middle of  $[0, 1]$  suggests that the change in  $Z(t)$  from the beginning to the end of the observation period is important. In this way, the linear functional  $Q$ -learning algorithm can uncover important relationships between the outcome and evolving patient characteristics including how these characteristics should be used to dictate treatment. However, in some cases, existing scientific theory may suggest a particular form for  $\mathbb{E}\{Y|\mathbf{X}, Z(\cdot), A\}$ , which may be nonlinear. We consider this case next.

### 3.2. Nonlinear Working Models for $\mathbb{E}\{Y|\mathbf{X}, Z(\cdot), A\}$

In some settings, nonlinear features of a patient's latent longitudinal process may be highly clinically informative. For example, in the context of depression and anxiety, an important feature of a patient's evolving health status is the degree of waxing and waning of patient symptomatology (Wittchen et al. 2000). In this section, we develop functional  $Q$ -learning nonlinear summaries of the latent process that are known and suggested by domain experts or are unknown but assumed to belong to a parametric class of nonlinear models; the extension to nonparametric classes of functionals is discussed in Remark 2. In the context of major depressive disorder, an important feature might be the variation in an individual's depression level as captured by  $f\{Z(\cdot); \eta\} = \eta \int_0^1 |Z'(t)| dt$ , where  $\eta \in \mathbb{R}$  and  $Z'(t)$  denotes the derivative of  $Z(t)$ . In this section, we construct estimators of  $\pi^{\text{opt}}$  under the assumption

$$(A1') \quad \mathbb{E}\{Y|\mathbf{X} = \mathbf{x}, Z(\cdot) = z(\cdot), A = a\} = \mathbf{x}^\top \alpha^* + f\{z(\cdot); \eta\} + a[\mathbf{x}^\top \delta^* + g\{z(\cdot); \rho\}],$$

where  $f$  and  $g$  are functionals from  $L^2[0, 1]$  in to  $\mathbb{R}$  indexed by unknown parameters  $\eta \in \mathbb{R}^k$  and  $\rho \in \mathbb{R}^q$ . If  $Z(\cdot)$  was observed, then (A1') reduces to a parametric nonlinear regression model. Under (A1') and (A2)–(A5)

$$\begin{aligned} Q\{\mathbf{x}, w(\mathbf{t}), a\} &= \mathbf{x}^\top \alpha^* + \mathbb{E}[f\{Z(\cdot); \eta\} | W(\mathbf{T}) = w(\mathbf{t})] \\ &\quad + a(\mathbf{x}^\top \delta^* + \mathbb{E}[g\{Z(\cdot); \rho\} | W(\mathbf{T}) = w(\mathbf{t})]). \end{aligned}$$

Let  $G\{\cdot|w(\mathbf{t})\}$  denote the conditional distribution of  $Z(\cdot)$  given  $W(\mathbf{T}) = w(\mathbf{t})$ , then

$$Q(\mathbf{x}, w(\mathbf{t}), a) = \mathbf{x}^\top \alpha^* + \int f\{u(\cdot); \eta\} dG\{u(\cdot)|w(\mathbf{t})\} + a \left[ \mathbf{x}^\top \delta^* + \int g\{u(\cdot); \rho\} dG\{u(\cdot)|w(\mathbf{t})\} \right]. \quad (2)$$

We construct an estimator of  $Q(\mathbf{x}, w(\mathbf{t}), a)$  by first estimating  $G\{\cdot|w(\mathbf{T})\}$  and then applying nonlinear least squares to estimate the remaining parameters in (2). To estimate  $G\{\cdot|w(\mathbf{t})\}$  we assume that the loadings  $\{\xi_k\}_{k \geq 1}$  are a discrete-time, mean-zero Gaussian process with independent components and  $\text{var}(\xi_k) = \lambda_k$  for all  $k \geq 1$ . Under this assumption it can be seen that  $Z(\cdot)$  given  $W(\mathbf{T}) = w(\mathbf{t})$  follows a Gaussian process with mean function  $m\{w(\mathbf{t})\}(t) = \mathbb{E}\{Z(t)|W(\mathbf{t}) = w(\mathbf{t})\}$  and covariance kernel

$$\mathcal{K}\{w(\mathbf{t})\}(t, v) = \mathbb{E}\{[Z(t) - m\{w(\mathbf{t})\}(t)][Z(v) - m\{w(\mathbf{t})\}(v)]|W(\mathbf{t}) = w(\mathbf{t})\}.$$

Expressions for the maximum likelihood estimators of  $m\{w(\mathbf{t})\}(t)$  and  $\mathcal{K}\{w(\mathbf{t})\}(t, v)$  are obtained using the joint normality of  $\{Z(t), Z(v), W(\mathbf{T})\}$ , however, these estimators depend on the unknown basis functions  $\{\phi_k(\cdot)\}_{k \geq 1}$  and loading variances  $\{\lambda_k\}_{k \geq 1}$ . Let  $\widehat{m}_n^K\{w(\mathbf{t})\}(t)$  and  $\widehat{\mathcal{K}}_n^K\{w(\mathbf{t})\}(t, v)$  denote the estimators obtained by replacing  $\{\phi_k(\cdot)\}_{k \geq 1}$  and  $\{\lambda_k\}_{k \geq 1}$  with the functional principal components estimators  $\{\widehat{\phi}_k(\cdot)\}_{k=1}^K$  and  $\{\widehat{\lambda}_k\}_{k=1}^K$ , where  $K$  is a finite truncation. Let  $\widehat{G}_n^K\{\cdot|w(\mathbf{t})\}$  denote the distribution of Gaussian process with mean  $\widehat{m}_n^K\{w(\mathbf{t})\}(t)$  and covariance kernel  $\widehat{\mathcal{K}}_n^K\{w(\mathbf{t})\}(t, v)$ . The nonlinear functional Q-learning algorithm is

(NFQ1) Apply (FQ1) to construct estimators  $\widehat{\phi}_{nk}(\cdot)$  of  $\phi_k(\cdot)$ , and  $\widehat{\ell}_{nk}\{w(\mathbf{t})\}$  for  $k = 1, \dots, K$  where  $K$  is a finite truncation.

(NFQ2) Let  $\zeta \triangleq (\alpha^\top, \delta^\top, \eta^\top, \rho^\top)^\top$ , define

$$\widehat{Q}_n^K\{\mathbf{x}, w(\mathbf{t}), a; \zeta\} \triangleq \mathbf{x}^\top \alpha + \int f\{u(\cdot); \eta\} d\widehat{G}_n^K\{u(\cdot)|w(\mathbf{t})\} + a \left[ \mathbf{x}^\top \delta + \int g\{u(\cdot); \rho\} d\widehat{G}_n^K\{u(\cdot)|w(\mathbf{t})\} \right],$$

and use nonlinear least squares to compute  $\widehat{\zeta}_n = \arg \min_{\zeta} \mathbb{P}_n[Y - \widehat{Q}_n^K\{\mathbf{X}, W(\mathbf{T}), A; \zeta\}]^2$ .

(NFQ3) The estimated optimal decision rule is  $\widehat{\pi}_n^{\text{NFQ}}\{\mathbf{x}, w(\mathbf{t})\} = \arg \max_a \widehat{Q}_n^K\{\mathbf{x}, w(\mathbf{t}), a; \widehat{\zeta}_n\}$ .

The above algorithm, can be implemented using existing statistical software with the requisite integrals being evaluated using Monte Carlo methods. Implementations of the algorithms using the R programming language (<http://cran.r-project.org/>) are available in the supplemental material. Under conditions on the smoothness of  $f$  and  $g$ , consistency of the nonlinear functional Q-learning estimator follows from an argument similar to proof of Theorem 1.

**Remark 1.** One criticism of regression-based methods is that the form of the estimated optimal ITR is dictated by the form of the estimated Q-function. Thus, a parsimonious ITR requires a parsimonious Q-function, which may be prone to misspecification. On the other hand, a complex and expressive Q-function

may lead to a complex ITR that is difficult to communicate to domain experts. This has led increased interest in direct-search, also known as policy-search, algorithms that model the marginal mean outcome as a function on the space of ITRs and then to use the maximizer of this fitted model to estimate the optimal ITR; the appeal of these methods is that the maximizer can be restricted to belong to a prespecified class of ITRs that are interpretable or satisfy cost or logistical constraints. It is possible to use functional Q-learning to construct direct-search estimators with sparse longitudinal data. Let  $\Pi$  denote a class of ITRs of interest. It can be shown (Qian and Murphy 2011b) that  $\mathbb{E}Y^*(\pi) = \mathbb{E}[Q\{\mathbf{X}, W(\mathbf{T}), A\}1_{A=\pi\{\mathbf{X}, W(\mathbf{T})\}}/P\{A|\mathbf{X}, W(\mathbf{T})\}]$ , where  $P\{a|\mathbf{x}, w(\mathbf{t})\}$  denotes the propensity score  $P\{A = a|\mathbf{X} = \mathbf{x}, W(\mathbf{T}) = w(\mathbf{t})\}$ . In a randomized clinical trial the propensity score is known, in observational studies this must be estimated from data. Let  $\widehat{P}\{a|\mathbf{x}, w(\mathbf{t})\}$  denote an estimator of the propensity score and  $\widehat{Q}\{\mathbf{x}, w(\mathbf{t}), a\}$  an estimator of the Q-function constructed using functional Q-learning. An estimator of  $\mathbb{E}Y^*(\pi)$  is  $\widehat{V}(\pi) = \mathbb{P}_n[\widehat{Q}\{\mathbf{X}, W(\mathbf{T}), A\}1_{A=\pi\{\mathbf{X}, W(\mathbf{T})\}}/\widehat{P}\{A|\mathbf{X}, W(\mathbf{T})\}]$ . A policy-search estimator of the optimal ITR is  $\widehat{\pi}_{\text{PS}} = \arg \max_{\pi \in \Pi} \widehat{V}(\pi)$ . In this formulation, it is possible to increase the complexity of the Q-function without increasing the complexity of the estimated optimal ITR. For this reason, with policy-search estimators it is typical to use a nonparametric estimator of the Q-function. We discuss nonparametric estimators in the next remark.

**Remark 2.** The preceding estimation algorithm can be extended to handle more general classes of models for the conditional mean of the outcome given covariates and the latent functional process. For example, one might postulate an additive model of the form

$$\mathbb{E}\{Y|\mathbf{X} = \mathbf{x}, Z(\cdot) = z(\cdot), A = a\} = b(\mathbf{x}) + f\{z(\cdot)\} + a[c(\mathbf{x}) + g\{z(\cdot)\}],$$

where  $c, b \in \mathcal{C}$  are unknown functions,  $f, g \in \mathcal{F}$ , and  $\mathcal{C} \subseteq \mathbb{R}^{\mathbb{R}^p}$  and  $\mathcal{F} \subseteq \mathbb{R}^{L^2[0,1]}$  are (possibly infinitely dimensional) unknown classes of functions. Assuming this model, let  $\mathfrak{U} = (b, c, f, g)$  and define  $\widehat{G}_n^K$  as in the nonlinear functional Q-learning algorithm. The Q-function is

$$\widehat{Q}_n^K\{\mathbf{x}, w(\mathbf{t}), a; \mathfrak{U}\} = b(\mathbf{x}) + \int f\{u(\cdot)\} d\widehat{G}_n^K\{u(\cdot)|w(\mathbf{t})\} + a \left[ c(\mathbf{x}) + \int g\{u(\cdot)\} d\widehat{G}_n^K\{u(\cdot)|w(\mathbf{t})\} \right].$$

Hence,  $\mathfrak{U}$  can be estimated using penalized least squares

$$\widehat{\mathfrak{U}}_n^\tau = \arg \min_{\mathfrak{U}} \mathbb{P}_n[Y - \widehat{Q}_n^K\{\mathbf{X}, W(\mathbf{T}), A; \mathfrak{U}\}]^2 + \mathcal{P}_\tau(\mathfrak{U}),$$

where  $\mathcal{P}_\tau$  is a penalty function indexed by a vector of tuning parameters  $\lambda \in \mathbb{R}^q$ . The estimated optimal ITR  $\widehat{\pi}_n^{\text{NFQ}}\{\mathbf{x}, w(\mathbf{t})\} = \arg \max_a \widehat{Q}_n^K\{\mathbf{x}, w(\mathbf{t}), a; \widehat{\mathfrak{U}}_n^\tau\}$ .

As discussed previously, nonparametric functional Q-learning can be used in the context of policy-search algorithms. In addition, one can compare the estimated policy using parametric functional Q-learning with the estimated policy using nonparametric functional Q-learning as a means of diagnosing model misspecification. However, developing a formal test of the parametric functional Q-learning model

**Table 1.** Estimated mean outcome under linear generative models for functional Q-learning, Q-learning with ad hoc features, and random guessing.

Example	Sparsity	$\mathbb{E}Y^*(\pi^{\text{opt}})$	$\mathbb{E}Y^*(\hat{\pi}^{FQ})$	$\mathbb{E}Y^*(\hat{\pi}^{Q_{\text{last}}})$	$\mathbb{E}Y^*(\hat{\pi}^{Q_{\text{mean}}})$	$\mathbb{E}Y^*(\pi^{\text{Rand}})$
(S1)	Sparse	3.0	2.81	1.83	2.74	0.90
(S1)	Moderate	3.0	2.86	1.79	2.86	0.90
(S2)	Sparse	3.0	2.61	1.88	1.75	0.92
(S2)	Moderate	3.0	2.81	1.88	1.80	0.92
(S3)	Sparse	3.0	2.43	2.29	1.86	0.94
(S3)	Moderate	3.0	2.58	2.40	1.87	0.94

NOTE: Results are computed from 1000 datasets of size  $n = 250$ . Monte Carlo standard errors are in the second decimal place.

against a nonparametric alternative would require extending the asymptotic theory presented in the preceding section to handle nonparametric estimators; we leave this to future work.

#### 4. Empirical Study

In this section, we examine the finite sample performance of functional Q-learning. Performance of an estimated ITR,  $\hat{\pi}$ , is measured in terms of the average marginal mean outcome  $\mathbb{E}\{Y^*(\hat{\pi})\}$ . To form a baseline for comparison, we implemented standard Q-learning with linear models working models in  $\mathbf{X}$  and a fixed summary function  $s\{W(\mathbf{T})\}$ . We consider two potential summary functions, the last observed value in the longitudinal trajectory  $s_L\{W(\mathbf{T})\} = W(T_M)$ , and the mean  $s_M\{W(\mathbf{T})\} = \bar{W}(\mathbf{T}) = M^{-1} \sum_{j=1}^M W(T_j)$ . Let  $\hat{\pi}_n^{Q_{\text{last}}}$  and  $\hat{\pi}_n^{Q_{\text{mean}}}$  denote the estimated optimal ITRs using Q-learning with  $s_L\{W(\mathbf{T})\}$  and  $s_M\{W(\mathbf{T})\}$ . Reported simulation results are based on 1000 Monte Carlo replications.

##### 4.1. Linear Generative Models

To evaluate functional Q-learning with a linear model for  $\mathbb{E}\{Y|\mathbf{X}, Z(\cdot), A\}$ , we use the following class of generative models:  $\mathbf{X} \sim \text{Normal}\{0, \Omega_p(r)\}$ ,  $\{\Omega_p(r)\}_{i,j} = r^{|i-j|}$ ,  $\zeta \sim \text{Normal}(0, I_K)$ ,  $Z(t) = \sum_{k=1}^K \zeta_k b_k(t)$ ,  $t \in [0, 1]$ ,  $\tau \sim \text{Normal}(0, 1)$ ,  $Y = \mathbf{X}^T \alpha^* + \int_0^1 Z(t) \beta^*(t) dt + A\{\mathbf{X}^T \delta^* + \int_0^1 Z(t) \gamma^*(t) dt\} + \tau$ ,  $M \sim \text{Uniform}\{c, c+1, \dots, c+d\}$ , and  $\mathbf{T}|M = m \sim \text{Uniform}[0, 1]^m$ , where  $\{b_k(t)\}_{k=1}^K$  is a cubic B-spline basis with six equally spaced knots. Thus, this class of generative models is indexed by: the dimension  $p$ , coefficient vectors  $\alpha^*$  and  $\delta^*$ , number of basis functions  $K$ , autocorrelation  $r$ , sparsity parameters  $c$  and  $d$ , and the coefficient functions  $\beta^*(t)$  and  $\gamma^*(t)$ . Throughout, we fix  $p = 5$ ,  $\alpha^* \equiv -1$ ,  $\delta^* = (-0.1, 0, \dots, 0)$ ,  $r = 1/2$ , and  $K = 10$ . We consider two sparsity settings:  $c = 5$ ,  $d = 15$ , which we term sparse, and  $c = 15$ ,  $d = 10$ , which we term moderate. We consider three settings for the coefficient functions: (S1)  $\gamma^*(t) = \beta^*(t) \propto 1$ ; (S2)  $\gamma^*(t) = \beta^*(t) \propto 1_{t>0.90}$ ; and (S3)  $\gamma^*(t) = \beta^*(t) \propto t - 1/2$ . Setting (S1) corresponds to a simple average of  $Z(t)$  and thus Q-learning using the summary  $s_M\{W(\mathbf{T})\}$  may be expected to perform well, whereas Q-learning with the summary  $s_L\{W(\mathbf{T})\}$  may be expected to perform poorly; setting (S2) corresponds to the average of  $Z(t)$  only over the last 10% of the time domain and therefore Q-learning using the summary  $s_L\{W(\mathbf{T})\}$  may be expected to perform well, whereas Q-learning with the summary  $s_M\{W(\mathbf{T})\}$  may be expected to perform poorly; and setting (S3) corresponds to a contrast between early and late values of  $Z(t)$  and thus neither summary  $s_M\{W(\mathbf{T})\}$  nor  $s_L\{W(\mathbf{T})\}$  captures salient

features of the functional predictor and therefore Q-learning with either of these features may be expected to perform poorly.

In each setting, to control the effect size, we scale the coefficient functions so that  $\mathbb{E}|\int_0^1 Z(t) \beta^*(t) dt| = 2$ . We assume that  $W(t) = Z(t) + \epsilon(t)$  where  $\epsilon(t)$  is a standard Gaussian white-noise process. We use training sets of size  $n = 250$ , results with larger sample sizes were qualitatively similar and are omitted.

Table 1 shows the marginal mean outcomes for functional Q-learning, Q-learning, and a random regime that assigns treatment by random guessing with  $P(A = 1) = P(A = -1) = 1/2$ . Functional Q-learning performs favorably compared with Q-learning in all examples considered. The performance of Q-learning appears to depend critically on the choice of summary, for example, as anticipated, Q-learning with the summary  $s_M\{W(\mathbf{T})\}$  performs well in (S1) but poorly in (S2) and (S3), whereas Q-learning with the summary  $s_L\{W(\mathbf{T})\}$  performs well in (S3) but poorly in (S1) and (S2).

##### 4.2. Nonlinear Generative Models

To evaluate functional Q-learning with a nonlinear model for  $\mathbb{E}\{Y|\mathbf{X}, Z(\cdot), A\}$ , we use the same class of generative models as in Section 4.1 except that  $Y = \mathbf{X}^T \alpha^* + f\{Z(\cdot); \eta\} + A\{\mathbf{X}^T \delta^* + g\{Z(\cdot); \rho\} + \tau$ , where  $f\{z(\cdot); \eta\}$  and  $g\{z(\cdot); \rho\}$  are functionals of  $z(\cdot)$  indexed by unknown parameters  $\eta$  and  $\rho$ , and all other parameters are as in Section 4.1. In the settings, we consider here  $f \equiv g$  and  $f\{z(\cdot); \eta\} = \eta f_0\{z(\cdot)\}$  where  $\eta \in \mathbb{R}$ , and  $f_0\{z(\cdot)\}$  is a known functional of  $z(\cdot)$ . We consider three settings for  $f_0$ : (S4)  $f_0\{z(\cdot)\} = \int_0^1 tz(t) dt$ ; (S5)  $f_0\{z(\cdot)\} = \int_0^1 \{z'(t)\}^2 dt$ , where  $z'(t)$  denotes the derivative of  $z(\cdot)$  at  $t$ ; and (S6)  $f_0\{z(\cdot)\} = \sup_{t \in [0,1]} z(t) - \inf_{t \in [0,1]} z(t)$ . To control the treatment effect size, the constant  $\eta$  is chosen so that  $\eta \mathbb{E}|f_0\{z(\cdot)\}| = 1$ .

Table 2 shows the marginal mean outcomes under functional Q-learning, Q-learning, and random guessing. Functional Q-learning consistently obtains a higher marginal mean outcome than competing methods in all settings. This suggests that if there is strong scientific theory to guide the form of the Q-function, then functional Q-learning may perform well. However, the poor performance of Q-learning with the last observation illustrates that misspecification of the form of  $f\{z(\cdot); \eta\}$  could lead to poor performance of functional Q-learning, for example, by incorrectly choosing  $f\{z(\cdot); \eta\}$  to approximate  $z(t_m)$ .

#### 5. Application to STAR\*D

Sequenced Treatment Alternatives to Relieve Depression (STAR\*D, [www.star-d.org](http://www.star-d.org); Rush et al. 2003; Fava et al. 2008) is

**Table 2.** Estimated mean outcome under nonlinear generative models for functional Q-learning, Q-learning with ad hoc features, and random guessing.

Example	Sparsity	$\mathbb{E}Y^*(\pi^{\text{opt}})$	$\mathbb{E}Y^*(\hat{\pi}^{FQ})$	$\mathbb{E}Y^*(\hat{\pi}^{Q_{\text{last}}})$	$\mathbb{E}Y^*(\hat{\pi}^{Q_{\text{mean}}})$	$\mathbb{E}Y^*(\pi^{\text{Rand}})$
(S4)	Sparse	3.0	2.98	2.73	2.58	2.53
(S4)	Moderate	3.0	2.99	2.73	2.61	2.53
(S5)	Sparse	3.0	2.55	1.99	1.99	2.05
(S5)	Moderate	3.0	2.60	2.05	2.06	2.05
(S6)	Sparse	3.0	2.86	2.60	2.59	2.55
(S6)	Moderate	3.0	2.91	2.55	2.54	2.55

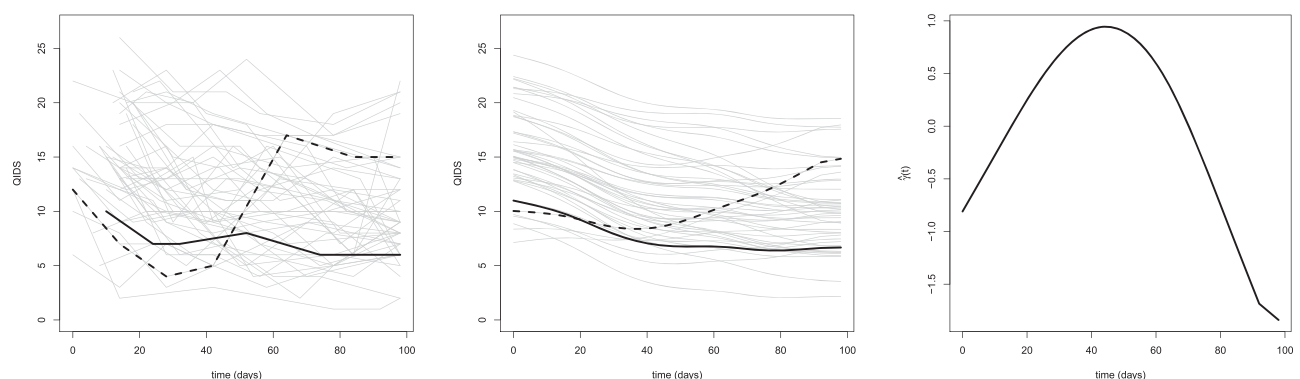
NOTE: Results are computed from 1000 datasets of  $n = 250$ . Monte Carlo standard errors are in the second decimal place.

a multi-stage, randomized trial of patients with major depressive disorder. We use data from the first stage of the trial to illustrate functional Q-learning (data can be obtained through [www.nimh.nih.gov](http://www.nimh.nih.gov)). For simplicity and to match the development of our methodology, we restrict attention to the first randomized stage of the trial and compare the two most prevalent treatments with complete outcome and covariate data: Citalopram augmented with Bupropion ( $A = 1$ ) and Citalopram augmented with Buspirone ( $A = -1$ ). Complete data are available on  $n = 192$  subjects. Subjects were observed for up to 14 weeks before the initial treatment, visit times are recorded in days from baseline; almost all (92%) of subjects in the complete data visited the clinic six times during this period. We use the prerandomization self-rated Quick Inventory of Depressive Symptomatology (QIDS) as our longitudinal predictor. The left panel of Figure 1 shows the raw QIDS score for 50 randomly sampled subjects, we have highlighted two subjects to emphasize heterogeneity in the observed patterns. Based on a survey of clinical papers and previous analyses on STAR\*D (see Pineau et al. 2007; Fava et al. 2008; Young et al. 2009; Chakraborty, Laber, and Zhao 2013; Schulte et al. 2014; Novick et al. 2015, and references therein), we include as covariates: mean Clinical Global Impression (CGI) score taken over prerandomization visits, sex, number of comorbidities at baseline, education level (indicator of some college or above), and the log number of major depressive episodes per year between diagnoses and enrollment. To be consistent with our paradigm of maximizing a desirable outcome, we use 27 minus the average QIDS score over the course of post-randomization follow-up in the first stage as our outcome. Fitting a linear working model, for example, Equation (1), to the observed data and using 50-fold cross-validation yields an estimated value of 19.35. The center panel of Figure 1 shows the smoothed longitudinal trajectories

**Table 3.** Estimated coefficients and 95% nonparametric bootstrap confidence intervals for nonfunctional predictors.

Term	Estimate	95% Bootstrap CI
Intercept	19.46	(17.24, 21.95)
CGI	-0.50	(-1.25, 0.26)
Sex	-0.52	(-1.17, 0.18)
Number Comorbidities	0.18	(-0.39, 0.80)
Depressive Episodes (Log)	-0.21	(-0.47, 0.11)
Education	0.45	(-0.22, 1.11)
Treatment	0.48	(-2.22, 2.50)
Treatment $\times$ CGI	-0.20	(-0.79, 0.64)
Treatment $\times$ Sex	0.23	(-0.47, 0.94)
Treatment $\times$ Number Comorbidities	-0.073	(-0.65, 0.53)
Treatment $\times$ Depressive Episodes (Log)	-0.031	(-0.31, 0.26)
Treatment $\times$ Education	-0.050	(-0.62, 0.72)

recovered from the functional principal components analysis and Table 3 shows the estimated coefficients for nonfunctional predictors. In contrast, using Q-learning and using the QIDS at the time of randomization in place of the QIDS longitudinal trajectory yields a cross-validated value of 16.69. Because the marginal mean outcome of an estimated optimal treatment regime is a nonsmooth functional of the underlying generative distribution standard methods for inference do not apply (van der Vaart 1991; Laber and Murphy 2011; Hirano and Porter 2012). To provide a rough sense of the uncertainty in the difference in value between functional and standard Q-learning, we calculated a Z-statistic using fifty differences in value computed across each fold of the cross-validation; the resultant Z-value was 1.88, so that functional Q-learning is nearly two (crudely estimated) standard errors above the mean using standard Q-learning. Thus, incorporating longitudinal information shows promising results in this limited example. Furthermore, the analysis also produces an estimate of  $\hat{\gamma}(\cdot)$ , which is shown in the right panel of Figure (1); the estimated coefficient function

**Figure 1.** Left: Observed self-rated QIDS scores for 50 randomly selected subjects. Center: Estimated trajectories using functional principal components. Right: Estimated coefficient function  $\hat{\gamma}(\cdot)$ .



weights recent observations most heavily and contrasts them with data in the first half of the study.

## 6. Discussion

We proposed functional Q-learning as means of incorporating sparse, irregularly spaced, noisy longitudinal data into the estimation of optimal ITRs. We proved that under regularity conditions, including correct specification of the outcome regression models, the estimated optimal ITR is consistent for the true optimal ITR. Furthermore, the proposed method requires less stringent causal assumptions than standard Q-learning used with an ad hoc summary of longitudinal information. The proposed methodology performed favorably compared with standard methods in simulation experiments.

The proposed methodology lays the groundwork for a number of important extensions. One such extension is to incorporate the proposed estimator into each step of multiple stage Q-learning (Schulte et al. 2014) to analyze data from sequential multiple assignment randomized studies or observational longitudinal data with multiple treatments. The proposed estimators could also be extended to incorporate multivariate longitudinal processes, for example, through additive regression models, and to discrete longitudinal processes, for example, through latent threshold models. Finally, as noted in Remark 1, the proposed methodology can be used to build direct-search algorithms to estimate the optimal ITR within a prespecified class of ITRs. We are currently pursuing a number of these extensions.

## Funding

Division of Mathematical Sciences: [1513579, 1557733, 1555141]; National Institutes of Health: [P01 CA142538]; U.S. Fish and Wildlife Service: [PR-W-F14AF0017].

## References

- Chakraborty, B., Laber, E. B., and Zhao, Y. (2013), "Inference for Optimal Dynamic Treatment Regimes Using an Adaptive m-Out-of-n Bootstrap Scheme," *Biometrics*, 69, 714–723. [1219,1220,1225]
- Chakraborty, B., and Moodie, E. E. (2013), "Statistical Reinforcement Learning," in *Statistical Methods for Dynamic Treatment Regimes*, New York: Springer, pp. 31–52. [1219,1220]
- Chakraborty, B., Murphy, S., and Strecher, V. (2010), "Inference for Non-Regular Parameters in Optimal Dynamic Treatment Regimes," *Statistical Methods in Medical Research*, 19, 317–343. [1219]
- Chakraborty, B., and Murphy, S. A. (2014), "Dynamic Treatment Regimes," *Annual Review of Statistics and Its Application*, 1, 447–464. [1219]
- Ciarleglio, A., Petkova, E., Ogden, R. T., and Tarpey, T. (2015), "Treatment Decisions Based on Scalar and Functional Baseline Covariates," *Biometrics*, 71, 884–894. [1220,1221]
- Ciarleglio, A., Petkova, E., Tarpey, T., and Ogden, R. T. (2016), "Flexible Functional Regression Methods for Estimating Individualized Treatment Rules," *Stat*, 5, 185–199. [1220]
- Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009), "Multilevel Functional Principal Component Analysis," *Annals of Applied Statistics*, 3, 458–488. [1221]
- Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002), *Analysis of Longitudinal Data*, Oxford: Oxford University Press. [1220]
- Fava, M., A John Rush, M., Alpert, J. E., Balasubramani, G., Wisniewski, S. R., Carmin, C. N., Biggs, M. M., Zisook, S., Leuchter, A., Howland, R., Warden, D., and Trivedi, M. H. (2008), "Difference in Treatment Outcome in Outpatients With Anxious Versus Nonanxious Depression: A STAR\* D Report," *American Journal of Psychiatry*, 165, 342–351. [1224]
- Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012), *Applied Longitudinal Analysis* (Vol. 998), New York: Wiley. [1220]
- Goldberg, Y., and Kosorok, M. R. (2012), "Q-Learning With Censored Data," *Annals of Statistics*, 40, 529. [1219]
- Hall, P., and Horowitz, J. L. (2007), "Methodology and Convergence Rates for Functional Linear Regression," *The Annals of Statistics*, 35, 70–91. [1222]
- Hall, P., and Hosseini-Nasab, M. (2006), "On Properties of Functional Principal Components Analysis," *Journal of the Royal Statistical Society, Series B*, 68, 109–126. [1222]
- Hamburg, M. A., and Collins, F. S. (2010), "The Path to Personalized Medicine," *New England Journal of Medicine*, 363, 301–304. [1219]
- Henderson, R., Ansell, P., and Alshibani, D. (2010a), "Regret-Regression for Optimal Dynamic Treatment Regimes," *Biometrics*, 66, 1192–1201. [1219]
- (2010b), "Regret-Regression for Optimal Dynamic Treatment Regimes," *Biometrics*, 66, 1192–1201. [1222]
- Hirano, K., and Porter, J. R. (2012), "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80, 1769–1790. [1225]
- Insel, T. R. (2009), "Translating Scientific Opportunity Into Public Health Impact: A Strategic Plan for Research on Mental Illness," *Archives of General Psychiatry*, 66, 128. [1219]
- Jiang, C.-R., and Wang, J.-L. (2010), "Covariate Adjusted Functional Principal Components Analysis for Longitudinal Data," *The Annals of Statistics*, 38, 1194–1226. [1221]
- Laber, E., Linn, K., and Stefanski, L. (2013a), "Interactive Q-Learning for Quantiles," *Journal of the American Statistical Association*, 112, 638–649. [1220]
- Laber, E., Lizotte, D., and Ferguson, B. (2014), "Set-Valued Dynamic Treatment Regimes for Competing Outcomes," *Biometrics*, 70, 53–61. [1219,1220]
- Laber, E., Qian, M., Lizotte, D. J., Pelham, W., and Murphy, S. (2013b), "Dynamic Treatment Regimes: Technical Challenges and Applications," *Electronic Journal of Statistics*, 8, 1225–1272. [1219,1220]
- Laber, E. B., and Murphy, S. A. (2011), "Adaptive Confidence Intervals for the Test Error in Classification," *Journal of the American Statistical Association*, 106, 904–913. [1225]
- Lai, T. L., and Robbins, H. (1985), "Asymptotically Efficient Adaptive Allocation Rules," *Advances in Applied Mathematics*, 6, 4–22. [1222]
- Laird, N. M., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974. [1220]
- Lauer, M. S., and Collins, F. S. (2010), "Using Science to Improve the Nation's Health System," *JAMA: The Journal of the American Medical Association*, 303, 2182–2183. [1219]
- Li, Y., and Hsing, T. (2010), "Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data," *The Annals of Statistics*, 38, 3321–3351. [1222]
- Lian, H. (2011), "Functional Partial Linear Model," *Journal of Nonparametric Statistics*, 23, 115–128. [1222]
- Lu, W., Zhang, H. H., and Zeng, D. (2011), "Variable Selection for Optimal Treatment Decision," *Statistical Methods in Medical Research*, 5, 493–504. [1219]
- McKeague, I. W., and Qian, M. (2011), "Sparse Functional Linear Regression With Applications to Personalized Medicine," in *Recent Advances in Functional Data Analysis and Related Topics*, New York: Springer, pp. 213–218. [1219]
- (2014), "Estimation of Treatment Policies Based on Functional Predictors," *Statistica Sinica*, 24, 1461–1485. [1220,1221]
- Moodie, E. E., Dean, N., and Sun, Y. R. (2013), "Q-Learning: Flexible Learning About Useful Utilities," *Statistics in Biosciences*, 6, 223–243. [1219]
- Moodie, E. E., and Richardson, T. S. (2010), "Estimating Optimal Dynamic Regimes: Correcting Bias Under the Null," *Scandinavian Journal of Statistics*, 37, 126–146. [1219]

- Murphy, S. A. (2003), "Optimal Dynamic Treatment Regimes" (with discussion), *Journal of the Royal Statistical Society, Series B*, 58, 331–366. [1219]
- Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W. E., Gnagy, B., Fabiano, G. A., Waxmonsky, J. G., Yu, J., and Murphy, S. A. (2012), "Q-Learning: A Data Analysis Method for Constructing Adaptive Interventions," *Psychological Methods*, 17, 478–494. [1220]
- Novick, D., Hong, J., Montgomery, W., Due nas, H., Gado, M., and Haro, J. M. (2015), "Predictors of Remission in the Treatment of Major Depressive Disorder: Real-World Evidence From a 6-Month Prospective Observational Study," *Neuropsychiatric Disease and Treatment*, 11, 197–205. [1225]
- Orellana, L., Rotnitzky, A., and Robins, J. (2010), "Dynamic Regime Marginal Structural Mean Models for Estimation of Optimal Dynamic Treatment Regimes, Part I: Main Content," *International Journal of Biostatistics*, 6, 8. [1219]
- Pineau, J., Bellemare, M. G., Rush, A. J., Ghizaru, A., and Murphy, S. A. (2007), "Constructing Evidence-Based Treatment Strategies Using Methods From Computer Science," *Drug and Alcohol Dependence*, 88, S52–S60. [1225]
- Piquette-Miller, M., and Grant, D. (2007), "The Art and Science of Personalized Medicine," *Clinical Pharmacology and Therapeutics*, 81, 311–315. [1219]
- Qian, M., and Murphy, S. A. (2011a), "Performance Guarantees for Individualized Treatment Rules," *Annals of Statistics*, 39, 1180. [1219,1222]
- (2011b), "Performance Guarantees for Individualized Treatment Rules," *Annals of Statistics*, 39, 1180–1210. [1223]
- Robins, J. (1986), "A New Approach to Causal Inference in Mortality Studies With a Sustained Exposure Period Application to Control of the Healthy Worker Survivor Effect," *Mathematical Modelling*, 7, 1393–1512. [1219]
- (1989), "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," *Health Service Research Methodology: A Focus on AIDS*, 113, 159. [1219]
- (1997), "Causal Inference From Complex Longitudinal Data," in *Latent Variable Modeling and Applications to Causality*, New York: Springer, pp. 69–117. [1219]
- (2004), "Optimal Structural Nested Models for Optimal Sequential Decisions," in *Proceedings of the Second Seattle Symposium in Biostatistics*, Springer, pp. 189–326. [1219]
- Robins, J. M., Hernan, M. A., and Brumback, B. (2000), "Marginal Structural Models and Causal Inference in Epidemiology," *Epidemiology*, 11, 550–560. [1220]
- Rubin, D. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34–58. [1220]
- Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., and Keller, M. B. (2003), "The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): A Psychometric Evaluation in Patients With Chronic Major Depression," *Biological Psychiatry*, 54, 573–583. [1220,1224]
- Schulte, P., Tsiatis, A., Laber, E., and Davidian, M. (2014), "Q- and A-Learning Methods for Estimating Optimal Dynamic Treatment Regimes," *Statistical Science*, 29, 640–661. [1219,1220,1225,1226]
- Shin, H., and Lee, M. H. (2012), "On Prediction Rate in Partial Functional Linear Regression," *Journal of Multivariate Analysis*, 103, 93–106. [1222]
- Shortreed, S. M., Laber, E., Lizotte, D. J., Stroup, T. S., Pineau, J., and Murphy, S. A. (2011), "Informing Sequential Clinical Decision-Making Through Reinforcement Learning: An Empirical Study," *Machine Learning*, 84, 109–136. [1220]
- Silverman, B., and Ramsay, J. (2005), *Functional Data Analysis*, New York: Springer. [1221]
- Splawa-Neyman, J., Dabrowska, D., and Speed, T. (1990), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Statistical Science*, 5, 465–472. [1220]
- Staicu, A.-M., Li, Y., Crainiceanu, C., and Ruppert, D. (2014), "Likelihood Ratio Tests for the Mean Structure of Correlated Functional Processes," *Scandinavian Journal of Statistics*, 41, 932–949. [1222]
- Topol, E. J. (2012), *The Creative Destruction of Medicine*, New York: Basic Books. [1219]
- van der Vaart, A. (1991), "On Differentiable Functionals," *The Annals of Statistics*, 19, 178–204. [1225]
- Wittchen, H.-U., Lieb, R., Pfister, H., and Schuster, P. (2000), "The Waxing and Waning of Mental Disorders: Evaluating the Stability of Syndromes of Mental Disorders in the Population," *Comprehensive Psychiatry*, 41, 122–132. [1222]
- Yao, F., Müller, H.-G., and Wang, J.-L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590. [1221,1222]
- Young, E. A., Kornstein, S. G., Marcus, S. M., Harvey, A. T., Warden, D., Wisniewski, S. R., Balasubramani, G., Fava, M., Trivedi, M. H., and Rush, A. J. (2009), "Sex Differences in Response to Citalopram: A STAR D Report," *Journal of Psychiatric Research*, 43, 503–511. [1225]
- Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M., and Laber, E. (2012a), "Estimating Optimal Treatment Regimes From a Classification Perspective," *Stat*, 1, 103–114. [1219,1221]
- Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012b), "A Robust Method for Estimating Optimal Treatment Regimes," *Biometrics*, 68, 1010–1018. [1219,1221]
- (2013), "Robust Estimation of Optimal Dynamic Treatment Regimes for Sequential Treatment Decisions," *Biometrika*, 100, 681–694. [1219,1221]
- Zhang, X., and Wang, J.-L. (2016), "From Sparse to Dense Functional Data and Beyond," *The Annals of Statistics*, 44, 2281–2321. [1222]
- Zhao, Y., Zeng, D., Socinski, M. A., and Kosorok, M. R. (2011), "Reinforcement Learning Strategies for Clinical Trials in Nonsmall Cell Lung Cancer," *Biometrics*, 67, 1422–1433. [1219]
- Zhao, Y. Q., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012), "Estimating Individualized Treatment Rules Using Outcome Weighted Learning," *Journal of the American Statistical Association*, 107, 1106–1118. [1219,1221]