



HHS Public Access

Author manuscript

J Comput Graph Stat. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

J Comput Graph Stat. 2022 ; 31(2): 553–562. doi:10.1080/10618600.2021.1923521.

A single-index model with a surface-link for optimizing individualized dose rules

Hyung Park,

Division of Biostatistics, Department of Population Health, New York University

Eva Petkova,

Division of Biostatistics, Department of Population Health, New York University

Thaddeus Tarpey,

Division of Biostatistics, Department of Population Health, New York University

R. Todd Ogden

Department of Biostatistics, Columbia University

Abstract

This paper focuses on the problem of modeling and estimating interaction effects between covariates and a continuous treatment variable on an outcome, using a single-index regression. The primary motivation is to estimate an optimal individualized dose rule and individualized treatment effects. To model possibly nonlinear interaction effects between patients' covariates and a continuous treatment variable, we employ a two-dimensional penalized spline regression on an index-treatment domain, where the index is defined as a linear projection of the covariates. The method is illustrated using two applications as well as simulation experiments. A unique contribution of this work is in the parsimonious (single-index) parametrization specifically defined for the interaction effect term.

Keywords

Single-index models; individualized dose rules; tensor product P-splines; heterogeneous dose effects

1 Introduction

In precision medicine, a primary goal is to characterize individuals' heterogeneity in treatment responses so that individual-specific treatment decisions can be made (Murphy, 2003; Robins, 2004). Most work on developing methods for individualized treatment decisions has focused on a finite number of treatment options. The focus of this paper is to develop individualized treatment decision methodology in the realm of a continuous treatment. Specifically, we consider a semiparametric regression approach for developing optimal individualized dosing rules based on baseline patient characteristics. Often in clinical practice, the maximum dose that a patient can tolerate is the most effective one, however, there are situations where this is not the case. In the example section, we present a study of warfarin (an anticoagulant), where too high doses lead to severe bleeding and thus the highest dose is not the optimal dose. In finding the optimal dose, there is an essential

non-monotone and nonlinear relationship that needs to be accounted for. A similar case is with insulin for controlling blood glucose levels.

To establish notation, let $X = (X_1, \dots, X_p)^\top \in \mathcal{X}$ be the set of baseline covariates, $Y \in \mathbb{R}$ be the outcome variable, and $A \in \mathcal{A}$ denote the dose. Let $Y^*(a)$ be the potential outcome when a dose level $a \in \mathcal{A}$ is given. Throughout, we assume: 1) consistency, i.e., $Y = \int_{\mathcal{A}} \delta(A = a) Y^*(a) da$, where $\delta(\cdot)$ is the Dirac delta function; 2) no unmeasured confoundedness, i.e., $\{Y^*(a), a \in \mathcal{A}\}$ is conditionally independent of A given X ; 3) positivity, i.e., $p(A = a | X = x) \geq c$, for all $a \in \mathcal{A}$, $x \in \mathcal{X}$, for some $c > 0$ (where $p(a/x)$ is the conditional density of $A = a$ given $X = x$), which are standard assumptions adopted in the causal inference literature (Gill and Robins, 2001). Without loss of generality, we assume that a larger value of the outcome Y is better. The goal is then to find an optimal individualized dose rule $f: \mathcal{X} \mapsto \mathcal{A}$ such that for a patient with covariate X , the dose assignment $A = f(X)$ maximizes the expected response, the so-called value function, $\mathcal{V}(f) = \mathbb{E}[Y^*(f(X))]$, that is,

$$\mathcal{V}(f) = \mathbb{E}[\mathbb{E}[Y | A = f(X), X]] \quad (1)$$

which holds and can be empirically approximated under the above three assumptions. In settings in which the treatment can be administered at continuous doses (i.e., when \mathcal{A} is an interval), Chen *et al.* (2016) proposed to optimize the individualized dosing rule f by maximizing a local approximation of the value function (1), optimized under the framework of outcome weighted learning (Zhao *et al.*, 2012). Laber and Zhao (2015) proposed a tree-based decision rule for treatment assignment with minimal impurity dividing patients into subgroups with different discretized doses. Kallus and Zhou (2018) developed an inverse propensity weighted estimator of (1) for continuous treatments with the doubly robust property (Dudík *et al.*, 2014), and recently, kernel assisted learning with linear dimension reduction (Zhou *et al.*, 2020; Zhu *et al.*, 2020) for direct optimization of (1) have been developed. However, implementation of these approaches for general exponential family distributions is not straightforward and has not been accomplished. In this paper, we consider a regression-based approach to optimizing f that uses a semiparametric regression model for $\mathbb{E}[Y | A, X]$. There is also extensive literature on multi-armed bandit (e.g., Lattimore and Szepesvari, 2019) problems in the context of reinforcement learning (e.g., Kaelbling *et al.*, 1996), incorporating context (i.e., feature $X \in \mathcal{X}$) (see, e.g., Lu *et al.* (2010); Perchet and Rigollet (2013); Slivkins (2014); Jun *et al.* (2017); Li *et al.* (2017); Kveton *et al.* (2020); Chen *et al.* (2020)) for making a sequential decision that minimizes the notion of cumulative regret, with relatively fewer works on contextual bandits with continuous actions (see, e.g., Krishnamurthy *et al.*, 2020; Kleinberg *et al.*, 2019; Majzoubi *et al.*, 2020). However, these works are focused on optimizing online performance (addressing the exploration issue) and are considerably different from personalized dose-finding focused on a single stage with feature X . Kennedy *et al.* (2017) considered a method for estimating the average dose effect allowing for flexible doubly robust covariate adjustment, but the method is not intended for optimal dose finding for individual patients. For multi-stage personalized dose-finding, Rich *et al.* (2014) proposed adaptive strategies, and more recently Schulz and Moodie (2020) developed a doubly robust estimation approach based on a linear

model. All these approaches are limited by the stringent linear model assumptions for the heterogeneous dose effects.

While the methods of directly optimizing the value function (1), including the outcome weighted learning of Chen *et al.* (2016) and the tree-based method of Laber and Zhao (2015), are highly appealing, the proposed semi-parametric regression modeling has the advantage of being easy to implement and readily generalizable to an exponential family response.

It is straightforward to see that, given X , the optimal dose $f_{\text{opt}}(X)$ (i.e., that which maximizes the value function (1)) is

$$f_{\text{opt}}(X) = \operatorname{argmax}_{a \in \mathcal{A}} m(a, X), \quad (2)$$

where $m(a, X) = \mathbb{E}[Y | A = a, X]$. If we estimate $m(a, X)$ with $\hat{m}(a, X)$, then the optimal rule f_{opt} in (2) can be approximated as

$$\hat{f}(X) = \operatorname{argmax}_{a \in A} \hat{m}(a, X). \quad (3)$$

Methodologies for optimizing individualized treatment rules f in the precision medicine literature are developed predominantly for the cases in which the treatment variable A is binary or discrete-valued. Regression-based methodologies first estimate the treatment a -specific mean response functions $m(a, X)$ and then obtain a treatment decision rule, i.e., the left-hand side of (3) (e.g., see Qian and Murphy, 2011; Zhang *et al.*, 2012; Gunter *et al.*, 2011; Lu *et al.*, 2013; Park *et al.*, 2020) given X . In particular, Qian and Murphy (2011) show that the optimal individualized treatment rule (2) depends only on the interaction between treatment A and covariates X , and not on the main effects of X in the mean models $m(a, X)$. For regression-based methodologies, a successful estimation of the function f_{opt} in (2) boils down to efficiently estimating the A -by- X interaction effects on the treatment response. In this paper, we consider a semi-parametric model that is useful for estimating such interactions in the case where A is a continuous dose variable.

2 Models

Our goal is to provide an interpretable and flexible approach to modeling the A -by- X interaction effects on Y . To achieve this goal, we consider the following additive single-index model:

$$\mathbb{E}[Y | X, A] = \mu(X) + g(\beta^\top X, A) \quad (4)$$

where $\mu(X)$ represents an unspecified main effect of X , and $g(\beta^\top X, A)$ models the A -by- X interaction effects. Here, $g(\cdot, \cdot)$ is an unspecified smooth two-dimensional *surface* link function of the variable A and a single index $\beta^\top X$. We shall call model (4) a *single-index model with a surface-link* (SIMSL). We restrict

$\beta \in \Theta := \{\beta = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p : \|\beta\|_2 = 1, \beta_1 > 0\}$, as β in (4) is only identifiable up to a scale constant without further constraint, due to the unspecified nature of g .

Without loss of generality, we assume $\mathbb{E}[\mu(X)] = 0$ and $\mathbb{E}[g(\beta^\top X, A)] = 0$ (where the expectation is with respect to X and A), i.e., each of the additive components in model (4) has mean 0, and that these components have finite variances, as are typically assumed in generalized additive models (GAM; Hastie and Tibshirani, 1999). That is, let \mathcal{H}_1 and $\mathcal{H}_2^{(\beta)}$ (for a fixed $\beta \in \Theta$) denote the L^2 spaces of measurable functions $\mu(X)$ on X and measurable functions $g(\beta^\top X, A)$ on $(\beta^\top X, A)$, respectively, and we assume $\mu \in \mathcal{H}_1$ and $g \in \mathcal{H}_2^{(\beta)}$.

Due to the unspecified nature of μ and g (and that X is involved in both μ and g), model (4) is not identifiable without further constraints. We will constrain the smooth function $g \in \mathcal{H}_2^{(\beta)}$ to satisfy:

$$\mathbb{E}[g(\beta^\top X, A) | X] = 0, \text{ a. s. } (X) \ X \in \mathcal{X}, \beta \in \mathbb{R}^p, \tag{5}$$

which acts as an identifiability condition of model (4).

Applying the constraint (5) to the function g in (4) essentially reparametrizes the model (4), by replacing $g(\beta^\top X, A)$ with $g_0(\beta^\top X, A) = g(\beta^\top X, A) - \mathbb{E}[g(\beta^\top X, A) | X]$, and $\mu(X)$ with $\mu_0(X) = \mu(X) + \mathbb{E}[g(\beta^\top X, A) | X]$. This yields an equivalent model of (4), $\mathbb{E}[Y | X, A] = \mu_0(X) + g_0(\beta^\top X, A)$, where the term $g_0(\beta^\top X, A)$ satisfies the identifiability condition (5). Since any arbitrary (μ, g) in (4) can be rearranged to give such reparametrized components (μ_0, g_0) , without loss of generality, we will represent (μ_0, g_0) as (μ, g) subject to (5).

Under the SIMSL (4) (subject to constraint (5)), the optimal individualized dose rule, f_{opt} , is specified as: $f_{\text{opt}}(X) = \underset{a \in \mathcal{A}}{\operatorname{argmax}} g(\beta^\top X, a)$, which does not involve the component μ . Therefore, in terms of estimating f_{opt} in (2), our modeling focus is on estimating (g, β) in (4).

Using the constrained least squares framework, the right-hand side of (4), subject to constraint (5), can be optimized by solving:

$$\begin{aligned} (\mu^*, g^*, \beta^*) = & \underset{\mu \in \mathcal{H}_1, g \in \mathcal{H}_2^{(\beta)}, \beta \in \Theta}{\operatorname{argmin}} \mathbb{E}[(Y - \mu(X) - g(\beta^\top X, A))^2] \\ & \text{subject to } \mathbb{E}[g(\beta^\top X, A) | X] = 0. \end{aligned} \tag{6}$$

Constraint (5) ensures that $\mathbb{E}[\mu(X)g(\beta^\top X, A)] = \mathbb{E}[\mu(X)\mathbb{E}[g(\beta^\top X, A) | X]] = 0$ (in which we apply the iterated expectation rule to condition on X), which implies the orthogonality,

$$\mu(X) \perp g(\beta^T X, A), \tag{7}$$

in L^2 . The orthogonality (7) implies that the optimization for μ^* and that for (g^*, β^*) on the left-hand side of (6) can be performed separately, without iterating between the two optimizations. Specifically, we can solve for the X main effect component:

$$\mu^* = \operatorname{argmin}_{\mu \in \mathcal{H}_1} \mathbb{E}[(Y - \mu(X))^2], \tag{8}$$

and separately solve for the A -by- X interaction effect component:

$$\begin{aligned} (g^*, \beta^*) = & \operatorname{argmin}_{g \in \mathcal{H}_2^{(\sigma)}, \beta \in \Theta} \mathbb{E}[(Y - g(\beta^T X, A))^2] \\ & \text{subject to } \mathbb{E}[g(\beta^T X, A) | X] = 0, \end{aligned} \tag{9}$$

without alternating between the two optimization procedures. We can fit f_{opt} by:

$$\hat{f}(X) = \operatorname{argmax}_{a \in \mathcal{A}} \hat{g}^*(\hat{\beta}^{*T} X, a), \text{ where } (\hat{g}^*, \hat{\beta}^*) \text{ denotes an estimate of } (g^*, \beta^*) \text{ in (9).}$$

This approach using optimization (9) is appealing, because, due to orthogonality (7), misspecification of the functional form for μ in (6) does not affect specification of (g^*, β^*) on the left-hand side of (9). If primary interest is in estimating f_{opt} in (2), then utilizing \hat{f} based on optimization (9) circumvents the need to estimate the term μ in (4), obviating the need to specify its form and thus avoiding the issue of model misspecification on the X main effect. The equivalence between (g^*, β^*) on the left-hand side of (9) and (g, β) in (4) is given in Proposition 1 in Section 4 (in the context where Y follows an exponential family response).

Since our primary focus is on estimating f_{opt} , we focus on solving (9). However, we also note that modeling the μ term in (4) can generally improve the performance of the estimator of (g, β) , as discussed in Supplementary Materials Section C.1. For each fixed β , the term $g(\beta^T X, A)$ depends the covariates $X \in \mathcal{X}$ only through the 1-dimensional projection $\beta^T X$. Therefore, for each fixed β , the distribution of $g(\beta^T X, A) | X$ is equal to that for $g(\beta^T X, A) | \beta^T X$, which implies $\mathbb{E}[g(\beta X, A) | X] = \mathbb{E}[g(\beta^T X, A) | \beta^T X]$, for each fixed β . Then, for each fixed $\beta \in \Theta$, the following constraint on $g \in \mathcal{H}_2^{(\beta)}$,

$$\mathbb{E}[g(\beta^T X, A) | \beta^T X] = 0, \quad X \in \mathcal{X}, \tag{10}$$

is a sufficient condition for the orthogonality constraint (5). Thus, for each fixed β , the constraint (5) can be simplified to (10). The following iterative procedure will be used to solve (9):

1. For fixed β , optimize the smooth $g(\cdot, \cdot)$ by solving:

$$\operatorname{argmin}_{g \in H_2^{(\theta)}} \mathbb{E} \left[\left(Y - g(\beta^\top X, A) \right)^2 \right], \quad (11)$$

subject to the constraint (10).

2. For fixed g , optimize the coefficient $\beta \in \Theta$ by minimizing the squared error criterion of (11).
3. Iterate steps (1) and (2) until convergence with $\beta \in \Theta$.

The data version of optimizing (g, β) can be derived as an empirical counterpart of the iterative procedure given above. Details on implementing this algorithm are given below.

3 Estimation

3.1 Representation of link surface

Suppose we have observed data (Y_i, A_i, X_i) ($i = 1, \dots, n$). For each candidate vector $\beta \in \Theta$, let

$$u_i = u_i^{(\beta)} = \beta^\top X_i \quad (i = 1, \dots, n),$$

where (on the left-hand side), for the notational simplicity, we suppress the dependence of the linear predictor $u^{(\beta)} \in \mathbb{R}$ on the candidate vector β .

Eilers and Marx (2003) have used tensor products of B -splines (de Boor, 2001) to represent two-dimensional surfaces, which they termed tensor product P -splines, with separate difference penalties applied to the coefficients of the B -splines along the covariate axes. Although alternative nonparametric methods could also be used to estimate the smooth function $g \in \mathcal{H}_2^{(\beta)}$ given each coefficient vector β in model (4), in this paper we focus on one smoother, the tensor-product P -splines, for the ease of presentation.

Specifically, for each $u = \beta^\top X$, to represent the 2-dimensional function $g(u, A)$ in (11), we consider the tensor product of the two sets of univariate cubic B -spline basis functions, say B and \check{B} , with N (and \check{N}) B -spline knots for the basis functions that are placed along the u (and A) axis. The number of knots N (and \check{N}) is chosen to be large, i.e., to allow the surface much flexibility. Associated with the basis representation defined by the marginal basis function B (resp., \check{B}) is an $N \times N$ (resp., $\check{N} \times \check{N}$) roughness penalty matrix, which we denote by \mathbb{P} (and $\check{\mathbb{P}}$). The penalty matrix \mathbb{P} (and $\check{\mathbb{P}}$) can be easily constructed, for example, based on a second-order difference matrix (e.g, see Eilers and Marx (2003)).

For each fixed $u_i = \beta^\top X_i$ ($i = 1, \dots, n$), let us write the $n \times N$ (and $n \times \check{N}$) B -spline evaluation matrix \mathbf{B} (and $\check{\mathbf{B}}$), in which its i th row is $\mathbf{B}_i = B(u_i)^\top$ (and $\check{\mathbf{B}}_i = \check{B}(A_i)^\top$). Given a knot grid, a flexible surface can be approximated (Marx, 2015) at n points (u_i, A_i) ($i = 1, \dots, n$):

$$g(u_i, A_i) = \sum_{r=1}^N \sum_{s=1}^{\tilde{N}} B_r(u_i) \check{B}_s(A_i) \gamma_{rs} = (\mathbf{B}_i \otimes \check{\mathbf{B}}_i) \boldsymbol{\theta} \quad (i = 1, \dots, n), \quad (12)$$

where the vector $\boldsymbol{\theta} = (\gamma_{11}, \dots, \gamma_{1\tilde{N}}; \dots; \gamma_{N1}, \dots, \gamma_{N\tilde{N}})^\top \in \mathbb{R}^{N\tilde{N}}$ corresponds to an unknown (vectorized) coefficient vector of the tensor product representation of g , and \otimes represents the usual Kronecker product. Equation (12) can be compactly written as:

$$\text{vec}\{g(u_i, A_i)\} = g(u_{n \times 1}, A_{n \times 1}) = \mathbf{D}\boldsymbol{\theta}, \quad (13)$$

where

$$\mathbf{D} = \mathbf{B} \square \mathbf{B} = (\mathbf{B} \otimes \mathbf{I}_N^\top) \odot (\mathbf{I}_N^\top \otimes \mathbf{B}), \quad (14)$$

in which the symbol \odot denotes element-wise multiplication of matrices. In Wood (2017), the symbol \square in (14) is called the row-wise Kronecker product, which results in a $n \times N\tilde{N}$ tensor product design matrix \mathbf{D} from the two marginal design matrices \mathbf{B} and $\check{\mathbf{B}}$.

Similarly, the roughness penalty matrices associated with the tensor product representation (12) can be constructed from the roughness penalty matrices \mathbb{P} and $\check{\mathbb{P}}$ associated with the univariate (marginal) basis matrices \mathbf{B} and $\check{\mathbf{B}}$, and are given by $\mathbf{P} = \mathbb{P} \otimes \mathbf{I}_{\tilde{N}}$ and $\check{\mathbf{P}} = \mathbf{I}_N \otimes \check{\mathbb{P}}$, for the axis directions u and A , respectively. Here, \mathbf{I} denotes the identity matrix, and both \mathbf{P} and $\check{\mathbf{P}}$ are square matrices with dimension $N\tilde{N}$.

We now need to impose the constraint (10) on the 2-dimensional smooth function g under the tensor product representation (13). For each fixed $\boldsymbol{\beta}$, the constraint (10) on g amounts to excluding the main effect of $u = \boldsymbol{\beta}^\top X$ from the function g . We deal with this by a reparametrization of the representation (13) for g . Consider the following sum-to-zero (over the n observed values) constraint for the marginal function of A :

$$\mathbf{1}^\top \check{\mathbf{B}} \check{\boldsymbol{\gamma}} = 0, \quad (15)$$

for any arbitrary $\check{\boldsymbol{\gamma}} \in \mathbb{R}^{\tilde{N}}$, where $\mathbf{1}$ is a length n vector of 1's. With constraint (15), the linear smoother associated with the basis matrix $\check{\mathbf{B}}$ cannot reproduce constant functions (Hastie and Tibshirani, 1999). That is, the linear constraint (15) removes the span of constant functions from the span of the marginal basis matrix $\check{\mathbf{B}}$ associated with A . Constraint (15) results in a tensor product basis matrix, $\mathbf{D} = \mathbf{B} \square \check{\mathbf{B}}$ in (13), that will not include the main effect of u that results from the product of the marginal basis matrix \mathbf{B} with the constant function in the span of the other marginal basis matrix $\check{\mathbf{B}}$. Therefore, the resultant fit, under representation (13) (subject to (15)) of the smooth function g , excludes the main effect of u . See Section 5.6 of Wood (2017) for some more details.

We impose the linear constraint (15) on the matrix $\check{\mathbf{B}}$, and consequently, the resulting basis matrix \mathbf{D} of representation of g in (13) becomes independent of the basis associated with the main effect of u . Imposition of such a linear constraint (15) on a basis matrix

is routine. The key is to find an (orthogonal) basis for the null space of the constraint (15), and then absorb the constraint into the basis construction (14). To be specific, we can create a $\tilde{N} \times (\tilde{N} - 1)$ matrix, which we denote as \mathbf{Z} , such that, given any arbitrary coefficient vector $\check{\gamma}_0 \in \mathbb{R}^{\tilde{N}-1}$, if we set $\check{\gamma} = \mathbf{Z}\check{\gamma}_0$, then we have $\mathbf{I}^\top \check{\mathbf{B}}\check{\gamma} = 0$, automatically satisfying the constraint (15). Such a matrix \mathbf{Z} is constructed using a QR decomposition of $\check{\mathbf{B}}^\top \mathbf{I}$. Then we can reparametrize the marginal function of A by setting its model matrix to $\check{\check{\mathbf{B}}} \leftarrow \check{\mathbf{B}}\mathbf{Z}$ (and its penalty matrix to $\check{\check{\mathbf{P}}} \leftarrow \mathbf{Z}^\top \check{\mathbf{P}}\mathbf{Z}$). From this point forward, for notational simplicity, we redefine the matrix $\check{\mathbf{B}}$ (and $\check{\mathbf{P}}$) to be this reparameterized, constrained marginal basis matrix (and the reparameterized constrained penalty matrix).

This sum-to-zero reparametrization of the marginal basis matrix $\check{\mathbf{B}}$ of A to satisfy (15) creates a term $\text{vec}\{g(u_i, A_i)\} \in \mathbb{R}^n$ in (13) that specifies such a pure A -by- X interaction (plus the A main effect) component, that is also orthogonal to the X main effect. In Wood (2006), this reparameterization approach is used to create an analysis of variance (ANOVA) decomposition of a smooth function of several variables. In this paper we use this same reparameterization to orthogonalize the interaction effect component $g(\beta^\top X, A)$ from the main effect, and to allow an unspecified/misspecified main effect for X in the estimation of the SIMSL (4). Provided that the orthogonality constraint (i.e., (15)) issue is addressed, the interaction effect term $g(\beta^\top X, A)$ of model (4), for each fixed β , can be represented using penalized regression splines and estimated based on penalized least squares, which we describe next.

3.2 Estimation algorithm

We define the criterion function for estimating (g, β) in the SIMSL (4):

$$\begin{aligned} Q(\theta, \beta) &= \|Y_{n \times 1} - g(\mathbf{X}\beta, A_{n \times 1})\|^2 + \lambda \|\mathbf{P}\theta\|^2 + \check{\lambda} \|\check{\mathbf{P}}\theta\|^2 \\ &= \|Y_{n \times 1} - \mathbf{D}\theta\|^2 + \lambda \|\mathbf{P}\theta\|^2 + \check{\lambda} \|\check{\mathbf{P}}\theta\|^2 \end{aligned} \tag{16}$$

subject to the constraint that the function $g(\cdot, \cdot)$ empirically satisfies (5). In (16), \mathbf{X} is a $n \times p$ matrix whose i th row is X_i^\top . Since both θ and β are unknown in (16), estimation of θ and β is conducted iteratively. We describe below the estimation procedure.

1. For a fixed estimate of β (that defines the linear predictor u), minimize the following criterion function over $\theta \in \mathbb{R}^{N\tilde{N}}$,

$$\|Y_{n \times 1} - \mathbf{D}\theta\|^2 + \lambda \|\mathbf{P}\theta\|^2 + \check{\lambda} \|\check{\mathbf{P}}\theta\|^2, \tag{17}$$

where \mathbf{D} is given by (14). Given tuning parameters $(\lambda, \check{\lambda})$, the minimizer θ of (17) is:

$$\hat{\theta} = \left(\mathbf{D}^\top \mathbf{D} + \lambda \mathbf{P}^\top \mathbf{P} + \check{\lambda} \check{\mathbf{P}}^\top \check{\mathbf{P}} \right)^{-1} \mathbf{D}^\top Y_{n \times 1}.$$

2. For a fixed estimate of the surface g (i.e., given Θ), perform a first-order Taylor approximation of $g(\mathbf{X}\beta, A_{n \times 1})$ in (16) with respect to β , around the current estimate, say, $\tilde{\beta} \in \Theta$,

$$g(\mathbf{X}\beta, A_{n \times 1}) \approx g(\mathbf{X}\tilde{\beta}, A_{n \times 1}) + \text{diag}\{\dot{g}_{\partial_1}(\mathbf{X}\tilde{\beta}, A_{n \times 1})\}\mathbf{X}(\beta - \tilde{\beta}), \tag{18}$$

where $\dot{g}_{\partial_1}(u, a)$ denotes the partial first derivative of $g(u, a)$ with respect to u , i.e., $\frac{\partial g(u, a)}{\partial u}$. Utilizing (18), we approximate the quadratic loss in (16) by (as a function of β given Θ):

$$\begin{aligned} & \|Y_{n \times 1} - g(\mathbf{X}\tilde{\beta}, A_{n \times 1}) - \text{diag}\{\dot{g}_{\partial_1}(\mathbf{X}\tilde{\beta}, A_{n \times 1})\}\mathbf{X}(\beta - \tilde{\beta})\|^2 \\ &= \|Y_{n \times 1} - g(\mathbf{X}\tilde{\beta}, A_{n \times 1}) + \text{diag}\{\dot{g}_{\partial_1}(\mathbf{X}\tilde{\beta}, A_{n \times 1})\}\mathbf{X}\tilde{\beta} - \text{diag}\{\dot{g}_{\partial_1}(\mathbf{X}\tilde{\beta}, A_{n \times 1})\}\mathbf{X}\beta\|^2 \\ &= \|Y_{n \times 1}^* - \mathbf{X}^*\beta\|^2, \end{aligned} \tag{19}$$

where $Y_{n \times 1}^* = Y_{n \times 1} - g(\mathbf{X}\tilde{\beta}, A_{n \times 1}) + \text{diag}\{\dot{g}_{\partial_1}(\mathbf{X}\tilde{\beta}, A_{n \times 1})\}\mathbf{X}\tilde{\beta}$, and $\mathbf{X}^* = \text{diag}\{\dot{g}_{\partial_1}(\mathbf{X}\tilde{\beta}, A_{n \times 1})\}\mathbf{X}$. The minimizer $\hat{\beta}$ of (19) is:

$$\hat{\beta} = (\mathbf{X}^{* \top} \mathbf{X}^*)^{-1} \mathbf{X}^{* \top} Y_{n \times 1}^*. \tag{20}$$

Then we scale $\hat{\beta}$ to have unit L^2 norm, i.e. $\hat{\beta}/\|\hat{\beta}\|$, and enforce a positive first element to restrict the estimate of β to be in Θ .

These two steps can be iterated until convergence to obtain an estimate of (g^*, β^*) in (9), which we denote as $(\hat{g}^*, \hat{\beta}^*)$. For Step 1, the tuning parameters $(\lambda, \check{\lambda})$ can be automatically selected, for example, by the generalized cross-validation (GCV) or the restricted maximum likelihood (REML) methods. In this paper, we use REML for the simulation examples and the applications.

Lastly, for model hierarchy, it is common practice to include all lower order effects of variables if there are higher-order interaction terms including that set of variables. Once convergence of the estimate $\hat{\beta}$ is reached in the above algorithm and the single-index $\beta^\top X$ in the term $g(\beta^\top X, A)$ of model (4) is estimated, we recommend fitting one final (unconstrained) smooth function g of A and $\hat{\beta}^\top X$, without enforcing the constraint (15) on g . Given the final estimate of β , the unconstrained final surface-link $g(\cdot, \cdot)$ retains the main effect of $\beta^\top X$ and preserves model hierarchy.

4 Generalized single-index models for optimizing dose rules

The proposed approach to optimizing the heterogeneous dose effect (i.e., the X -by- A interaction effect) term of model (4) can be extended to a more general setting in which the response Y follows an exponential family distribution. Given (X, A) , we assume an additive single-index model (4):

$$m(X, A) = \mathbb{E}[Y | X, A] = \mu_0(X) + g_0(\beta_0^\top X, A), \tag{21}$$

(in which subscript (0) is used to indicate the “true” value), and the variance of Y is determined based on the exponential family density of the form:

$$\exp\{[Yh(m(X, A)) - b(h(m(X, A)))]/a(\phi) + c(Y, \phi)\}, \tag{22}$$

where h is the *canonical* link function associated with the assumed distribution of Y , and the functions a , b and c are distribution-specific known functions. For model identifiability, we assume $\beta_0 \in \Theta$, and $g_0 \in \mathcal{H}_2^{(\beta_0)}$ to satisfy $\mathbb{E}[g_0(\beta_0^\top X, A) | X] = 0$. The dispersion parameter $\phi > 0$ in (22) takes on a fixed, known value in some families (e.g., $\phi = 1$, in Bernoulli and Poisson), while in other families, it is an unknown parameter (e.g., in Gaussian).

We estimate (g_0, β_0) in model (21) by utilizing the following working mean model:

$$m(X, A) = h^{-1}(g(\beta^\top X, A)) \left(g \in \mathcal{H}_2^{(\beta)}; \beta \in \Theta \right), \tag{23}$$

subject to the constraint $\mathbb{E}[g(\beta^\top X, A) | X] = 0$ on g , and the distribution of Y is described by the exponential family chosen in (22). We optimize the population logarithmic version of (22) over the unknowns (g, β) of (23):

$$\begin{aligned} (g^*, \beta^*) = & \operatorname{argmax}_{g \in \mathcal{H}_2^{(\beta)}, \beta \in \Theta} \mathbb{E}[Yg(\beta^\top X, A) - b(g(\beta^\top X, A))] \\ & \text{subject to } \mathbb{E}[g(\beta^\top X, A) | X] = 0, \end{aligned} \tag{24}$$

where (we use the convention that the terms (i.e., $a(\phi)$ and $c(Y, \phi)$) that do not contain the parameter of interest can be dropped from the expression of a log likelihood) the expectation in the criterion function is with respect to the joint distribution of (Y, A, X) . Thus, the solution (g^*, β^*) on the left-hand side of (24) corresponds to the minimizer of the Kullback-Leibler (KL) divergence between the distributions with the working mean model (23) and the true mean model (21). In (24), $b(s) = s^2/2$ for a Gaussian Y (for which the optimization (9) is a special case of (24)), $b(s) = \log\{1 + \exp(s)\}$ for a Bernoulli Y , and $b(s) = \exp(s)$ for a Poisson Y .

The constraint $\mathbb{E}[g(\beta^\top X, A) | X] = 0$ in optimization (24) implies:

$$\begin{aligned} \mathbb{E}[Yg(\beta^\top X, A) - b(g(\beta^\top X, A))] &= \mathbb{E}[\{\mu_0(X) + g_0(\beta_0^\top X, A)\}g(\beta^\top X, A) - b(g(\beta^\top X, A))] \\ &= \mathbb{E}[\mathbb{E}[\mu_0(X)g(\beta^\top X, A) | X]] \\ &+ \mathbb{E}[g_0(\beta_0^\top X, A)(g(\beta^\top X, A) - b(g(\beta^\top X, A)))] \\ &= \mathbb{E}[g_0(\beta_0^\top X, A)(g(\beta^\top X, A) - b(g(\beta^\top X, A)))] \end{aligned} \tag{25}$$

which is free of μ_0 in model (21). Therefore, the left-hand side, (g^*, β^*) , of (24) does not depend on the unspecified X “main” effect function μ_0 .

Proposition 1.

The solution (g^*, β^*) of the constrained optimization problem (24) satisfies:

$$g_0 = h^{-1} \circ g^* \text{ and } \beta_0 = \beta^* \quad (26)$$

(a.s.), where $g_0 \in \mathcal{R}_2^{(\beta_0)}$ and $\beta_0 \in \Theta$ are given from the true mean model (21), and the function h^{-1} is the inverse of the canonical link function associated with the assumed exponential family distribution and the operator \circ represents the composition of two functions.

The proof of Proposition 1 is in Supplemental Materials Section A.1. In (26), $h^{-1}(s)=s$ (the identity function) for a Gaussian Y , $h^{-1}(s)=\exp(s)/\{1+\exp(s)\}$ for a Bernoulli Y , and $h^{-1}(s)=\exp(s)$ for a Poisson Y .

To solve (24) based on observed data (Y_i, A_i, X_i) ($i = 1, \dots, n$), we replace the quadratic loss term in (16) by the negative of the log likelihood of the data. For a fixed $\beta \in \Theta$, given the representation of g with $D\theta$ in (13), the basis coefficient θ is estimated by the *inner* iteratively re-weighted least squares (IRLS) and the associated smoothing parameters (λ and λ in (16)) are estimated by the outer optimization of, e.g., REML or GCV, as part of Step 1 (of Section 3.2) of the model fitting. The only adjustment to be made to the conventional GAM estimation is to enforce the constraint $E[g(\beta^\top X, A) | \beta^\top X] = 0$ on g . As in Section 3.1, this constraint can be absorbed into the tensor product basis representation (13). For Step 2 (of Section 3.2), once g is profiled out (by Step 1), we replace the *normal residual* vector in (19), i.e., $Y_{n \times 1} - g(X\tilde{\beta}, A_{n \times 1})$, by the *working residual* from the final IRLS fit of Step 1. Then we perform a weighted least squares (instead of the least squares (20)) for β , where the weights are given from the final IRLS fit of Step 1. We alternate between the Steps 1 and 2 until convergence of $\hat{\beta}$, as in Section 3.2. The resulting estimate for (g^*, β^*) in (24) is then used to estimate (g_0, β_0) in (21), based on the relationship (26). We also note that the framework (24) can be extended to incorporate a multinomial response, an instance of which we discuss in the following remark.

Remark 1.

The approach (24) to optimizing the component (g, β) of SIMSL can be generalized to the context of a proportional odds single-index model, where we have an ordinal response Y , in which its value exists on an arbitrary scale (in K categories), with only the relative ordering between different values being important. To deal with such a case, we introduce a length- K response vector $\mathbf{Y} = (Y_1, Y_2, \dots, Y_K)^\top$ where its component Y_j denotes the indicator for category j , with associated vector of probabilities $(p_1, p_2, \dots, p_K)^\top$, in which $p_K = 1 - \sum_{j=1}^{K-1} p_j$ (and $p_j > 0$), together with their cumulative probabilities: $P(Y \leq j) = q_j = \sum_{s=1}^j p_s$ ($j = 1, 2, \dots, K-1$). We can model these cumulative probabilities $(q_1, q_2, \dots, q_{K-1}, 1)^\top$ a by cumulative logit SIMSL:

$$h(P(Y \leq j | X, A)) = h(q_j(X, A)) = \alpha_j + \mu(X) + g(\beta^\top X, A) \quad (j = 1, 2, \dots, K-1) \quad (27)$$

where $\alpha_j \in \mathbb{R}$ ($\alpha_1 < \alpha_2 < \dots < \alpha_{K-1}$) are unknown cut-point parameters associated with the ordered response categories $j=1,2,\dots,K-1$, and $h(s) = \log(s/(1-s))$ is the logit link. For the (K -category) multinomial exponential family response \mathbf{Y} , let us consider its canonical parameter $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_{K-1}, 0)^\top \in \mathbb{R}^K$, where its nonzero components are specified by: $\eta_j = \log(p_j/p_K) = \log((q_j - q_{j-1})/(1 - q_{K-1}))$ ($j = 1, \dots, K-1$) (with $q_0 = 0$), in which the cumulative probabilities $(q_1, q_2, \dots, q_{K-1}, 1)^\top$ are specified by model (27). This multinomial exponential family representation for the distribution of Y allows us to use the optimization framework (24), with its criterion function extended to incorporate the multivariate response: $\mathbb{E}[\mathbf{Y}^\top \boldsymbol{\eta} - b(\boldsymbol{\eta})]$, where $b(\boldsymbol{\eta}) = \log\left(1 + \sum_{j=1}^{K-1} \exp(\eta_j)\right)$, for optimization of the cumulative logit SIMSL (27). The threshold-point parameters $\alpha_j \in \mathbb{R}$ ($\alpha_1 < \alpha_2 < \dots < \alpha_{K-1}$) in model (27) are estimated as part of Step 1 of model fitting in Section 3.2 (alongside the model smoothing parameters), in which the model (27), for each fixed $\boldsymbol{\beta}$ which is estimated as part of Step 2, is optimized via the penalized IRLS with an empirical version of $-\mathbb{E}[\mathbf{Y}^\top \boldsymbol{\eta} - b(\boldsymbol{\eta})]$; the Steps 1 and 2 are iterated until convergence. The heterogeneous treatment effect $g(\boldsymbol{\beta}^\top X, A)$ in model (27) does not depend on j nor $\mu(X)$; this allows us to develop an individualized dose rule independently of the arbitrary categorization of Y , and of the unspecified X main effect $\mu(X)$ that does not influence the treatment effect. Given X , the cumulative logit SIMSL-based individualized dose rule is $f(X) = \operatorname{argmax}_{a \in A} g(\boldsymbol{\beta}^\top X, a)$.

In Supplemental Materials Section C, we provide a real data example illustrating the approach (24) to modeling interaction effects between X and A on a count response variable Y , and a simulation illustration for the utility of the proportional odds SIMSL (27) when the treatment response Y is ordinal, which is common in biomedical and epidemiological studies.

5 Simulation example

In this section, we consider a set of simulation studies with data generated from the four scenarios described in Chen *et al.* (2016). We generate p -dimensional covariates $X = (X_1, \dots, X_p)^\top$, where each entry is generated independently from Uniform $[-1, 1]$. In Scenarios 1 and 2, the treatment A is generated from Uniform $[0, 2]$ independently of X , mimicking a randomized trial. In Scenarios 3 and 4, the distribution of A (described below) depends on X , mimicking an observational study setting. In each scenario, the outcome Y given (X, A) is generated from the standard normal distribution, with the following four different mean function scenarios:

1. Scenario 1: $\mathbb{E}[Y | X, A] = 8 + 4X_1 - 2X_2 - 2X_3 - 25(f_{\text{opt}}(X) - A)^2$, where $f_{\text{opt}}(X) = 1 + 0.5X_1 + 0.5X_2$. Here, the optimal individualized dose rule is a linear function of X .
2. Scenario 2: $\mathbb{E}[Y | X, A] = 8 + 4\cos(2\pi X_2) - 2X_4 - 8X_5^3 - 15|f_{\text{opt}}(X) - A|$, where

$$f_{\text{opt}}(X) = 0.6(-0.5 < X_1 < 0.5) + 1.2(X_1 > 0.5) + 1.2(X_1 < -0.5) + X_4^2 + 0.5\log(|X_7| + 1) - 0.6.$$

Here, the optimal individualized dose rule is a nonlinear function of X .

3. Scenario 3 is the same as in Scenario 2, except that A depends on X as follows:

$$A \sim \begin{cases} \text{TruncN}(-0.5 + 0.5X_1 + 0.5X_2, 0, 2, 0.5), & \text{when } X_3 < 0 \\ \text{TruncN}(|0.5 + 1.5X_2|, 0, 2, 1), & \text{when } X_3 > 0 \end{cases}$$

where $\text{TruncN}(\mu, a, b, \sigma)$ denotes the truncated normal distribution with mean μ , lower bound a and upper bound b , and standard deviation σ .

4. Scenario 4 is the same as in Scenario 2, except that A depends on X as follows:

$$A \sim \text{TruncN}(f_{\text{opt}}(X), 0, 2, 0.5).$$

Following Chen *et al.* (2016), we set $p = 30$ in Scenario 1, and $p = 10$ for Scenarios 2, 3 and 4. (In Section A.2 of Supplementary Materials, we describe how these scenarios' mean models can be reparametrized to fit in the framework (4).) For each simulated dataset, we apply the proposed method of estimating the A -by- X interaction term in the SIMSL (4) based on (9), and the optimal dose rule f_{opt} by $\hat{f}(X) = \underset{a \in A}{\operatorname{argmax}} \hat{g}^*(\hat{\beta}^* \top X, a)$.

We simulated 200 data sets for each scenario. For comparison, we report results of the estimation approaches considered in Chen *et al.* (2016), including their Gaussian kernel-based outcome-weighted learning (K-O-learning) and linear kernel-based outcome-weighted learning (L-O-learning). We also report a support vector regression (SVR; Vapnik, 1995; Smola and Scholopf, 2004) with a Gaussian kernel to estimate the nonlinear relationship between Y and (X, A) (Zhao *et al.*, 2009) that was used for comparison. In Scenario 1, we used (X, A) as the predictors for the outcome in the SIMSL. In Scenarios 2, 3 and 4, we used (X, X^2, A) (i.e., including a quadratic term in X) as the predictors for the SIMSL.

Since we are simulating data from known models in which the true relationship $\mathbb{E}[Y | X, A]$ is known, we can compare the estimated dose rules \hat{f} derived from each method in terms of the value (1). Specifically, an independent test set of size $\tilde{n} = 5000$ generated and the value (1) of \hat{f} was approximated using $\mathcal{V}(\hat{f}) = \tilde{n}^{-1} \sum_{i=1}^{\tilde{n}} \mathbb{E}[Y_i | X_i, A_i = \hat{f}(X_i)]$, for each simulation run. Given each scenario and a training sample size n , we replicate the simulation experiment 200 times, each time estimating the value. Again, following Chen *et al.* (2016), we report the averaged estimated values (and standard deviations) for the cases where \hat{f} is estimated from a training set of size $n = 50, 100, 200, 400$ and 800 for Scenario 1 and 2, and the cases with $n = 200$ and 800 for Scenario 3 and 4. The simulation results are given in Tables 1 and 2.

The results in Tables 1 and 2 indicate that the proposed regression method for optimizing dose rules outperforms the alternative approaches presented in Chen *et al.* (2016) in all cases

except when the training sample size is very small ($n = 50$). In Table 2, K-O-learning(Prp) refers to the propensity score-adjusted K-O-learning of Chen *et al.* (2016). When $n = 50$, the outcome-weighted learning methods outperform the regression-based approaches (i.e., SIMSL and SVR), especially for Scenario 1 (with $p = 30$) where the regression approaches exhibit large variances. However, when $n = 100$, the performance of SIMSL improves dramatically in terms of both value and small variance. We also note that using (A, X, X^2) instead of (A, X) as predictors of SIMSL in Scenario 2 lead to a substantial improvement in performance. If (X, A) is used for SIMSL in Scenario 2, the estimated values (and sd) are: $-0.98(2.15), 0.56(1.54), 1.91(0.89), 2.70(0.64)$ and $3.23(0.41)$, for $n = 50, 100, 200, 400$ and 800 , respectively.

6 Application to optimization of the warfarin dose with clinical and pharmacogenetic data

In this section, the utility of the SIMSL approach to personalized dose finding is illustrated from an anticoagulant study. Warfarin is a widely used anticoagulant to treat and prevent blood clots. The therapeutic dosage of warfarin varies widely across patients. Our analysis of the data will broadly follow that of Chen *et al.* (2016). After removing patients with missing data, the dataset provided by International Warfarin Pharmacogenetics Consortium *et al.* (2009) (publicly available to download from <https://www.pharmgkb.org/downloads/>) consists of 1780 subjects, including information on patient covariates (X), final therapeutic dosages (A), and patient outcomes (INR, International Normalized Ratio). INR is a measure of how rapidly the blood can clot. For patients prescribed warfarin, the target INR is around 2.5. To convert the INR to a measurement responding to the warfarin dose level, we construct an outcome $Y = -|2.5 - \text{INR}|$, and a larger value of Y is considered desirable.

There were 13 covariates $X = (X_1, \dots, X_{13})^T$ in the dataset (both clinical and pharmacogenetic variables): weight (X_1), height (X_2), age (X_3), use of the cytochrome P450 enzyme inducers (X_4 ; the enzyme inducers considered in this analysis includes phenytoin, carbamazepine, and rifampin), use of amiodarone (X_5), gender (X_6 ; 1 for male, 0 for female), African or black race (X_7), Asian race (X_8), the VKORC1 A/G genotype (X_9), the VKORC1 A/A genotype (X_{10}), the CYP2C9 *1/*2 genotype (X_{11}), the CYP2C9 *1/*3 genotype (X_{12}), and the other CYP2C9 genotypes (excluding the CYP2C9 *1/*1 genotype which is taken as the baseline genotype) (X_{13}). Further details on these covariates are given in International Warfarin Pharmacogenetics Consortium *et al.* (2009). The first 3 covariates (weight, height, age) were treated as continuous variables, standardized to have mean zero and unit variance; the other 10 covariates are indicator variables.

In estimating the optimal individualized dose rule f_{opt} , modeling the drug (dose level A) interactions with the patient covariates X is essential. Under the proposed SIMSL approach (4), $f_{\text{opt}}(X) = \underset{a \in \mathcal{A}}{\text{argmax}} g(\beta^T X, a)$ and thus the A -by- X interaction effect term $g(\beta^T X, A)$ is

the target component of interest. In SIMSL, due to orthogonality (7), we can solve (9) for (g, β) , without having to model the μ term in (4). However, modeling the μ term, even with a misspecified working model, can generally improve the efficiency of the estimator of

(g, β) (i.e., yielding smaller variances for estimators; see Park *et al.* (2020) for a theoretical justification in the case where treatment A is a discrete or binary variable), which leads to improve the efficiency of the estimator of f_{opt} (see Supplementary Materials Section C.1 for a simulation illustration in which an enhanced estimation performance is illustrated when the X main effect is incorporated to the estimation of f_{opt}).

Thus, in this application, we model the μ term of (4) with a possibly misspecified working model, which consists of a set of linear terms for the indicators X_4, \dots, X_{13} and a set of cubic P -spline smooth terms for the continuous covariates X_1, X_2 and X_3 . These terms are estimated alongside the heterogeneous treatment effect term $g(\beta^T X, A)$ by the procedure described in Supplementary Materials Section C.1, which is a slight modification to that described in Section 3.2. (The estimated coefficient β , with or without incorporating the μ term in the estimation, along with their bootstrap confidence intervals, are provided in Section D of Supplementary Materials.) The third panel in Figure 1 displays a surface plot of the estimated 2-dimensional link function $g(\beta^T X, A)$, showing an interactive relationship on the index-treatment domain. The first two panels in Figure 1 display the estimated marginal effect function for the dose A and that for the estimated index $\beta^T X$.

We construct a 95% normal approximation bootstrap confidence interval for β , based on 500 bootstrap replications (see Supplementary Materials Section C.3 for the confidence interval construction and for a coverage probability simulation). The confidence intervals for the β_j 's associated with the covariates weight (X_1), height (X_2), the use of the cytochrome P450 enzyme inducers (X_4), Asian race (X_8), the VKORC1 A/G genotype (X_9), and the CYP2C9 *1/*2 genotypes (X_{11}) do not include 0. We infer that these covariates are potentially clinically important drug effect modifiers, interacting with warfarin in their effects on INR.

Chen *et al.* (2016) noted that the analysis results from International Warfarin Pharmacogenetics Consortium *et al.* (2009), as well as their linear kernel-based outcome-weighted learning results, suggest increasing the dose if patients are taking Cytochrome P450 enzyme (X_4). Roughly speaking, the interaction surface g (the right-most panel) in Figure 1 indicates that for a smaller value of $\beta^T X$ (e.g., $\beta^T X < -0.3$), a moderate or a relatively low dose A (e.g., $A < 50$) may be preferred, whereas for a larger value of $\beta^T X$ (e.g., $\beta^T X > 0$), a relatively high dose A (e.g., $A > 50$) may be preferred. Considering the sign of the estimated coefficient ($\hat{\beta}_4 = -0.60$) associated with X_4 , this is roughly consistent with International Warfarin Pharmacogenetics Consortium *et al.* (2009) and Chen *et al.* (2016).

To evaluate the performance of the individualized dose rules estimated from the 6 methods, including the propensity score-adjusted outcome-weighted learning with a linear/Gaussian kernel, denoted as L-O-learning(Prp) and K-O-learning(Prp), respectively) considered in Section 5, we randomly split the dataset at a ratio of 1-to-1 into a training set and a testing set, replicated 100 times, each time estimating f_{opt} using the 6 methods based on the training set, and estimating the value (1) of each estimated f_{opt} based on the testing set. Unlike the simulated data in Section 5, the true relationship between the covariate-specific dose and the response is unknown. Therefore, for each dose rule f , we need to estimate the value (1) from the testing data. Given a dose rule f , only a very small proportion (or none) of

the observations will satisfy $A_j = f(X_j)$, and thus only a very small proportion (or none) of the observations in the testing data will contribute information to estimate the value (1). However, Cai and Tian (2016) noted that the value (1) for each f can also be written as $\mathcal{V}(f) = \mathbb{E}[\mathbb{E}[Y | A = f(X), f(X)]]$. Therefore, using a 2-dimensional smoother of A and $f(X)$ for Y , one may first obtain a nonparametric estimate of $\mathbb{E}[Y | A, f(X)]$, denoted as $\hat{m}(A, f(X))$, and then $\mathcal{V}(f)$ may be estimated as $\mathcal{V}(f) = n^{-1} \sum_{i=1}^n \hat{m}(f(X_i), f(X_i))$. Specifically, given a dose rule f estimated from a training set, we can estimate $\mathbb{E}[Y | A, f(X)]$ based on $(Y_i, A_i, f(X_i))$ from a test set, using a set of thin plate regression spline bases obtained from a rank-100 eigen-approximation to a thin plate spline, with the smoothness parameter selected by REML, implemented via the R (R Core Team, 2019) function `mgcv::gam` (Wood, 2019). A thin plate spline is an isotropic smooth; isotropy is often appropriate for two variables observed on the same scale, which is the case here.

Figure 2 displays boxplots describing the distributions for the estimated values (1) of the approaches “SIMSL(w. X main)” (SIMSL *with* the μ term in the estimation) and “SIMSL(w.o. X main)” (SIMSL *without* the μ term in the estimation), and the other five estimation methods described in Section 5, obtained from the aforementioned 100 random training/testing splits. The boxplots indicate that the proposed SIMSL methods (note that “SIMSL(w. X main)” slightly outperforms “SIMSL(w.o. X main)”) and the propensity-score adjusted K-O-learning of Chen *et al.* (2016) perform at a similar level, while outperforming all other approaches, illustrating the potential utility of the proposed method. In comparison to the outcome-weighted learning approach of Chen *et al.* (2016), one advantage of the proposed approach is that it allows visualization of the estimated interactive structure on the dose-index domain, as illustrated in the right panel of Figure 1. Additionally, if each of the covariates is standardized to have, say, unit variance, then the relative importance of each covariate in characterizing the heterogeneous dose response can be determined by the magnitude of the estimated coefficients in β , rendering a potentially useful interpretation when examining the drug-covariates interactions.

7 Discussion

In this paper, we proposed a single-index model that utilizes a surface link-function as a function of a linear projection of covariates and a continuous “treatment” variable, which parsimoniously represents the interaction effect between covariates and a treatment defined on a continuum. The model provides an intuitive tool for investigating personalized dose finding in precision medicine, without the need for a significant change in the established generalized additive regression modeling framework.

One important limitation is that the confidence band associated with the estimated surface g is computed conditional on the estimated $\beta^T X$, and the uncertainty in β is not accounted for. The fact that the domain of g varies depending on the estimate of β complicates the confidence band construction for g . One potential approach is to consider a Bayesian framework and a posterior distribution of $g(\beta^T X, A)$, and make probabilistic statements about the prediction of the component $g(\beta^T X, A)$ given (X, A) . Furthermore, depending on context, for scientific interpretability of the model, it may be sometimes desirable to

consider shape constraints such as monotonicity or convexity/concavity (see Supplementary Materials Section C.4 for discussion on shape constraints), as well as optimization under safety constraints (Laber *et al.*, 2018). The development of a Bayesian model estimation and inference, with potential monotonicity or convexity/concavity constraints on the link surface is currently under investigation. In many applications, only a subset of variables may be useful in determining an optimal individualized dose rule. Also, high-dimensional settings can lead to instabilities and issues of overfitting. Forthcoming work will introduce a regularization method that can both avoid overfitting and choose among multiple potential covariates by obtaining a sparse estimate of the single-index coefficient β . Future extensions of this work could also include an extension to incorporate a functional covariate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institute of Mental Health grant 5 R01 MH099003 and National Center for Advancing Translational Sciences grant 3 UL1TR001445–06A1S2.

References

- Cai T and Tian L (2016). Comment: Personalized dose finding using outcome weighted learning. *Journal of the American Statistical Association* 111:1521–1524.
- Chen G, Zeng D, and Kosorok MR (2016). Personalized dose finding using outcome weighted learning. *Journal of the American Statistical Association* 111:1509–1547. [PubMed: 28255189]
- Chen H, Lu W, and Song R (2020). Statistical inference for online decision making: In a contextual bandit setting. *Journal of the American Statistical Association*. doi:10.1080/01621459.2020.177009.
- de Boor C (2001). *A Practical Guide to Splines*. Springer-Verlag, New York.
- Dudík M, Erhan D, Langford J, and Li L (2014). Doubly robust policy evaluation and optimization. *Statistical Science* 29:485–511.
- Eilers P and Marx B (2003). Multivariate calibration with temperature interaction using 2-dimensional penalized signal regression. *Chemometrics and Intelligence Laboratory Systems* 66:159–174.
- Gill RD and Robins JM (2001). Causal inference for complex longitudinal data: the continuous case. *Annals of Statistics* 29:1785–1811.
- Gunter L, Zhu J, and Murphy S (2011). Variable selection for qualitative interactions in personalized medicine while controlling the family-wise error rate. *Journal of Biopharmaceutical Statistics* 21:1063–1078. [PubMed: 22023676]
- Hastie T and Tibshirani R (1999). *Generalized Additive Models*. Chapman & Hall Ltd.
- International Warfarin Pharmacogenetics Consortium, Klein T, Altman R, Eriksson N, Gage B, Kimmel S, Lee M, Limdi N, Page D, Roden D, Wagner M, Caldwell M, and Johnson J (2009). Estimation of the warfarin dose with clinical and pharmacogenetic data. *The New England Journal of Medicine* 360:753–674. [PubMed: 19228618]
- Jun KS, Bhargava A, Nowak R, and Willett R (2017). Scalable generalized linear bandits: Online computation and hashing. In *Advances in Neural Information Processing Systems* 30:99–109.
- Kaelbling L, Littman M, and Moore A (1996). Reinforcement learning: a survey. *Journal of Artificial Intelligence Research* 4:237–285.
- Kallus N and Zhou A (2018). Policy evaluation and optimization with continuous treatments. In *International Conference on Artificial Intelligence and Statistics* 84:1243–1251.

- Kennedy EH, Ma Z, McHugh MD, and Small DS (2017). Nonparametric methods for doubly robust estimation of continuous treatment effects. *Journal of Royal Statistical Society: Series B* 79:1229–1245.
- Kleinberg R, Slivkins A, and Upfal E (2019). Bandits and experts in metric spaces. *Journal of the ACM* 66:77.
- Krishnamurthy A, Langford J, Slivkins A, and Zhang C (2020). Contextual bandits with continuous actions: Smoothing, zooming, and adapting. *Journal of Machine Learning Research* 21:1–45. [PubMed: 34305477]
- Kveton B, Zaheer M, Szepesvari C, Li L, Ghavamzadeh M, and Boutilier C (2020). Randomized exploration in generalized linear bandits. *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics* 108:2066–2076.
- Laber E, Wu F, Munera C, Lipkovich I, Colucci S, and Ripa S (2018). Identifying optimal dosage regimes under safety constraints: An application to long term opioid treatment of chronic pain. *Statistics in Medicine* 37:1407–1418. [PubMed: 29468702]
- Laber EB and Zhao Y (2015). Tree-based methods for individualized treatment regimes. *Biometrika* 102:501–514. [PubMed: 26893526]
- Lattimore T and Szepesvari C (2019). *Bandit algorithms*. Cambridge University Press.
- Li L, Lu Y, and Zhou D (2017). Provably optimal algorithms for generalized linear contextual bandits. In *Proceedings of the 34th International Conference on Machine Learning* 70:2071–2080.
- Lu T, Pal D, and Pal M (2010). Contextual multi-armed bandits. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings* 9:485–492.
- Lu W, Zhang H, and Zeng D (2013). Variable selection for optimal treatment decision. *Statistical Methods in Medical Research* 22:493–504. [PubMed: 22116341]
- Majzoubi M, Zhang C, Chari RK, Langford A, J., and Slivkins A (2020). Efficient contextual bandits with continuous actions. *ArXiv, abs/2006.06040*.
- Marx B (2015). Varying-coefficient single-index signal regression. *Chemometrics and Intelligence Laboratory Systems* 143:111–121.
- Murphy SA (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65:331–355.
- Park H, Petkova E, Tarpey T, and Ogden R (2021). *simsl: Single-index models with a surface-link*. R package version 0.2.1.
- Park H, Petkova E, Tarpey T, and Ogden RT (2020). A constrained single-index regression for estimating interactions between a treatment and covariates. *Biometrics*. doi:10.1111/biom.13320.
- Perchet V and Rigollet P (2013). The multi-armed bandit problem with covariates. *The Annals of Statistics* 41:693–721.
- Qian M and Murphy SA (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics* 39:1180–1210. [PubMed: 21666835]
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rich B, Moodie EE, and Stephens DA (2014). Simulating sequential multiple assignment randomized trials to generate optimal personalized warfarin dosing strategies. *Clinical Trials* 11:435–444. [PubMed: 24464036]
- Robins J (2004). *Optimal Structural Nested Models for Optimal Sequential Decisions*. Springer, New York.
- Schulz J and Moodie EE (2020). Doubly robust estimation of optimal dosing strategies. *Journal of the American Statistical Association*. doi:10.1080/01621459.2020.1753521.
- Slivkins A (2014). Contextual bandits with similarity information. *Journal of Machine Learning Research* 15:2533–2568.
- Smola A and Schölkopf B (2004). A tutorial on support vector regression. *Statistics and Computing* 14:199–222.
- Vapnik VN (1995). *The Nature of Statistical Learning Theory*, volume 8. Springer, New York.

- Wood SN (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* 62:1025–1036. [PubMed: 17156276]
- Wood SN (2017). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, second edition.
- Wood SN (2019). *mgcv: Mixed GAM computation vehicle with automatic smoothness estimation*. R package version 1.8.28.
- Zhang B, Tsiatis AA, Davidian M, Zhang M, and Laber E (2012). Estimating optimal treatment regimes from classification perspective. *Stat* 1:103–114. [PubMed: 23645940]
- Zhao Y, Kosorok M, and Zeng D (2009). Reinforcement learning design for cancer clinical trials. *Statistics in Medicine* 28:3294–3315. [PubMed: 19750510]
- Zhao Y, Zeng D, Rush AJ, and Kosorok MR (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association* 107:1106–1118. [PubMed: 23630406]
- Zhou W, Zhu R, and Zeng D (2020). A parsimonious personalized dose-finding model via dimension reduction. *Biometrika*. doi:10.1093/biomet/asaa087.
- Zhu L, Lu W, Kosorok MR, and Song R (2020). Kernel assisted learning for personalized dose finding. arXiv:2007.09811.

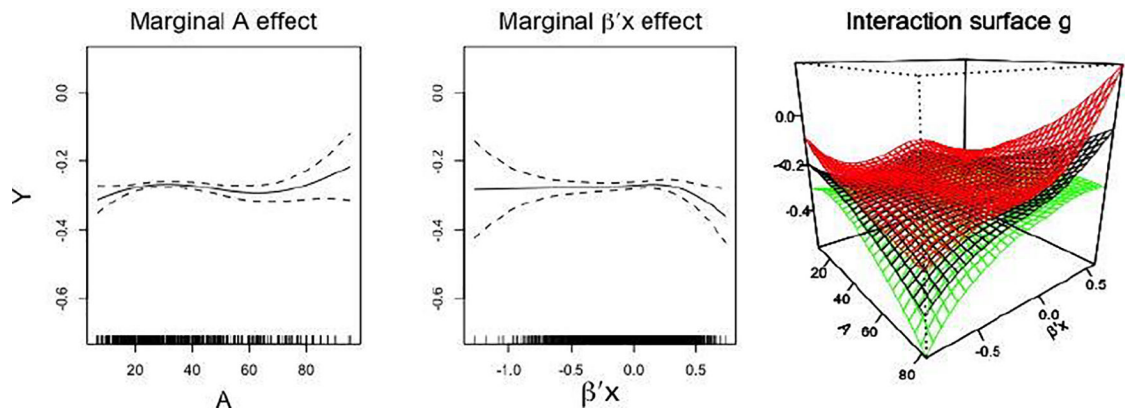


Fig. 1.

The first two panels: the marginal effect of dose A (left panel) and that of the estimated single-index (middle panel) with 95% confidence bands (dashed curves) given the estimated $\beta^T X$. The third panel: the estimated link surface (g) for the dose (A) and index ($\beta^T X$) interaction; the red and green surfaces are at ± 2 standard error from the estimated surface (the black) in the middle, conditioning on the estimated single-index.

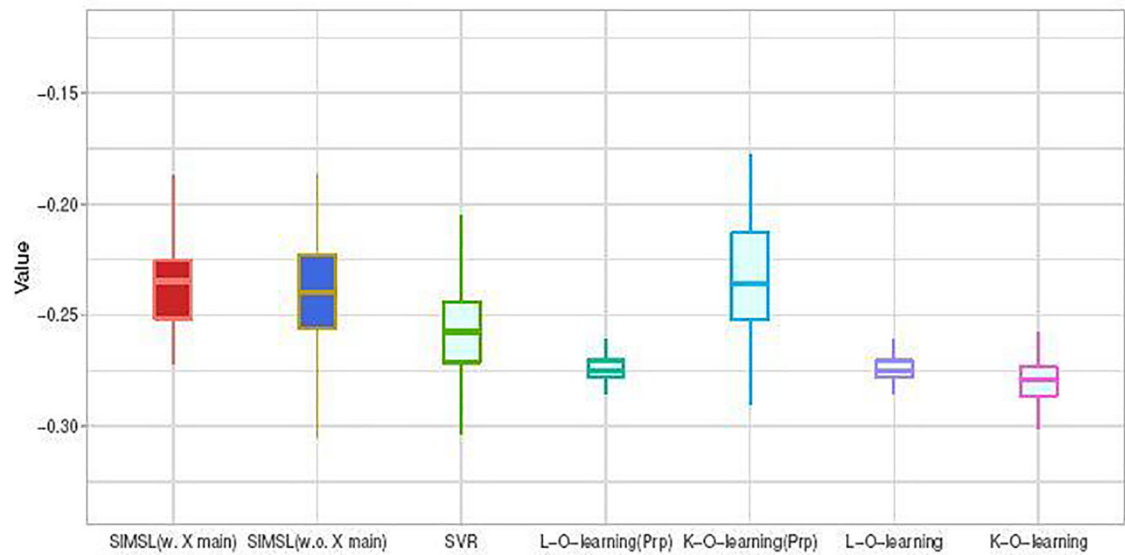


Fig. 2. Boxplots of the estimated values of the individualized dose rules using 7 approaches, obtained from 100 randomly split testing sets. Mean (and standard deviation) of the value estimates: SIMSL(w. X main): -0.231 (0.03); SIMSL(w.o. X main): -0.237 (0.03); SVR: -0.254 (0.02); L-O-learning(Prp): -0.274 (0.01); K-O-learning(Prp): -0.234 (0.03); L-O-learning: -0.274 (0.01); K-O-learning: -0.279 (0.01).

Table 1

Average (sd) value $\mathcal{V}(f)$ from 200 replicates from randomized trials. In both settings, the oracle f_{opt} attains a value $\mathcal{V}(f_{\text{opt}}) = 8$ (boldface denotes the largest in each row).

	n	SIMSL	K-O-learning	L-O-learning	SVR
Scenario 1	50	1.04 (4.06)	4.78 (0.48)	4.83 (1.40)	-12.21 (7.53)
	100	6.63 (0.63)	5.69 (0.40)	5.39 (0.93)	-2.57 (6.34)
	200	7.45 (0.20)	6.68 (0.26)	6.85 (0.34)	3.46 (1.97)
	400	7.77 (0.08)	7.28 (0.15)	7.41 (0.14)	6.13 (0.47)
	800	7.88 (0.04)	7.54 (0.08)	7.67 (0.08)	7.36 (0.12)
Scenario 2	50	0.90 (2.04)	2.00 (0.29)	1.16 (0.71)	-1.96 (1.70)
	100	3.65 (0.76)	2.19 (0.43)	1.57 (0.52)	0.24 (1.42)
	200	4.71 (0.41)	2.84 (0.37)	2.02 (0.30)	2.01 (0.84)
	400	5.25 (0.20)	3.69 (0.27)	2.30 (0.18)	3.47 (0.37)
	800	5.59 (0.12)	4.41 (0.19)	2.49 (0.10)	4.35 (0.19)

Table 2

Average (sd) value $\mathcal{V}(f)$ from 200 replicates from observational studies. In both settings, the oracle f_{opt} attains a value $\mathcal{V}(f_{\text{opt}}) = 8$ (boldface denotes the largest in each row).

	n	SIMSL	K-O-learning	K-O-learning(Prp)	SVR
Scenario 3	200	4.03 (0.97)	2.68 (0.30)	2.74 (0.29)	1.99 (0.83)
	800	5.46 (0.20)	4.06 (0.30)	4.19 (0.20)	4.09 (0.28)
Scenario 4	200	4.07 (0.72)	3.29 (0.28)	3.23 (0.28)	-0.95 (1.57)
	800	5.51 (0.19)	4.91 (0.14)	4.73 (0.17)	3.04 (0.52)