Personalized Treatment Plans with Multivariate Outcome Measures

Chathura Siriwardhana, Ph.D.

Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA

Somnath Datta, Ph.D.

Department of Biostatistics, University of Florida, Gainesville, FL, USA

K.B. Kulasekera, Ph.D.

Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA

Technical Report BST2017-02

Department of Bioinformatics and Biostatistics
School of Public Health and Information Science
University of Louisville
Louisville, KY, USA

Personalized Treatment Plans with Multivariate Outcome Measures

Chathura Siriwardhana[1], Somnath Datta[2] and K.B. Kulasekera[3,*]

[1]*Office of Biostatistics & Quantitative Health Sciences, University of Hawaii John A. Burns School of Medicine, Honolulu, HI 96813, USA,* [2]*Department of Biostatistics, University of Florida, Gainesville, FL 32603, USA,* [3]*Department of Bioinformatics & Biostatistics, University of Louisville, Louisville, KY 40202, USA*

Abstract

In this work we propose a novel method for individualized treatment selection when the treatment response is multivariate. Our method covers any number of treatments and it can be applied for a broad set of models. The proposed method uses a rank aggregation technique to estimate an ordering of treatments based on ranked lists of treatment performance measures such as smooth conditional means and conditional probability of a response for one treatment dominating others. An empirical study demonstrates the performance of the proposed method in finite samples. We also present a data analysis using a HIV clinical trial data to show the applicability of the proposed procedure for real data.

Key Words: Design variables; Personalized Treatments; Single Index Models; Rank Aggregation

1. Introduction

Designing optimal treatment regimes based on individual patient characteristics has gained a momentum over the last few years (see for example van't Veer and Bernards, 2008; Varquez, 2013). Dynamic treatment regimes that are geared towards the "best" outcome (eg: survival time) for a patient based on his/her genetic and genomic markers are of high importance. Literature on this topic only deals with deciding between treatments based on a single outcome measure modeled against patient characteristics. Assuming without any loss of generality that a larger outcome is better, the methods developed in the literature essentially determine the best treatment as the one associated with the largest of a measure of dominance. Existing literature use either a conditional location parameter or a measure based on a conditional probability of an outcome for one treatment exceeding the outcomes for others given the set of markers for the patient; see Siriwardhana et al. (2017) and references therein.

In many practical situations the success of a treatment cannot necessarily be measured via a single outcome as a variety of factors may compel both patients and clinicians to consider recovery in a rather broad view. For example, in deciding a treatment for a cancer, a clinician may use multiple values of gene expressions from different families of genes (Kelly et al., 2011,) as endpoint indicators of a successful treatment. Situations where the disease is not curable, eg: Multiple Myeloma, may require monitoring multiple measurements such as immunoglobulins, creatinine level etc. as outcome measures in planning optimal long term treatment regimes. Also, in many cancer treatment regimes

while longer remission times are highly desired, the impact of drug side effects/reactions, long term effects from drug combinations, the quality of life, social, family and economic factors etc. can also play an important role in deciding on treatment protocols. Hence, selecting the best treatment considering multiple outcome measures becomes a relevant issue for most patient populations.

In this paper we consider selection of the optimal treatment among K possible treatments for a patient using his or her baseline characteristics when multivariate outcomes (responses) are to be considered. First, to handle statistical issues arising due to high dimensional covariates, each patient is assigned a score based on his/her covariate values. Then we use a weighted rank aggregation method (see for example Pihur et al., 2007 and Pihur et al., 2009) to combine ranks(orderings) assigned to treatments based on each response. These ranks can be determined for each response using an existing criteria such as ordered conditional mean for each response given the patient score or quantities based on conditional probability of one treatment dominating others given the patient score. Additionally, the rank aggregation method in Pihur et al. (2007) is flexible to assign different importance factors to each response variable. This allows one to use apriori opinions on the importance of each response in determining the best treatment procedure. Our simulations studies show that the proposed method has very desirable properties in terms of selection frequency of the best treatment. A real data analysis show differences in the selection of the best treatment using multi responses compared with the selection using a single response.

The remainder of the article is organized as follows. In Section 2, we discuss the proposed methodology. Section 3 includes simulation results followed by a real data illustration in Section4. The main body of the paper ends with a discussion in Section 5.

2. Treatment Selection

In this section we describe the proposed procedure. Suppose we observe q response variables for each patient undergoing a treatment selected from K possible treatments and, without loss of generality, suppose larger values of the each component of the q dimensional response vectors are indicative of better outcomes. Let $\mathbf{Y}_i^* = (Y_{1i}^*, ..., Y_{qi}^*)'$ indicate the vector of responses for the ith treatment with an associated r dimensional covariate vector $\boldsymbol{X}$. Assume further that a patient's covariate value $\boldsymbol{X}$ is used to obtain a lower dimensional composite patient score $U(\boldsymbol{X})$. It should be noted that in practice, one cannot observe the full $q \times K$ matrix of counterfactuals $(\mathbf{Y}_1^*, ..., \mathbf{Y}_K^*)$ for a single patient.

In the single response case ($q = 1$) using iid observations of type $(\tilde{Y}_{1j}, \boldsymbol{X}_j, A_j)$, $j = 1, ..., n$ where A_j is the binary treatment indicator for two treatments and $\tilde{Y}_{1j}$ is the observed single response for the jth patient, previous authors have proposed the estimated difference in conditional means given a score U to compare two treatments. For example, Zhang et al. (2012) use robust estimators of $E[Y_{11}^*|A = 0, U(\boldsymbol{X})] - E[Y_{12}^*|A = 1, U(\boldsymbol{X})]$ where $U(\boldsymbol{X}) = \boldsymbol{X}$ and $A = 0$, 1 assign treatments 1 and 2 respectively. For

the K treatment case with a single outcome measure, one may use the largest index corresponding to estimated values of

$$\mu_{1i} = E\left[Y_{1i}^{*} | U(\boldsymbol{X}) = u\right]$$

$i = 1, ..., K$ for a suitable score function U as the best treatment. In contrast to this approach, because one cannot directly sample from the joint distribution of counterfactuals $(Y_{11}^{*}, ..., Y_{1K}^{*})'$, Siriwardhana et al. (2017) consider pairs of independent observations $(Y_{1i}, \boldsymbol{X}_i)$ from the marginal distribution of $(Y_{1i}^{*}, \boldsymbol{X}), i = 1, ..., K$ for the treatment selection for K treatments and proposed a method based on estimators of a set of probabilities defined as

$$p_{1i}(u) = P[Y_{1i} > \max_{1 \leq m \leq K; i \neq m} Y_{1m} | U(\boldsymbol{X_l}) = u; l = 1, ..., K]; i = 1, ..., K \qquad (1)$$

and compared that method against the criteria that uses the largest index corresponding to estimated values of

$$\mu_{1i} = E\left[Y_{1i} | U(\boldsymbol{X_i}) = u\right]; i = 1, ..., K \qquad (2)$$

as the best treatment. Note that in (1) and (2), the Ys do not denote the set of true counterfactuals for a patient (given the set of $\boldsymbol{X}$) but are independently distributed with the same marginal distributions (given the set of $\boldsymbol{X}$). The function $p_{1i}(u)$ above is

actually a surrogate for

$$p^*_{1i}(u) = P[Y^*_{1i} > \max_{i \neq m} Y^*_{1m} | U(\boldsymbol{X_i}) = u]; i = 1, ..., K. \tag{3}$$

However, in real life one cannot sample from the joint distribution of $(Y^*_{11}, ..., Y^*_{1K})'$ for a patient with covariate value $\boldsymbol{X}$ and therefore unable to estimate p^*_{1i}s. Using an extensive set of simulations, Siriwardhana et al. (2017) provide a convincing argument that p_{1i} above nevertheless gives a very competitive index compared to p^*_{1i} to measure the dominance of the ith treatment over others and hence can be used in selecting the best treatment. This was suggested as an alternative to measures based on conditional expectations $(\mu_{1i}, i = 1, ..., K)$ which require restrictive moment assumptions on the error distribution for inferential aspects in a regression context. The method based on the p_{1i}s was very competitive against methods based on μ_is for a variety of models as shown from their empirical studies. Readers are directed to Siriwardhana et al. (2017) for further details.

In dealing with multiple responses, we consider pairs of independent observations $(\mathbf{Y}_i, \boldsymbol{X}_i)$ from the marginal distribution of $(\mathbf{Y}^*_i, \boldsymbol{X}), i = 1, ..., K$ where $\mathbf{Y}_i = (Y_{1i}, ..., Y_{qi})'$. We propose to select the optimum treatment for K treatments using either vectors of smoothed conditional means for each treatment or sets of probabilities defined in a similar fashion in (1) above. For example, in generalizing the approach of using the

conditional means for the response vector $\mathbf{Y}_i = (Y_{1i}, ..., Y_{qi})'$ we define

$$\mu_{ki}(u_k) = E[Y_{ki}|U_k(\boldsymbol{X}_i) = u_k]; k = 1, ..., q; i = 1, ..., K \tag{4}$$

and vectors $\boldsymbol{\mu}_i(\mathbf{u}) = (\mu_{1i}(u_1), ..., \mu_{qi}(u_q))'$ for $\mathbf{u} = (u_1, ..., u_q)'$ where components of these vectors correspond to each response. Now we rank the K values for each component of $\boldsymbol{\mu}_i$ vectors (i=1,...,K) to get size K vectors $\mathbf{v}_k(\mathbf{u}) = (v_{k1}(\mathbf{u}), ..., v_{kK}(\mathbf{u}))'$ where $v_{ki}(\mathbf{u})$ is the rank of μ_{ki} among $\mu_{ki}, i = 1, ..., K$ for each k (here $k = 1, .., q$) with the largest μ_{ki} value given the rank 1. Then, we use an aggregation method to combine these rank vectors to get an overall ranking of treatments. In this article we use the method proposed in Pihur et al. (2007, 2009) to aggregate these rank vectors for a given score vector $\mathbf{U}_0 = (U_{10}, ..., U_{q0})'$. In particular, for a suitably chosen set of weights $\omega_k; k = 1, ..., q$ and a distance measure γ (Pihur et al. 2007), we minimize a quantity

$$\psi(\mathbf{v}) = \sum_{k=1}^{q} \omega_k \gamma(\mathbf{v}, \mathbf{v}_k(\mathbf{U}_0)) \tag{5}$$

over P_K, the set of all permutations of $\{1, ..., K\}$, to get a vector $\mathbf{v}^* = (v_1^*, ..., v_K^*)'$ where

$$\mathbf{v}^* = \arg\min_{\mathbf{v} \in P_K} \psi(\mathbf{v}). \tag{6}$$

Among possible distance measures for γ is the weighted Spearman's Footrule distance (Pihur et al., 2007) which was used in our empirical work. We then define the optimal

treatment as

$$i^*(\mathbf{U}_0) = \arg \min_{1 \le i \le K} \{v_i^*\} \tag{7}$$

We illustrate the proposed procedure with a simple example. Suppose we have a situation with $K = 3$ treatments with $q = 4$ responses with μ_{ik}s and corresponding ranks as

$$\begin{pmatrix} 30 & 35 & 28 \\ 10 & 18 & 30 \\ 14 & 12 & 8 \\ 22 & 18 & 31 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 & 1 & 3 \\ 3 & 2 & 1 \\ 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

For example, the first row of the second matrix above indicates that with respect to the first response, the second treatment as the best followed by treatments 1 and 3 . Now, if we use the aggregation algorithm in Pihur et al., (2007, 2009) which uses both the values of $\mu_{ki}, k = 1, ..., 4; i = 1, ..., 3$ and their corresponding ranks in the weighted Spearman's Footrule distance γ combined with $\omega_k = 1, k = 1, ..., 4$ we get the aggregated rank vector $\mathbf{v}^* = (3, 2, 1)'$ indicating that the treatment 3 is the best among the three competitors. On the other hand, use of $\omega_1 = 0.4, \omega_2 = 0.3, \omega_3 = \omega_4 = 0.15$, in the same aggregation algorithm results in $\mathbf{v}^* = (3, 1, 2)'$ indicating that treatment 2 is optimal.

If we are to use conditional probabilities as in Siriwardhana et al. (2017) to rank the treatments, we consider

$$p_{ki}(u_k) = P[Y_{ki} > \max_{m \neq i} Y_{km} | U_k(\boldsymbol{X_l}) = u_k; l = 1, ..., K]; k = 1, ..., q; i = 1, ..., K \tag{8}$$

and use the same aggregation method above to vectors of ranks corresponding to $\boldsymbol{p}_i(\mathbf{u}) = (p_{1i}(u_1), ..., p_{qi}(u_q))'$, $i = 1, ..., K$ in a similar fashion.

Remark 1. In the case of μ_{ki}s, as far as selecting using the conditional means is concerned, it really does not matter whether we have observations from the joint distribution of $(\mathbf{Y}_1^*, ..., \mathbf{Y}_K^*, \boldsymbol{X})'$ or marginal distributions of $(\mathbf{Y}_i^*, \boldsymbol{X}_i)'$s, because $\mu_{ki}(u) = E[\mathbf{Y}_{ki}^*|U(\boldsymbol{X}) = u] = E[\mathbf{Y}_{ki}|U(\boldsymbol{X}_i) = u]$. However, this is not the case for p_{ki}^*s versus p_{ki}s; in the case of p_{ki}^*s, we cannot obtain data to estimate it in real life whereas we can obtain samples to estimate p_{ki}s.

We base our discussion on a set of Single Index Models relating the kth component Y_{ki} of the response vector $\mathbf{Y}_i$ for the ith treatment and covariates $\boldsymbol{X}_i$ via

$$Y_{ki} = g_{ki}\left(\boldsymbol{\beta}_{ki}'\boldsymbol{X}_i\right) + \epsilon_{ki} \tag{9}$$

for $k = 1, \ldots, q$ and $i = 1, ..., K$ where each $\boldsymbol{\beta}_{ki}$ is a r-vector of parameters, g_{ki}s are unknown link functions for which we assume some reasonable smoothness conditions to hold, and ϵ_{ki} are error terms with $E[\epsilon_{ki}|\boldsymbol{X}_i] = 0$. Furthermore we assume independence of ϵ_{ki}s across $i = 1, ..., K$ for a fixed k where these terms are correlated across ks for any given i. The Single Index formulation provides flexibility and reasonable efficiency in modeling many types of data.

Our observations are of the following form. Let Y_{kij} indicate the kth component of the jth response from a group of n_i individuals under treatment i with associated covariate values $\boldsymbol{X}_{ij}, j = 1, \ldots, n_i$.A technical requirement for consistency of the proposed

procedure is that sample sizes n_i are assumed to satisfy the condition that n_i/N tends to a positive number where $N = \sum n_i$. Then, for this data, relationship (9) is written as

$$Y_{kij} = g_{ki}(\boldsymbol{\beta}'_{ki}\boldsymbol{X}_{ij}) + \epsilon_{kij}, j = 1, \ldots, n_i. \tag{10}$$

Following Siriwardhana et al. (2017) we define overall score vector $\mathbf{U}$ as follows. Our approach to define an appropriate overall score vector $\mathbf{U}$ is first to use a reasonable model to obtain a treatment specific score for each patient. The score for treatment i measures how favorable it is for a patient to receive this treatment when compared to if he or she were to receive other treatments for each response. To be specific, we first define

$$S_{ki}\left(\boldsymbol{X}\right) = g_{ki}\left(\boldsymbol{\beta}'_{ki}\boldsymbol{X}\right) - \max_{l \neq i}\left\{g_{kl}\left(\boldsymbol{\beta}'_{kl}\boldsymbol{X}\right)\right\}.$$

Next, define the kth components of the combined overall score vectors as

$$\begin{aligned} S_k\left(\boldsymbol{X}\right) &= \max_i\left\{S_{ki}\right\} \\ \delta_k\left(\boldsymbol{X}\right) &= \arg\max_i\left\{S_{ki}\right\}. \end{aligned} \tag{11}$$

Then, for a patient with covariate value $\boldsymbol{X}$ we define the patient score as $\mathbf{U}(\boldsymbol{X}) = (U_1, ..., U_q)'$ where $U_k = (S_k, \delta_k)'$ for $k = 1, ..., q$.

In practice one does not know the error distributions and model functions for models defined in (9) and therefore we cannot directly calculate either the $\boldsymbol{\mu}_i$s or $\boldsymbol{p}_i$s at a given score $\mathbf{u}$. Thus, to apply the proposed selection method, we first need to estimate

components of these vectors using a standard function estimation method. This requires observed Y_{kij} values as well as observed $U_k, k = 1, ..., q$ values corresponding to those responses. However, **U**s defined above are hypothetical scores for a covariate value $\boldsymbol{X}$ as we do not know link functions g_{ki}s and index vectors $\boldsymbol{\beta}_{ki}$s. Hence, in estimating p_{ki}s and μ_{ki}s, we propose to use "estimated" $\mathbf{U}(\boldsymbol{X}_{ij})$ values , $\hat{\mathbf{U}}(\boldsymbol{X}_{ij})$, say, corresponding to responses $Y_{kij}, k = 1, ..., q; j = 1, ..., n_i; i = 1, ..., K$.

Now, to obtain $\hat{\mathbf{U}}(\boldsymbol{X}_{ij})$ values, suitable estimators of link functions g_{ki}s and index vectors $\boldsymbol{\beta}_{ki}$s can be used to construct estimators $\hat{S}_k(\boldsymbol{X})$ and $\hat{\delta}_k(\boldsymbol{X})$ of $S_k(\boldsymbol{X})$ and $\delta_k(\boldsymbol{X})$, respectively. Estimators of the link functions and index vectors can be obtained using responses for each k,$k = 1, ..., q$ coupled with the corresponding covariate observations for any given i, $i = 1, ..., K$, since this estimation amounts to estimating the mean function of a vector random variable with covariates. There is a vast literature on estimating the link function and the index vector of a single index model (see, for example, Ichimura et al., 1993, Hristache et al., 2001, Yu and Ruppert, 2002 and references therein) allowing us to use one out of a several available reasonable estimation methods to estimate the gs and the $\boldsymbol{\beta}$s. We used the procedure given in Hristache et al. (2001) in our simulations and data analysis in the sequel. In the sequel these estimators will be generically denoted by $\hat{g}_{ki}$ and $\hat{\boldsymbol{\beta}}_{ki}$, respectively, for $k = 1, \ldots, q$ and $i = 1, \ldots, K$. In particular, for any

given vector $\boldsymbol{x}$, let

$$\hat{S}_{ki}(\boldsymbol{x}) = \hat{g}_{ki}\left(\hat{\boldsymbol{\beta}}'_{ki}\boldsymbol{x}\right) - \max_{l\neq i}\left\{\hat{g}_{kl}\left(\hat{\boldsymbol{\beta}}'_{kl}\boldsymbol{x}\right)\right\}$$

$$\hat{S}_k\left(\boldsymbol{x}\right) = \max_i\left\{\hat{S}_{ki}(\boldsymbol{x})\right\}$$

$$\hat{\delta}_k\left(\boldsymbol{x}\right) = \arg\max_i\left\{\hat{S}_{ki}(\boldsymbol{x})\right\}$$

and

$$\hat{U}_k(\boldsymbol{x}) = (\hat{S}_k\left(\boldsymbol{x}\right), \hat{\delta}_k\left(\boldsymbol{x}\right))'; k = 1, ..., q. \tag{12}$$

Our development assumes that there is at least one optimal treatment given a score. In the unlikely event that multiple treatments produce the same $\hat{S}_k$ value, we randomly select an index $\hat{\delta}_k$ among the indexes corresponding to that $\hat{S}_k$. Two or more treatments, while dominating all others, vary only by chance for a given score is the same as assuming that the conditional distribution of corresponding responses given the score are identical for those treatments. In that situation, the probability of an index corresponding to those treatments being selected is equal. Furthermore, even if one treatment dominates others for a given covariate value, the conditional means of responses given the score corresponding to that covariate value can be the same for multiple treatments. We have not explored such situations theoretically or empirically in this article.

Now, for a given k, $k = 1, ..., q$, we construct estimators for $\mu_{ki}(u)$ and $p_{ki}(u)$, $i = 1, ..., K$ at a given $u = (s, d)'$ as follows. For estimating μ_{ki} for a given k and i, we first let w be a kernel function with $w \geq 0$ and $\int w(t)dt = 1$, and let $h_i, i = 1, ..., K$ be a set

of smoothing parameters. We define an estimator of $\mu_{ki}(u), i = 1, ..., K$ as

$$\hat{\mu}_{ki}(u) = \frac{\sum_{j=1}^{n_i} Y_{kij} w\left((s - \hat{S}_k(\boldsymbol{X}_{ij}))/h_i\right) I\left(\hat{\delta}_k(\boldsymbol{X}_{ij}) = d\right)}{\sum_{j=1}^{n_i} w\left((s - \hat{S}_k(\boldsymbol{X}_{ij}))/h_i\right) I\left(\hat{\delta}_k(\boldsymbol{X}_{ij}) = d\right)} \tag{13}$$

where $I(A)$ is the indicator of A. The estimator $\hat{p}_{ki}(u)$ of $p_{ki}(u)$ for a given u is obtained by the formula given in Siriwardhana et al. (2017) which is reproduced in the Appendix.

For a realization $\boldsymbol{x}_0$ of the covariate $\boldsymbol{X}$, if we knew the corresponding realizations of the scores, $u_{k0} = (S_k(\boldsymbol{x}_0), \delta_k(\boldsymbol{x}_0))'$, we can estimate $\mu_{ki}(u_{k0})$ and $p_{ki}(u_{k0})$ by $\hat{\mu}_{ki}(u_{k0})$ and $\hat{p}_{ki}(u_{k0})$ respectively. However, due to the aforementioned reasons, we can only find an estimate $\hat{u}_{k0}$ of u_{k0} using (12) above. Thus, we use $\hat{\mu}_{ki}(\hat{u}_{k0})$ and $\hat{p}_{ki}(\hat{u}_{k0})$ as our estimates of $\mu_{ki}(u_{k0})$ and $p_{ki}(u_{k0})$ respectively for $k = 1, ..., q; i = 1, ..., K$.

Finally, for either using means or the probabilities, for a given estimated score vector $\hat{\mathbf{u}}_0 = (\hat{u}_{10}, ..., \hat{u}_{q0})'$, the estimated best treatment for a patient with covariate value $\boldsymbol{x}_0$ is defined via the minimization of

$$\psi(\mathbf{v}) = \sum_{k=1}^{q} \omega_k \gamma(\mathbf{v}, \mathbf{v}_k(\hat{\mathbf{u}}_0)) \tag{14}$$

over P_K and defining

$$\hat{\mathbf{v}^*} = \arg\min_{\mathbf{v} \in P_K} \psi(\mathbf{v}). \tag{15}$$

followed by defining

$$\hat{i}^* = \arg \min_{1 \leq i \leq K} \{\hat{v^*}_i\} \tag{16}$$

where $\hat{v^*}_i, i = 1, ..., K$ are the ranks obtained by the minimization of the distance function (14) for the corresponding procedure.

Remark 2. Using arguments similar to those given in Siriwardhana et al. (2017), we can show that the estimators $\hat{\mu}_{ki}(\hat{u}_{k0})$, $k = 1, ..., q; i = 1, ..., K$ and $\hat{p}_{ki}(\hat{u}_{k0})$ converge in probability to $\mu_{ki}(u_{k0})$ and $p_{ki}(u_{k0})$ respectively as $\min_i n_i \to \infty$. Now, if there are no ties in in the rankings produced by $v_k(u_0)$ for $k = 1, ..., q$ and that the function ψ has a unique minimizer $\mathbf{v}^*$, we can conclude that $\hat{\psi}$ converges in probability to ψ uniformly on the finite set P_K. This gives that $\hat{v^*}$ converges in probability to v^* which in turn implies that $\hat{i^*}$ converges in probability to i^*.

Remark 3. Choice of weights $\omega_k, k = 1, ..., q$ decides the role the vitality of each response in selecting the ideal treatment for a patient. The method we propose is at an individual patient level and hence the importance of each response is also at an individual level. While some responses are clinical/biological measurements, others may be measures that are directly patient related. Thus, while we can always use $\omega_k = 1$ for all k, we offer the user (clinician/patient/medical researcher) the flexibility of selecting the weights for their specific application.

Bandwidth selection for estimating the link functions, p_{ki}s and μ_{ki}s is a challenging issue which has not been investigated in this work. However, methods suggested in Wand and Jones (1995) for kernel smoothing seemed to perform reasonably well in our simulations and data analysis.

3. Empirical Studies

In this section we present a simulation study that investigates the properties of the proposed procedure in finite samples.

Although it is impossible to obtain counterfactual data in real life, one can simulate them under reasonable assumptions. Simulations in Siriwardhana et al. (2017) showed that, in the single response case, the selection using estimated p_{1i}s based on observations from marginal distributions of $(Y^*_{1i}, \boldsymbol{X})$, $i = 1, ..., K$ was very effective compared with the selection using estimated p^*_{1i}. Given that one cannot obtain real life data to estimate p^*_{1i}s, the surrogate method of using p_{1i}s to select treatments was therefore recommended in that article.

Guided by those results, in this study we primarily focused on the accuracy of treatment assignment of a new (test) observation using estimated values of μ_{ki} and p_{ki} functions from a set of training data obtained from marginal distributions of $(\mathbf{Y}^*_i, \boldsymbol{X})'$. This simulation study was performed for treatment groups $K = 2$ and 3 with response dimension $q = 2, 3$ and 4. We selected our model sets such that each model in a set dominates other competing models for some combination of covariate values; in other words, none of considered models fully dominate other models within the whole covariate space. This signifies, subjects with distinct covariates vectors, could experience corresponding highest response from different treatments illustrating the personalized medicine concept.

In our study, we first simulated K independent multivariate (dimension q) samples of size n ($n = 100$ or $n = 200$) per group. The components of the r dimensional covariate vectors $\boldsymbol{X}$ were generated from a r dimensional multivariate normal distribution with zero mean and a covariance matrix with the ijth element equal to $\rho^{|i-j|}$ where ρ was chosen from the set $\{0.1, 0.5, 0.9\}$ and r was chosen as $3, 8$ and 10. Using various link functions and index vectors, we obtained the treatment responses from model (9) for each i. Here, for a given i, $i = 1, ..., K$, the errors were generated from either a q dimensional multivariate normal distribution or a multivariate double exponential distribution with zero mean and a correlation matrix with the ijth element given by $\xi^{|i-j|}$ where ξs were chosen from the set $\{0.1, 0.5, 0.9\}$. The R package *mvtnorm* (Genz et al., 2015) was used for the generation of these random vectors where in the multivariate normal case, the dispersion parameter σ was chosen from the set $\{0.1, 0.3, 0.5\}$. We used the package *LaplacesDemon* (Hall et al., 2106) for generating double exponential random variables with dispersion parameters 0.1 and 0.5. We examined the performance of the proposed methodology under a variety of both linear and nonlinear regression models with the Single Index Model (SIM) structure.

Once K samples were generated, we estimated the corresponding SIMs followed by an estimation of scores at each covariate value. SIMs were estimated by the procedure given in Ichimura et al. (1993) using Gaussian kernels. Then, a new covariate value $\boldsymbol{X}_0$ was generated in the same manner as previous covariates above, and for its corresponding estimated score $\hat{\mathbf{u}}_0$, we calculated $\hat{\mu}_{ki}(\hat{u}_{k0})$ and $\hat{p}_{ki}(\hat{u}_{k0})$ for $k = 1, ..., q; i = 1, ..., K$

and the corresponding $\hat{i}^*$ values for equal weights ($\omega_k = 1$ for all k) cases and few unequal weights cases. The kernel function in this estimation was taken to be a $U(-1, 1)$ probability density function. We chose all bandwidths by the algorithm given by Wand and Jones (1995) for each $i, i = 1, ..., K$.

Next, we generated K new response vectors, $\mathbf{Y}_i^{**}, i = 1, ..., K$, each with mean vector $(g_{1i}(\boldsymbol{\beta}_{1i}'\boldsymbol{X}_0), ..., g_{qi}(\boldsymbol{\beta}_{qi}'\boldsymbol{X}_0))'$ for $i = 1, ..., K$, corresponding to this $\boldsymbol{X}_0$ using model (9) where the errors were generated independently from the same error distribution that was used to generate the K original samples. Then we obtain rank vectors $\tilde{\mathbf{v}}_k$,$k = 1, ..., q$, say, for each row of the data matrix $(\mathbf{Y}_1^{**}, ..., \mathbf{Y}_K^{**})$, and minimize

$$\psi(\mathbf{v}) = \sum_{k=1}^{q} \omega_k \gamma(\mathbf{v}, \tilde{\mathbf{v}}_k) \tag{17}$$

over P_K for same corresponding weights $\omega_k = 1$ above to get the corresponding aggregated vector $(\hat{v^*}_1, ..., \hat{v^*}_K)'$ and define the treatment assignment to be correct if

$$\hat{i}^* = \arg \min_{1 \le k \le K} \{\hat{v^*}_k\}$$

for the $\hat{i}^*$ corresponding to the criteria using $\hat{\mu}_{ki}$s or $\hat{p}_{ki}$s.

We repeated this procedure 1000 times for each model and error distribution combination. Frequencies of correct treatment assignments for a representative set of cases are given in the Tables (2) and (3). We provide additional results in several tables in the supplementary material. The results presented below are for model functions and

index vectors given in Table (1) below and for covariate dimension $r = 10$.

An examination of these tables reveal that the selection accuracy drops when the number of treatments increase as well as when the error distribution has a high variability. The method based on smoothed means and the method based on p_{ki} both have comparable selection frequencies in all cases where the frequency appears to be marginally higher with p_{ki} for a few cases. This pattern was seen even in the single dimensional response case. The selection frequency appears to be slightly lower when the covariate correlation is higher although the drop is minimal in most cases. In all simulations, model functions were chosen so that they would dominate all other model functions at some covariate values. We have not investigated cases where two or more models were identical and dominating all others because in that case those dominating models will have equal probability of being selected.

In the supplementary material, we also provide frequencies of overlap between treatment assignments by methods based on p_{ki}'s and μ_{ki}'s for a selected case. This table indicates high agreement between the two approaches.

4. ACTG-175 HIV Clinical Trial

In this section we illustrate our proposed method using a real clinical trial dataset.

The data resulted from the ACTG 175 clinical trial (Hammer et al. 1996). This trial

was a randomized, double-blinded, placebo-controlled clinical trial that was conducted for comparing antiviral medications for HIV-1 patients whose T-cell CD4 counts were in the range of 200 to 500 per cubic millimeter. The dataset (Juraska et al. 2012) contains information on 2136 HIV-1 infected individuals who were randomized into four treatment arms; those treated with Zidovudine (arm-0), combination of Zidovudine and Didanosine (arm-1), combination of Zidovudine and Zalcitabine (arm-2), Didanosine (arm-3). Arms 0, 1, 2, and 3 contain 532, 519, 524, and 561 patients, respectively. T cell CD4 and CD8 are critical components in the human immune system. Frequently, the severity of HIV progression is measured through a decline in CD4 counts.

Clearly, the original study was a randomized clinical trial where no optimality was considered when patients were assigned to treatments. This trial periodically measured both these cell counts for each patient and defined a survival time based on those counts as the end point which was severely censored. Our intention in this data analysis is merely to illustrate what would be the optimal treatment for a new patient if one were to use a formal strategy like ours for treatment selection based on individual patient characteristics when training samples of uncensored responses with corresponding covariate values for multiple treatments are available in advance. In this situation, rather than assuming there is a standard treatment that is being compared with experimental treatment(s) as in a typical clinical trial, we take the stance that given multiple treatments can be used to treat a patient, what would be optimal for the individual based on his/her characteristics.

Between CD4 and CD8 T cell types, the scientific literature on HIV/AIDS often declare CD4 cell as the primary T cell type that is suppressing the HIV cell replication, whereas the critical role of the CD8 cell is typically referred to as the antibody reaction against cancers and various types of other viruses. However some studies have illustrated the important role of the CD8 during early stages of HIV progression (Eg: Streeck and Nixon, 2010).

In our analysis, we considered the log transformed CD4 and CD8 counts of a patient after 20 weeks of treatment as the bivariate clinical response. As covariates, we used log-CD4, log-CD8 counts at baseline, age, weight, and the number of months a patient received the pre-antiviral therapy. All these variables were significant in linear models for the counts after 20 weeks for one or more arms.

We fist selected 200 patients at random from each group and considered their data as outcomes of a clinical trial in which those patients were randomly assigned to one of the four treatments. Based on these data we estimated the corresponding SIMs for the response pair $\log \mathrm{CD4}$, and $\log$ CD8 coupled with above covariates. Then, pretending that the remaining 1336 cases as "new" patients, we applied the proposed treatment selection for those cases using combinations of weights ranging from 0 to 1 for each response.

We summarize the resulted assignments for test patients set in Table (4). For example, when we chose CD4:CD8 weights to be 0.6:0.4, using the proposed treatment

selection based on p_{ki}'s, 56, 815, 182, and 283 test patients are proposed to be assigned to arms 0, 1, 2, and 3 respectively, whereas the corresponding assignment using μ_{ki}'s is 11, 814, 368, and 143. The overall pattern of these assignments indicates that the majority of the 1336 test patients are proposed to be assigned to arm-1. Especially this number is relatively high in cases where a more priority is given to CD4 counts. In the supplementary materials, we also present the overlap in the number of cases by p_{ki}'s and μ_{ki}'s.

Furthering our analysis we conducted the following study to examine the effectiveness of proposed treatment selection for test patients compared to their original random assignment.

Consider the t th subject in the test set with covariate value $\boldsymbol{X}_t$ who is assigned to a particular arm by the random assignment mechanism. Suppose the individual's estimated score is $\hat{u}_t = (\hat{u}_{t1}, \hat{u}_{t2})$, where $\hat{u}_{t1} = (s_1, d_1)$ and $\hat{u}_{t2} = (s_2, d_2)$ are the marginal estimated scores based on CD4 and CD8 responses respectively. Let v be the group our procedure would assign this patient based on his/her estimated score $\hat{u}_t$ and let v_0 be the treatment group patient was assigned in the original trial by the random assignment. We estimated the average conditional gain in log-CD4 and log-CD8 values for an individual, say Δ_t^{CD4} and Δ_t^{CD8}, by the proposed assignment versus his/her original assignment. We define Δ_t^{ind} where $ind = CD4$ and $CD8$ to be,

$$\Delta_t^{ind} = E(Y_{iv}|\hat{u}_i) - Y_{iv_0},$$

where Y_{iv_0} is the observed value by the original assignment. We estimated $E(Y_v|\hat{u}_j)$ using a smooth mean estimator given by,

$$E(Y_{iv}|\hat{u}_i) = \frac{\sum_{j=1}^{n_v} Y_{ivj} \prod_{l=1}^{2} w\left((s_l - \hat{S}_l(\boldsymbol{X}_{vj}))/h_{lv}\right) I\left(\hat{\delta}_l\left(\boldsymbol{X}_{vj}\right) = d_l\right)}{\sum_{j=1}^{n_v} \prod_{l=1}^{2} w\left((s_l - \hat{S}_l(\boldsymbol{X}_{vj}))/h_{lv}\right) I\left(\hat{\delta}_l\left(\boldsymbol{X}_{vj}\right) = d_l\right)}, \tag{18}$$

where n_v is the number of patients originally assigned to treatment arm v, $v = 0, .., 3$, w is a kernel function with $w \geq 0$ and $\int w(t)dt = 1$, and h_{lv}'s, $l = 1, 2$ are corresponding smoothing parameters. We chose w to be a Gaussian kernel and the corresponding h_{lv}'s were determined by the method suggested in Wand and Jones (1995) for kernel smoothing.

Finally we estimate the overall treatment selection efficiency as the averaged $\hat{\Delta}^i$ for all test patients,

$$\hat{\Delta}^{ind} = \frac{1}{N} \sum_{i=1}^{N} \hat{\Delta}_t^i, \tag{19}$$

where N is the number in the test set and where $ind = CD4$ and $CD8$. Note that positive values for $\hat{\Delta}_t^{CD4}$ and $\hat{\Delta}_t^{CD8}$ indicate overall effective treatment selection compared to the original random assignment.

In this portion of the analysis we used a higher weight for the CD4 response based on suggestions in the literature. Table (5) provides these $\hat{\Delta}_t^{CD4}$ and $\hat{\Delta}_t^{CD8}$ values for the proposed procedure with p_{ki}'s and μ_{ki}'s. These observed positive values for $\hat{\Delta}_t^{CD4}$ and $\hat{\Delta}_t^{CD8}$ seemingly suggesting that if one were able to use prior data to estimate the personal score for a new patient, then the proposed assignment is more beneficial for

him/her than an assignment based on a clinical trial which may only find the best treatment for the general population.

5. Discussion

In this article we proposed a novel personalized treatment plan to select the optimal treatment from a set of multiple treatments when the outcome measures are multivariate. Our method assumes that there are prior data with respect to each treatment and we try to garner information from that data to determine the best option for an individual. This method is a single step procedure which can be easily applied. Our empirical studies show that the proposed method performs very satisfactorily in selecting the optimal treatment in a multiple treatment setting. Our analysis of a real clinical trials dataset which has the multiple treatment option reveals possible changes if one were to use multiple outcome measures as opposed to a single measure.

The proposed method is based on semi parametric Single Index Models which add great flexibility in modeling real life situations. In practice, there may be occasions where one has to deviate from Gaussian framework due to issues with error structures. In such situations, the proposed method can also be applied using quantile regression SIMs providing additional model flexibility compared with existing methods based on conditional expectations.

As seen in simulations and real data analysis, the choice of ωs can be influential in the

selection. We feel this is more of a subjective/qualitative issue than a quantitative issue where one can guide the choice of optimal weights based on patient qualities and advise by clinicians. If all the responses are related to factors that govern the progression of a patient, then a weight selection using more of a data based method might be appropriate. While it is a worthwhile endeavor to find such data based guidelines, we have not investigated those aspects here.

Also, this article deals with complete responses. However, censoring is very common in practice. An extension of the proposed methodology to a covariate dependent censoring setting and various lifetime aspects such as multi state models is forthcoming.

Response	Mean functions		
	Group-1	Group-2	Group-3
1	$2\sin\left\{\pi\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$	$2\sin\left\{\frac{\pi}{4}+\pi\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$	$2\sin\left\{\frac{\pi}{6}+\pi\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$
2	$2\cos\left\{\pi\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$	$2\cos\left\{\frac{\pi}{4}+\pi\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$	$2\cos\left\{\frac{\pi}{6}+\pi\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$
3	$2\sin\left\{\frac{\pi}{2}\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$	$2\sin\left\{\frac{\pi}{4}+\frac{\pi}{2}\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$	$2\sin\left\{\frac{\pi}{6}+\frac{\pi}{2}\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$
4	$2\cos\left\{\frac{\pi}{2}\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$	$2\cos\left\{\frac{\pi}{4}+\frac{\pi}{2}\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$	$2\cos\left\{\frac{\pi}{6}+\frac{\pi}{2}\left(\boldsymbol{C}'\boldsymbol{X}\right)\right\}$

Table 1: Sets of mean functions used to generate treatment responses, with $\boldsymbol{C'} = \left(1/\sqrt{10},...1/\sqrt{10}\right)_{1\times 10}$.

Error dist.	Sample size per group	Error dist. parameter	Error correlation	Correlation between covariates					
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
				Prob of dominance	Smooth means	Prob of dominance	Smooth means	Prob of dominance	Smooth means
Normal	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	786	695	749	655	733	675
			$\rho_e = 0.5$	792	686	767	687	725	681
			$\rho_e = 0.9$	788	677	769	679	714	632
		$\sigma = 0.5$	$\rho_e = 0.1$	514	519	491	494	484	477
			$\rho_e = 0.5$	459	452	512	493	468	469
			$\rho_e = 0.9$	485	471	429	445	439	425
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	822	740	850	744	799	731
			$\rho_e = 0.5$	853	758	820	722	824	753
			$\rho_e = 0.9$	882	755	831	752	813	731
		$\sigma = 0.5$	$\rho_e = 0.1$	550	549	514	525	519	531
			$\rho_e = 0.5$	521	530	513	517	521	512
			$\rho_e = 0.9$	493	497	505	498	506	495
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
DE	$n = 100$	$\sigma_{DE} = 0.1$	$\rho_e = 0.1$	783	702	749	683	694	644
			$\rho_e = 0.5$	796	686	751	699	706	645
			$\rho_e = 0.9$	787	718	752	671	700	643
		$\sigma_{DE} = 0.5$	$\rho_e = 0.1$	507	518	480	521	504	512
			$\rho_e = 0.5$	526	503	524	494	474	470
			$\rho_e = 0.9$	522	512	479	454	479	472
	$n = 200$	$\sigma_{DE} = 0.1$	$\rho_e = 0.1$	869	749	838	726	784	734
			$\rho_e = 0.5$	865	749	813	738	793	732
			$\rho_e = 0.9$	836	738	826	725	789	724
		$\sigma_{DE} = 0.5$	$\rho_e = 0.1$	536	560	528	534	534	558
			$\rho_e = 0.5$	541	536	530	513	516	512
			$\rho_e = 0.9$	552	520	554	545	543	511

Table 2: Frequencies of correct treatment assignments in 1000 test cases by the proposed method. Three treatments with four responses and $w_k = 1; k = 1, ..., 4$

Error dist.	Sample size per group	Error dist. parameter	Error correlation	Correlation between covariates					
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
				Prob of dominance	Smooth means	Prob of dominance	Smooth means	Prob of dominance	Smooth means
Normal	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	797	735	771	731	703	649
			$\rho_e = 0.5$	825	747	761	702	688	677
			$\rho_e = 0.9$	797	710	755	692	678	640
		$\sigma = 0.5$	$\rho_e = 0.1$	516	508	465	475	478	465
			$\rho_e = 0.5$	482	472	475	478	422	410
			$\rho_e = 0.9$	504	508	465	469	451	445
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	880	819	858	790	821	765
			$\rho_e = 0.5$	858	788	840	749	800	748
			$\rho_e = 0.9$	864	793	843	764	802	738
		$\sigma = 0.5$	$\rho_e = 0.1$	579	588	550	549	514	515
			$\rho_e = 0.5$	519	522	531	513	517	503
			$\rho_e = 0.9$	523	527	514	510	491	490
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
DE	$n = 100$	$\sigma_{DE} = 0.1$	$\rho_e = 0.1$	809	735	755	691	713	649
			$\rho_e = 0.5$	801	729	742	706	703	663
			$\rho_e = 0.9$	807	742	745	697	684	611
		$\sigma_{DE} = 0.5$	$\rho_e = 0.1$	503	491	489	486	471	475
			$\rho_e = 0.5$	505	522	491	493	449	447
			$\rho_e = 0.9$	524	503	451	456	467	468
	$n = 200$	$\sigma_{DE} = 0.1$	$\rho_e = 0.1$	878	783	835	755	819	753
			$\rho_e = 0.5$	858	798	841	756	809	761
			$\rho_e = 0.9$	859	781	855	760	814	768
		$\sigma_{DE} = 0.5$	$\rho_e = 0.1$	519	517	537	532	547	542
			$\rho_e = 0.5$	543	545	520	525	532	530
			$\rho_e = 0.9$	565	553	556	545	510	515

Table 3: Frequencies of correct treatment assignments in 1000 test cases by the proposed method, for three treatments with four responses, using weights $\omega_1 = 0.4$, $\omega_2 = 0.3$, $\omega_3 = 0.2$, and $\omega_4 = 0.1$, for responses 1, 2, 3, and 4, respectively.

Weights		Probability of Dominance				Smooth Means			
ω_{CD4}	ω_{CD8}	Arm-0	Arm-1	Arm-2	Arm-3	Arm-0	Arm-1	Arm-2	Arm-3
1	0	59	760	252	265	4	731	426	175
0.8	0.2	59	760	252	265	4	731	426	175
0.6	0.4	56	815	182	283	11	814	368	143
0.5	0.5	285	710	168	173	132	867	261	76
0.4	0.6	265	571	161	339	130	755	224	227
0.8	0.2	310	476	109	441	242	547	181	366
0	1	311	476	109	440	242	547	181	366

Table 4: Treatment assignment summary for ACTG-175 clinical trial data, by the proposed method selecting both CD4 and CD8 counts as clinical response, using weights ω_{CD4} and ω_{CD8} for CD4 and CD8 counts, respectively.

Weights		Probability of Dominance		Smooth Means	
$\hat{\omega}_{CD4}$	$\hat{\omega}_{CD8}$	$\hat{\Delta}^{CD4}$	$\hat{\Delta}^{CD8}$	$\hat{\Delta}^{CD4}$	$\hat{\Delta}^{CD8}$
0.8	0.2	0.0598	0.0282	0.0672	0.0215
0.6	0.4	0.0595	0.0328	0.0624	0.0255

Table 5: The estimated gain in average log-CD4 ($\hat{\Delta}^{CD4}$) and log-CD8 ($\hat{\Delta}^{CD8}$) by the proposed method compared to the original assignment for test patients, using weights ω_{CD4} and ω_{CD8} for CD4 and CD8 counts, respectively.

Appendix

In this section we give the formula for estimating p_{ki}. First let

$$\mathcal{J} = \{(j_1, \ldots, j_K) \,|j_l \in \{1, \ldots, n_l\}, l = 1, \ldots, K\}\,.$$

and, for $J \in \mathcal{J}$ we let

$$\hat{w}_J^i(s) = \prod_{l=1}^{K} \frac{1}{h_l} w\left(\frac{s - \hat{S}_i(\boldsymbol{X}_{lj_l})}{h_l}\right)$$

Also, let

$$\hat{\eta}_J^i(d) = \prod_{l=1}^{K} I\left(\hat{\delta}_i(\boldsymbol{X}_{lj_l}) = d\right).$$

Now, the estimator of p_{ik} is defined as

$$\hat{p}_{ki}(u) = \frac{\sum_{J\in\mathcal{J}} I\left[Y_{kj_k} > \max_{k\neq l}\{Y_{lj_l}\}\right] \hat{w}_J^i(s)\,\hat{\eta}_J^i(d)}{\sum \hat{w}_J^i(s)\,\hat{\eta}_J^i(d)}. \tag{20}$$

References

Cai, T., Tian, L., Wong, P. H., and Wei, L. J. (2011),"Analysis of Randomized Comparative Clinical Trial Data for Personalized Treatment Selections", *Biostatistics*, pp. 270-282.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Bornkamp, B., Maechler, M., and Hothorn, T. (2015), "mvtnorm: Multivariate Normal and t Distributions", R package version 1.0-3.

Hall, B., Hall, M., Statisticat, L. L. C., Brown, E., Hermanson, R., Charpentier, E., and Singmann, H. (2016), "LaplacesDemon: a complete environment for Bayesian inference within R", R package version 16.0.1.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair J. P., Niu M. , Hirsch M. S., and Merigan T. C. (1996), "A Trial Comparing Nucleoside Monotherapy with Combination Therapy in HIV-Infected Adults with CD4 Cell Counts from 200 to 500 per Cubic Millimeter", *New England Journal of Medicine*, pp. 1081-1090.

Hristache, M., Juditsky, A., Polzehl, J., and Spokoiny, V. (2001), "Structure Adaptive Approach for Dimension Reduction", *The Annals of Statistics*. pp. 1537-1566.

Ichimura, H., Hall, P., Hardle, W. (1993), "Optimal smoothing in single index models", *The Annals of Statistics*, Vol. 21, No 1, pp. 157-178.

Kelly, Zo L, Agnieszka Michael, Butler-Manuel,Simon, Pandha,Hardev S, and Morgan, Richard G.L. (2011), "HOX genes in ovarian cancer", *Journal of Ovarian Res.*, pp 4-16.

Juraska, M., Gilbert, P. B., Lu, X., Zhang, M., Davidian, M., and Tsiatis, A. A. (2012), "speff2trial: Semiparametric Efficient Estimation for a Two-Sample Treatment Effect". R package version 1.0.4.

Liang, H., Liu, X., Li, R., amd Tsai, C-H. (2010), "Estimation and Testing for Partially Linear Single Index Models", *Annals of Statistics*, pp. 3811-3836.

Pihur, V., Datta, S., and Datta, S. (2007) "Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach" *Bioinformatics*, 23(13):1607-1615.

Pihur, V., Datta, S., and Datta, S. (2009) "RankAggreg, an R package for weighted rank aggregation" *BMC Bioinformatics*, 10:62.

Qian, M., and Murphy, S. A. (2011), "Performance Guarantees for Individualized Treatment Rules." *The Annals of Statistics*, pp. 1180-1210.

Schulte, P.J., Tsiatis, A.A., Laber, E.B., and and Davidian, M. (2014), "Q- and A-learning Methods for Estimating Optimal Dynamic Treatment Regimes", *Statistical Science*, To Appear.

Siriwardhana, C., Zhao, M., Datta, S., and Kulasekera, K.B. (2015),

"Personalized Plans with Multiple Treatments", Technical Report: http://louisville.edu/sphis/departments/bioinformatics-biostatistics/pdfs/BST2017-01.pdf, *Department of Bioinformatics and Biostatistics, The University of Louisville.*

Streeck, H. and Nixon, D. F. (2010), "T Cell Immunity in Acute HIV-1 Infection", *The Journal of Infectious Diseases*, 202(S2), pp. 302-308.

Stute, W (1991), "Conditional U-Statistics", *Annals of Probability*, pp. 812-825.

Robins J. (2004), "Optimal structured nested models for optimal sequential decisions", In: Lin DY, Heagerty PJ, editors. *Proceedings of the Second Seattle Symposium on Biostatistics. New York: Springer*, pp. 189326.

Robins,J., Orellana, L., and Rotnitzky, A. (2008), "Estimation and extrapolation of optimal treatment and testing strategies", *Statist. Med.*, pp. 46784721

van't Veer LJ, Bernards R.(2008), "Enabling Personalized Cancer Medicine Through Analysis of Gene-Expression Patterns", *Nature*, Apr 3;452(7187), pp. 564-70.

Vazquez, A. (2013), "Optimization of Personalized Therapies for Anticancer Treatment", *BMC Systems Biology*, pp. 7-31

Wand, M. P. and Jones, M. C. (1995), "Kernel Smoothing", *Chapman and Hall*, London.

Wang, L. and Yang, L. (2009), "Spline Estimation of Single Index Models", *Statistica Sinica*, 19, pp 765-783

Yu, Y. and David Ruppert, D. (2002), "Penalized Spline Estimation for Partially Linear Single-Index Models", *Journal of the American Statistical Association*, pp. 1042-1054.

Zhang, B., Tsiatis A. A., Laber, E. B., and Davidian, M. (2012), "A Robust Method for Estimating Optimal Treatment Regimes", *Biometrics*, pp. 1010-1018.

Zhao, Y., Donglin, Z., Rush, A. J., and Kosorok, M. K. (2012), "Estimating Individualized Treatment Rules Using Outcome Weighted Learning", *Journal of the American Statistical Association*, pp. 1106-1118.