5-2016

Semi-parametric methods for personalized treatment selection and multi-state models.

Chathura K. Siriwardhana
University of Louisville

SEMI-PARAMETRIC METHODS FOR PERSONALIZED TREATMENT SELECTION AND MULTISTATE MODELS

By

Chathura Siriwardhana
MSc, University of Peradeniya, 2012

A Dissertation
Submitted to the Faculty of the
School of Public Health and Information Sciences
of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in
Biostatistics: Decision Science

Department of Bioinformatics and Biostatistics
University of Louisville
Louisville, KY 40202

May 2016

SEMI-PARAMETRIC METHODS FOR PERSONALIZED TREATMENT SELECTION AND MULTISTATE MODELS

Submitted by

Chathura Siriwardhana

A Dissertation Approved on

April 14th 2016

by the Following Dissertation Committee:

K.B Kulasekera, Ph.D
Dissertation Director

Somnath Datta, Ph.D
Dissertation Co-Director

Rodney Folz, MD, PhD

Jeremy Gaskins, PhD

Riten Mitra, PhD

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my dissertation directors Drs. K.B Kulasekera and Somanth Datta for their excellent guidance towards the completion of my dissertation. I also thank my dissertation committee members Drs. Rodney Folz, Jeremy Gaskins, and Riten Mitra for their continuous support and direction.

I am especially grateful for my parents and sisters who have given me immense encouragement throughout my whole academic life.

I am earnestly thankful to the Department of Bioinformatics and Biostatistics and Chronic Disease Management Program at University of Louisville for providing the financial support throughout the Ph.D program. My sincere thank to Dr. Meng Zhao for assisting with theoretical proofs in project-1. I would like to thank Dr. Doug Lorenz for providing Spinal Cord Injury Data. Finally, I thank the all faculty and students of the Department of Biostatistics and Bioinformatics as well as the administrative staff for their support, direction and collaboration.

ABSTRACT

SEMI-PARAMETRIC METHODS FOR PERSONALIZED TREATMENT SELECTION AND MULTISTATE MODELS

Chathura Siriwardhana

April 14th 2016

This dissertation contains three research projects on personalized medicine and a project on multi-state modelling.

The idea behind personalized medicine is selecting the best treatment that maximizes interested clinical outcomes of an individual based on his or her genetic and genomic information. We propose a method for treatment assignment based on individual covariate information for a patient. Our method covers more than two treatments and it can be applied with a broad set of models and it has very desirable large sample properties. An empirical study using simulations and a real data analysis show the applicability of the proposed procedure.We then extend this idea for treatment section for survival outcomes under right-censoring by introducing re-weighted estimation to adjust the bias caused by censoring. Series of empirical studies using simulations show the desirable performance of re-weighted estimation concept in treatment selection in finite sample cases. We provide a real data application of the proposed procedure to illustrate the applicability for right-censored data. Next we propose a novel method for individualized treatment selection when the treatment response is multivariate. 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 very desirable performances of the proposed method in finite sample cases. We also present a data analysis using a HIV clinical trial data to show the applicability of the proposed procedure for real data.

Multi-state models are extensions of simple survival models that incorporate the progression of a subject in an interconnected system such as a disease network. An important measure arising from a mutistate model is the subjects' state occupational probabilities given baseline covariates. In the final portion of this dissertation we introduce an inverse censoring probability re-weighted semi-parametric single index model based approach to estimate conditional state occupation probabilities of a given individual in an acyclic multistate model under right-censoring. Besides obtaining a temporal regression function, we also test the potentially time varying effect of a baseline covariate on future state occupations. We show that the proposed technique has desirable finite sample performances. Its performance is competitive when compared with two other existing approaches. We illustrate the proposed methodology using two different data sets. First we re-examine a well known data set on various event times tracking the progression of a sample of leukemia patients undergoing bone marrow transplant. Our second illustration is based on the functional status of a set of spinal cord injured patients undergoing a rehabilitation program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Personalized Medicine

This section introduces three projects focused on different aspects in personalized medicine. In particular we address the following topics.

- Personalized plans with multiple treatments
- Personalized treatment selection for survival outcome
- Personalized treatment plans with multivariate outcome measures

We provide introductory outlines of these projects as below.

1.1.1 Personalized plans with multiple treatments

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 for a patient based on his/her genetic and genomic markers are of high importance. Rather limited literature on this topic mainly deals with deciding between two treatments based on patient characteristics. Assuming without any loss of generality that a larger outcome is desirable, the methods developed in the literature essentially determine the larger conditional expectation

of the outcome given the set of markers for the patient. Cai et al. (2011) use a smoothed sub-group mean in the comparison of two treatments. Here the sub-groups are determined via a set of contours (scores) that define overall similarities among patients. For continuous responses, these scores have been defined via linear models. Qian and Murphy (2011) discuss a two step procedure that is based on an estimation of a conditional mean followed by a maximization of that mean over a set of possible treatments. In a different approach for treatment assignments, Zhao et al. (2012) consider an optimization technique to select between two treatments where the binary optimization procedure is within a class of pre-specified model functions. Drawing parallels to the support vector machine technology, these authors show decision optimality of the treatment selection procedure within the binary framework showing that the procedure discussed in Qian and Murphy (2011) is inferior to theirs in the two treatments case. In a more recent article, Zhang et al. (2012) use a robust conditional mean estimation method to alleviate possible wrong model postulation when one estimates the conditional mean for each patient's profile. Schulte et al. (2014) provide details of using Quality learning (Q-learning) and Advantage learning (A-learning) concepts in devising sequential rules based on a set of pre-specified decision points. The optimality of the decision algorithm, based on the conditional sequential mean, has been discussed by these authors. While mathematically and computationally tedious, it gives a sequential decision rule that self updates the changing patient behavior in switching to a different treatment. Additional references on dynamic treatment regimes can be found in Schulte et al. (2014). Treatment selection based on observational studies has been treated by many authors. Readers are referred to Robins et al. (2004, 2008) and references therein for additional details of such procedures.

In many treatment selection situations clinicians have more than two treatments to select from and the decision of assigning the treatment protocol based on

individual patient characteristics is highly desirable. In this work, we discuss the K treatment ($K \geq 2$) scenario where we compare quantities that are suitable approximations to true conditional probabilities of outcome variable of each treatment dominating other treatments given patient specific scores constructed from covariates. In particular, instead of estimated marginal conditional expectations, we examine estimated conditional probability of each treatment dominating the others based on K independent pairs of outcomes and covariates, one for each treatment. We choose the optimal treatment as the one that has the highest estimated probability of dominating every one else for a given patient score. This allows one to compare treatments for a wide variety of distributions of outcome measures. As seen in our empirical investigations, the performance of this method is comparable to selection using conditional means when responses have finite means. In our approach, scores are defined via a set of Single Index Models (SIMs) or Partially Linear Single Index Models (PLSIMs) and our scoring system simplifies to the same type of scoring as in Cai et al. (2011) if $K = 2$. The method we propose is general where the above SIMs (PLSIMs) used to obtain scores can be quantile regression models rather than mean regression models, thus allowing a broad class of structures to get a suitable score. Empirical evaluations of this new mechanism using a detailed simulation study to assess the accuracy of treatment selection show that the proposed method is comparable with existing methodology in the two treatment option with linear models, it has a higher accuracy in the two treatment case with SIMs and performs very well in the multiple treatment case. Furthermore, we applied our method to an existing dataset with multiple treatment arms to examine the use of treatment assignment based on patient characteristics. The results show that one arm is highly preferred over the others for patients in this study with respect to the primary outcome variable which was a blood count. We also assessed possible gains or losses of patient survival had the patients were assigned according to the rule proposed here.

Interestingly our study reveals that there could have been an advantage in terms of survival also to have used our selection method in the treatment assignment.

1.1.2 Personalized treatment selection for survival outcome

In many severe illnesses, for examples in cancer or HIV, a patients' survival time is usually considered as the primary clinical outcome when one is investigating different treatment options/protocols. When the clinical response of interest is a survival outcome, patient survival times are often subject to censoring due to dropouts, competing risks or administrative reasons forcing treatment selection methods developed for complete observations to become inapplicable. Exiting literature on personalized treatment rules is often limited to completely observed responses. In this part of our research we develop a treatment selection method that addresses the multiple treatments selection issue with right-censored survival outcome.

Frequently, censoring adds an additional complexity to any statistical problem. It becomes more complicated when the censoring mechanism is a non-random process. Variety of methods have been developed to address this issue. Often, in regression methodology, the bias caused by partially observed observations is handled by utilizing a weighting scheme; specifically the inverse censoring probability weighting (ICPW), a widely applied technique in such problems.

This idea was first developed by Koul et. al (1981) for survival outcomes in regression. Subsequently this idea has been widely applied in many survival related studies. Robins and Rotinizky (1992), and Robins (1993) discussed a new class of tests and estimators for Cox model, accelerated failure time models and a model for the mean treatment effect, in case of dependent censoring, using a re-weighted scheme. Satten and Datta (2001) derived Kaplan-Meier estimator as an IPCW average. Satten et al. (2001) implemented this idea to estimate the marginal survival

function in the presence of time dependent covariates, calculating covariate dependent censoring probabilities using Aalens additive hazard model (Aalen, 1980, 1989). Datta and Satten (2002) estimated integrated transition hazards and stage occupation probabilities for non-Markov systems under dependent censoring utilizing IPC weights calculated using Aalen's linear model. Similar to these studies, in this work, we use a weighting scheme in all estimation steps of personalized treatment selection methods proposed above when responses are censored. This involves two steps: first estimating treatment specific SIM models to calculate patients' scores followed by estimating the probability of one treatment dominating all others for a given score. Lopez et al. (2013) discussed adjusting the single index estimator in the case of right censored observations using Cox model based weights for covariate dependent censoring. Our proposed re-weighted single index estimator is an extension of Ichimura et al. (1993) single index estimator, weighted by IPCW obtained by Aalens additive hazard model under covariate dependent censoring. In the case of random censoring, a Kaplan-Meier estimator based weighting scheme can be used. In the same fashion a new re-weighted estimator is proposed to estimate the treatment selection probability. We evaluate the re-weighted single index estimator and the proposed treatment selection concept via an extensive empirical study. We compare our method with an alternative method based on Cox's (1972) approach. To demonstrate practical applications, we apply this method to a real dataset with multiple treatments, where the survival outcome is heavily right-censored.

1.1.3 Personalized treatment plans with multivariate outcome measures

Current methods in personalized medicine 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 (Cai et al., 2011; Qian and Murphy, 2011; Zhao et al. 2012; Zhang et al. 2012; Zhao et al., 2015) 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, which described in the first project.

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 work 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 (eg: Zhang et al. 2012) or quantities based on conditional probability of one treatment dominating others given the patient score, as described in projects 1 and 2. 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.

1.2 Multistate models

In this section we introduce a project on estimating conditional state occupation probabilities of an individual given covariates under right-censored data in a disease network.

1.2.1 Flexible semi-parametric regression of state occupational probabilities in a multistate model with right-censored data

Multi-state models represent subjects' movement along time in terms of state occupation starting from an initial state to a final (absorbing) state. It can be a simple survival model that describes transition between two states or a more complex model which contains several intermediate and final states. The well studied bone marrow transplant data described by Copelan et al. (1991) is an example of such a system, which illustrates the transition of acute leukemia subjects in numerous different clinical states in time, after the bone marrow transplant. An important in-

vestigation for such a system centers around future state occupancy of an individual at a specific time since enrollment into the system given subject specific information. In reality, the complete movement of a subject in a multistate model may not fully be observable due to censoring, which restricts of using typical regression concepts for this problem. In this work, we propose a new method to estimate the conditional state occupation probability given a subject's covariates, in the presence of right-censoring. In our model, transitions between states are allowed to follow a dynamically varying nonlinear relationship with the individuals' covariates. Furthermore, the functional form of this non-linear relationship is semi-parametrically estimated at every time point using two single index models and thereby offering great flexibility in practice.

The literature of multistate models has been fairly dominated by parametric approaches over a long period. See Anderson and Keing (2002) for examples. However several works based on fully nonparametric concepts have added a great momentum to this area. In the past, Aalen (1976, 1978) and Aalen Johanson (1978) introduced nonparametric estimators of state occupational probabilities of a multistate model based on Nelson-Aalen type transition hazards. Datta and Satten (2001, 2002) showed the validity of these estimators under non-Markovian setting and extended their work further for subject dependent censoring. Introducing a new avenue to the parametric approach in multistate models, Anderson and Klein (2007) introduced a pseudo-values based regression approach starting with a marginal estimator which could be both parametric or nonparametric. Mostajabai and Datta (2013) developed a fully nonparametric approach to estimate conditional 'state-to-state transitions counts' and 'number at-risk' processes of a progressive multistate model under right-censoring given a value of a covariate. They incorporated the inverse probability censoring weighting (IPCW) concept described in Datta and Satten (2002), Satten and Datta (2002) to adjust the selection bias caused by the

censoring mechanism. However, limiting the practicability, their method handles only a single contentious covariate. Recently Chakrabory, Datta, and Datta (2015) extended this approach to multiple covariate cases, using the generalized additive model (Hastie and Tibshirani, 1990). However, this approach considers binary outcomes as contentious values between [0,1] interval for the model estimation, which may cause instability in the estimation. Furthermore, the robustness of their method under departure from the pre-assumed additive structure is uncertain. We propose a novel method to estimate underlining temporal processes of a multistate model, conditionally on given a covariate vector, introducing IPCW re-weighted binary choice Single Index Model (SIM). This approach allows one to estimate the conditional transition matrix for a given covariate vector at a specific time point, even when the transition mechanism has a highly nonlinear and rapidly varying dependency with multivariate baseline covariates. This is followed by a product limit calculation as in Datta and Satten (2002) to produce estimated conditional state occupying probabilities given the baseline covariates. Series of simulation studies show that the proposed method has desirable finite sample properties and it is robust under departure SIM from. We show that our method is fairly competitive for both estimation of the regression function and testing the effect of a baseline covariate, for future state occupation at a given time, by numerical comparisons with existing methods suitable for each of these purposes. We demonstrate the applicability of the proposed methodology in real life using two data sets resulting bone marrow transplantation and spinal cord injury studies.

CHAPTER 2
PERSONALIZED PLANS WITH MULTIPLE TREATMENTS

2.1 Treatment Selection

In this section we describe the proposed procedure and list some of its desirable large sample properties. Let $(Y_i^*, \boldsymbol{X})$ be the hypothetical (counterfactual) response and covariate pair for treatment i, $i = 1, \ldots, K$ where larger values of the response are indicative of better outcomes and $\boldsymbol{X}$ is a vector of r covariates. Assume further that a patient's covariate value $\boldsymbol{X}$ is used to obtain a lower dimensional composite patient score $U(\boldsymbol{X})$. In practice one cannot observe the whole vector $(Y_1^*, ..., Y_K^*)'$ for a single patient. However, using iid observations of type $(\tilde{Y}_i, \boldsymbol{X_i}, A_i)$, $i = 1, ..., n$ where A_i is the binary treatment indicator for two treatments and $\tilde{Y}_i$ is the observed response for the ith 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 $E[Y_1^*|A = 0, U(\boldsymbol{X})] - E[Y_2^*|A = 1, U(\boldsymbol{X})]$ where $U(\boldsymbol{X}) = \boldsymbol{X}$ and $A = 0$, 1 assign treatments 1 and 2 respectively.

In our approach, we consider pairs of independent observations $(Y_k, \boldsymbol{X}_k)$ from the marginal distribution of $(Y_k^*, \boldsymbol{X}), k = 1, ..., K$ to extend the treatment selection for K treatments using a set of probabilities defined as

$$p_i(u) = P[Y_i > \max_{i \neq j} Y_j | U(\boldsymbol{X_k}) = u; k = 1, ..., K]; i = 1, ..., K \qquad (2.1)$$

for a suitable score defined via a score function U. Note that in (2.1), 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 $\boldsymbol{X}$). Although the function $p_i(u)$ does not use the joint distribution of $(Y_1^*, ..., Y_K^*, \boldsymbol{X})$ for a patient with covariate value $\boldsymbol{X}$, we argue that p_i above nevertheless gives a measure of dominance for the ith treatment over the others and hence can be used in selecting the best treatment. This is an alternative to measures based on conditional expectations which require restrictive moment assumptions on the error distribution for all inference aspects in a regression context, the natural framework of handling such data. On the other hand, estimation of quantities like p_is can be done using conditional U-statistics with minimal assumptions. In our approach, for a given set of functions $p_1(.), ..., p_K(.)$, we define the best treatment for patients with a score U_0 as the treatment given by

$$k^*(U_0) = \arg \max_{1 \leq i \leq K} \{p_i(U_0)\}. \tag{2.2}$$

This procedure can be thought of as maximizing a value function that is the joint conditional expectation of an indicator of one treatment dominating the others given the score rather than evaluating $E[Y|U]$ for each treatment and picking the largest. For example, in Zhao et al. (2012), the best treatment was in principle defined as the index corresponding to the larger of $E[Y_1|U]$ and $E[Y_2|U]$ where Y_1 and Y_2 are the responses for each treatment. In practice, we propose to use estimators of $p_i(U_0)$ based on clinical data and then choose the best treatment as the one that is given by the corresponding estimator of $k^*(U_0)$.

The above approach can be meaningfully used for any set of models that is appropriate for relating responses and covariates provided that those models define an ordering of the above p_is for at least one score so that one of the treatments stands out. If several treatments have the same largest p_i value for a given score, one may pick one of those at random. As shown below, one set of models that can

provide such an ordering are Single Index Models (SIMs). In the sequel we base our discussions on Single Index Models relating response Y_i for the ith treatment and covariates $\boldsymbol{X}_i$ via

$$Y_i = g_i\left(\boldsymbol{\beta}_i'\boldsymbol{X}_i\right) + \epsilon_i \tag{2.3}$$

for $i = 1, \ldots, K$ where each $\boldsymbol{\beta}_i$ is a r-vector of parameters, g_is are unknown link functions for which we assume some reasonable smoothness conditions to hold, and ϵ is an error term with $E[\epsilon|\boldsymbol{X}] = 0$. This model can also be taken as a quantile regression model with suitable modifications.

In methods based on conditional means, one would ideally use $E[Y_i|\boldsymbol{X}]$ to select the best treatment. However, when $\boldsymbol{X}$ has very high dimension, a natural choice is to use a composite score $U(\boldsymbol{X})$ that has a much smaller dimension. We show in the sequel that if $g_i\left(\boldsymbol{\beta}_i'\boldsymbol{X}\right) > g_j\left(\boldsymbol{\beta}_j'\boldsymbol{X}\right)$ for all $i \neq j$, then the corresponding $p_i(u) > p_j(u), i \neq j$ for the realization $U = u$ for our proposed score. Hence, using p_is to choose the best treatment is somewhat more general than using conditional expectations. Although the properties of the proposed approach discussed in the sequel are for mean SIMs, they all also hold for quantile SIMs models. Additionally, those properties extend to PLSIMs as the parameters of the linear part of PLSIMs can be estimated at a $\sqrt{n}$ rate (see for example Liang et al., 2010).

If the model relating Y_i to $\boldsymbol{X}_i$ is not a SIM, we can still implement the same mechanism of obtaining the scores via a single index model approximation to the mean or the median of the responses and then estimate the corresponding p_is. This can be thought of as using a first order Projection Pursuit Regression to model the responses. Since nonparametric estimation of p_is require minimal model assumptions, our approach is applicable for a very wide class of models. For notational simplicity, we only list properties of the procedure for conditions that are appropriate for mean SIMs . Modifications in these conditions needed for other models are minimal.

Our data are of the following form. Let Y_{ij} indicate the jth responses from a group of n_i individuals under treatment i with covariate values $\boldsymbol{X}_{ij}, j = 1, \ldots, n_i$. The 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 2.3 is written as

$$Y_{ij} = g_i(\boldsymbol{\beta}_i'\boldsymbol{X}_{ij}) + \epsilon_{ij}, j = 1, \ldots, n_i. \tag{2.4}$$

Our approach to define an appropriate overall score 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. To be more specific, we first define

$$S_i(\boldsymbol{X}) = g_i(\boldsymbol{\beta}_i'\boldsymbol{X}) - \max_{j \neq i} \left\{ g_j\left(\boldsymbol{\beta}_j'\boldsymbol{X}\right) \right\}.$$

Next, we define the overall score to be the combination of the maximum of these treatment specific scores, and an index that indicates for which treatment the maximum has been achieved for the particular covariate value. That is, we define

$$\begin{aligned} S(\boldsymbol{X}) &= \max_i \{S_i\} \\ \delta(\boldsymbol{X}) &= \arg\max_i \{S_i\}. \end{aligned} \tag{2.5}$$

Then, for a patient with covariate value $\boldsymbol{X}$ we define the patient score as $U(\boldsymbol{X}) = (S(\boldsymbol{X}), \delta(\boldsymbol{X}))'$. Note that the score $U(\boldsymbol{X})$ reduces to the score used in the two treatment case by Cai et al. (2011) if we restrict gs to be linear. Also, if $K = 2$ and errors are symmetric about 0, δ becomes the index for the treatment with the larger location parameter for a given $\boldsymbol{X}$. However, when $K > 2$, this is not necessarily the case.

In practice one does not know the error distributions and model functions for models defined in 2.3 and therefore we cannot directly calculate p_is at a given score u. Thus, to apply the proposed selection method, we first need to estimate

each p_i using a standard function estimation method. This requires observed Y_{ij} values as well as observed U values corresponding to those responses. However, Us defined above are hypothetical scores for a covariate value $\boldsymbol{X}$ as we do not know g_is and $\boldsymbol{\beta}_i$s. Hence, in estimating p_is, we propose to use "estimated" $U(\boldsymbol{X}_{ij})$ values , $\hat{U}(\boldsymbol{X}_{ij})$, say, corresponding to responses $Y_{ij}, j = 1, ..., n_i; i = 1, ..., K$.

Now, to obtain $\hat{U}(\boldsymbol{X}_{ij})$ values, suitable estimators of link functions g_is and index vectors $\boldsymbol{\beta}_i$s can be used to construct estimators $\hat{S}(\boldsymbol{X})$ and $\hat{\delta}(\boldsymbol{X})$ of $S(\boldsymbol{X})$ and $\delta(\boldsymbol{X})$, respectively. There is a vast literature on estimating the link function and the index vector of a single index model (see, for example, 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 $\hat{g}_i$ and $\hat{\boldsymbol{\beta}}_i$, respectively, for $i = 1, \ldots, K$. In particular, for any given vector $\boldsymbol{x}$, let

$$\begin{aligned}
\hat{S}_i(\boldsymbol{x}) &= \hat{g}_i\left(\hat{\boldsymbol{\beta}}_i'\boldsymbol{x}\right) - \max_{j\neq i}\left\{\hat{g}_j\left(\hat{\boldsymbol{\beta}}_j'\boldsymbol{x}\right)\right\} \\
\hat{S}(\boldsymbol{x}) &= \max_i\left\{\hat{S}_i(\boldsymbol{x})\right\} \\
\hat{\delta}(\boldsymbol{x}) &= \arg\max_i\left\{S_i(\boldsymbol{x})\right\} \\
&\quad \text{and} \\
\hat{U}(\boldsymbol{x}) &= (\hat{S}(\boldsymbol{x}), \hat{\delta}(\boldsymbol{x}))' \qquad (2.6)
\end{aligned}$$

We randomly select an index $\hat{\delta}$ in the unlikely event that multiple treatments produce the same $\hat{S}$. Now, we construct our estimator for $p_i(u), i = 1, ..., K$ at a given $u = (s, d)'$ as follows. Define

$$\mathcal{J} = \{(j_1, \ldots, j_K) \,|\, j_i \in \{1, \ldots, n_i\}, i = 1, \ldots, K\}.$$

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

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

where w is a kernel function with $w \geq 0$ and $\int w(t)dt = 1$, and h_is are a set of smoothing parameters. Also, let

$$\begin{aligned}\hat{\eta}_J(d) &= \prod_{k=1}^{K} I\left[\hat{g}_d\left(\hat{\boldsymbol{\beta}}_d' \boldsymbol{X}_{kj_k}\right) = \max_m \left\{\hat{g}_m\left(\hat{\boldsymbol{\beta}}_m' \boldsymbol{X}_{kj_k}\right)\right\}\right] \\ &= \prod_{k=1}^{K} I\left(\hat{\delta}(\boldsymbol{X}_{kj_k}) = d\right).\end{aligned}$$

Now, taking an approach similar to the construction of conditional U-statistics (Stute, 1991), an estimator of $p_i(u), i = 1, ..., K$ can be defined as

$$\hat{p}_i(u) = \frac{\sum_{J \in \mathcal{J}} I\left[Y_{ij_i} > \max_{k \neq i}\{Y_{kj_k}\}\right] \hat{w}_J(s)\, \hat{\eta}_J(d)}{\sum \hat{w}_J(s)\, \hat{\eta}_J(d)}. \tag{2.7}$$

For a realization $\boldsymbol{X}_0$ of the covariate $\boldsymbol{X}$, if we knew the corresponding realization of the score, $u_0 = (S(\boldsymbol{x}_0), \delta(\boldsymbol{x}_0))'$, we can estimate $p_i(u_0)$ by $\hat{p}_i(u_0)$. However, due to the aforementioned reasons, we can only find an estimate $\hat{u}_0$ of u_0 using 2.6 above. Thus, we use $\hat{p}_i(\hat{u}_0)$ as our estimate of $p_i(u_0)$ for $i = 1, ..., K$. Finally, the estimated best treatment for a patient with estimated score $\hat{u}_0$ is defined as

$$\hat{k^*} = \arg \max_{1 \leq i \leq K} \{\hat{p}_i(\hat{u}_0)\}. \tag{2.8}$$

Under reasonable conditions stated below, we can show that

$$\hat{p}_i(\hat{u}_0) \to p_i(u_0) \tag{2.9}$$

in probability for each i. Hence, if for some k^*, $p_{k^*}(u_0) > \max_{j \neq k^*}\{p_j(u_0)\}$, then the treatment selection procedure described above is consistent since the best treatment is defined as the treatment corresponding to the largest p_i and, given the property $\hat{p}_i(\hat{u}_0) \to p_i(u_0)$, our procedure selects the best treatment with probability tending to one. The ordering of the p_is depends on models that relate the responses and the covariates.

Bandwidth selection for estimating the link functions and p_is is a challenging issue. Method suggested in Wand and Jones (1995) seemed to perform reasonably well in our simulations and data analysis. However, these choices may not be optimal. We do not investigate the optimal bandwidth selection issue in this work.

2.1.1 Theoretical Properties

In this section we list a few results that show the consistency of the proposed procedure. We begin by introducing some conditions that are needed to develop these theoretical results. In the sequel we assume that the random variables $(Y_i, \boldsymbol{X}_i), i = 1, ..., K$ are independent and further assume that $\boldsymbol{X_i}, i = 1, ..., K$ are iid. Let $F(s, d)$ be the common joint distribution function of $(S(\boldsymbol{X}_1), \delta(\boldsymbol{X}_1))$. We define $T_i = g_i(\boldsymbol{\beta}_i' \boldsymbol{X}_1)$, $\mathbf{T} = (T_1, \ldots, T_K)'$ and we let $f_{\mathbf{T}}(\mathbf{t})$ be the joint pdf of $\mathbf{T}$. We need following additional assumptions.

Assumption 1. *$F(s, d)$ is absolutely continuous in s for fixed d and has a density function $f(s, d)$, which is bounded.*

Assumption 2. *The kernel function w is symmetric, has bounded support, Riemann-integrable, nonnegative, bounded away from zero at 0, and has bounded derivative and finite total variation.*

Assumption 3. *$f_{\mathbf{T}}$ is continuous.*

Assumption 4. *The errors ϵ_{ij}, $j = 1, ..., n_i; i = 1, ..., K$ are i.i.d with a continuous pdf $f_\epsilon(\varepsilon)$ and $f_\epsilon(0) > 0$.*

Remark 1. *All distributional assumptions above are very reasonable and easily satisfied for many error distributions. Assumptions regarding the kernel function w are standard in nonparametric smoothing literature.*

The following lemma shows that the orderings of the p_is exist under models specified in 2.3. The proofs of some of these results use techniques similar to those used in the proofs of generalized U-statistics theory (Stute, 1991). However,since the generalized U-statistics theory is not directly applicable here, we give outlines of the proofs in the Appendix.

Lemma 1. *Under Assumption 1-4 and models 2.3, for a realization* $u = (s, d)'$ *of the score* $U(\boldsymbol{X})$ *defined above, functions* $p_i(u), i = 1, ..., K$ *are continuous in* s *and* $p_d(u) > \max_{1 \leq k \leq K; k \neq d} p_k(u)$.

The above lemma shows that under the SIM structure, if there is a link function function dominating others at a given covariate value, then there is a corresponding p function that dominates the other p functions over a non trivial set of scores. We now illustrate the consistency of $\hat{p}_i(\hat{u}_0)$ as an estimator for $p_i(u_0)$ at a given score u_0.

For our next result which shows that the estimator $\hat{p}(\hat{u}_0)$ converges to $p_i(u_0)$ we need the following assumption. In light of Remark 2 in the Appendix where the proof of Lemma 2 is provided and uniform convergence properties of nonparametric estimators of the link function in Single Index Models (see Wang and Yang (2007) and references therein), we see that this is a reasonable assumption.

Assumption 5. *For each* $i = 1, \ldots, K$, *smoothing parameters* $h_i \propto N^{-1/5}$ *and* $\sup_{x \in S_X} \left| \hat{g}_i \left(\hat{\boldsymbol{\beta}}_i' \boldsymbol{x} \right) - g(\boldsymbol{\beta}_i \boldsymbol{x}) \right| = O_p\left(N^{-2/5} \log N \right)$.

Now we have the following.

Theorem 1. *Under Assumptions 1-5, for* u_0 *and* $\hat{u}_0$ *defined above, we have* $\hat{p}_i(\hat{u}_0) - p_i(u_0) = o_p(1)$ *for* $i = 1, ..., K$.

This result shows that the selection of the appropriate treatment is consistent where we define consistency as being able to identify the index associated with the

largest p function in 2.1. In the next section we will provide an empirical assessment of the proposed procedure.

2.2 Empirical Studies

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

We conducted a series of simulations with the proposed procedure under various settings. Primarily, we focused on the accuracy of treatment assignment of a new (test) observation by using estimated values of the p_i functions from a set of training data. This simulation study was performed for both the two and multiple ($K > 2$) treatment groups cases. Results for the two groups cases were compared with the corresponding results for existing methods. However, such comparisons were not possible with multiple treatments since there is currently no other method covering more than two treatments. We select 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 samples with sample size n ($n = 50$ or $n = 100$) per group. The components of the r dimensional covariate vectors $\boldsymbol{X}$ were generated independently from a $U(-1, 1)$ distribution, where r ranged from 3 to 8. Using various link functions and index vectors, where a selected few are listed in Tables 2.1-2.3, we obtained the treatment responses from model 2.3. Here the errors were generated from $N(0, \sigma^2)$ and $DE(0, \sigma)$ where the dispersion parameter σ was chosen from the set $\{0.1, 0.2, 0.3, 0.4, 0.5, 1.0\}$. We have considered

the performance under both linear and nonlinear regression models. We discuss additional details of the structures of these models in the sequel.

Once the 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 Hristache et al. (2001) using Epanechnicov kernels (see Polzehl, 2013). Then, a new covariate value $\boldsymbol{X_0}$ was generated in the same manner as previous covariates above, and for its corresponding estimated score $\hat{u}_0$, we calculated $\hat{p}_i(\hat{u}_0)$ for $i = 1, ..., K$. The kernel function in this estimation was taken to be a $U(-1, 1)$ probability density function (pdf). We chose the bandwidths by the algorithm given by Wand and Jones (1995) for each $i, i = 1, ..., K$. We then generated K new responses, Y_i^*, each with mean $g_i(\boldsymbol{\beta}_i'\boldsymbol{X_0})$ for $i = 1, ..., K$, corresponding to this $\boldsymbol{X_0}$ using model 2.3 where the errors were generated independently from the same error distribution that was used to generate the K original samples. We define the treatment assignment to be correct if

$$\arg\max_i\{\hat{p}_i(\hat{u}_0)\} = \arg\max_i\{Y_i^*\}.$$

We repeated this procedure 1000 times for each model and error distribution combination. The frequency of correct treatment assignment for a selected set of cases are given in the Tables 2.1-2.3 and a few additional tables are provided in the supplemental materials.

In the analysis of the two groups case (Table 2.1), we used $N(0, \sigma^2)$ errors with $\sigma = 0.1$ and 0.2. We also compared these results with corresponding results for the two groups assignment methods proposed by Cai et al. (2011), Zhang et al. (2012), and Zhao et al. (2012). We report the number of cases in which their selection (using the highest conditional mean) matched with the group with the largest response. We chose to compare only with these three methods because these methods highly differ in their approaches and dominate other existing methods in

the literature for the two groups case. Here, we highlighted the settings in which other methods underperformed against our method by an asterisk sign. Out of 48 cases the new method competed well with the existing methods in 40 cases. Clearly, the proposed method has a high accuracy in nonlinear models compared to the three existing treatment selection methods. In the case of linear treatment models, which is represented by Model 1 in Table 2.1, the new method performed comparably to the best method. Model 4 in Table 2.1 was chosen to demonstrate the robustness of the proposed method, where the requirement of SIM's is violated. Even in these cases, the accuracy remained fairly high, showing that the proposed method is rather robust.

We studied the multiple treatment groups case for $K = 3$ and 4, using a variety of models generated from several nonlinear model families. All considered cases produced results that are generally anticipated in a study of this nature. Cases involving highly nonlinear curves with minor differences in the mean value function performed somewhat poor compared with cases where the nonlinearity is less severe or the differences between the signals is higher. Our discussion in the sequel focuses on two families

$$Y_i = g_i\Big\{\pi k_i + \pi\big(\boldsymbol{\beta}'\boldsymbol{X}\big)\Big\} + \varepsilon,\ i = 1, ..., K, \quad \text{(Type I)},$$

and

$$Y_i = g_i\Big\{\pi k_i + \pi\big(\boldsymbol{\beta}_i'\boldsymbol{X}\big)\Big\} + \varepsilon,\ i = 1, ..., K, \qquad \text{(Type II)}.$$

In each type above, g_i is either a *sine* or a *cosine* function. In Type I models, the same single index vector $\boldsymbol{\beta}$ has been used for the treatment groups where the g_i function varies across the groups. In our simulations we chose this common vector to be $\boldsymbol{C'} = \big(1/\sqrt{r}, ..., 1/\sqrt{r}\big)_{1\times r}$ Table 2.2. In Type II models we used a variety of $\boldsymbol{\beta}_i$ index vectors whose components were selected in an arbitrary fashion. These components are given in Table 2.4. For example, in the three treatment case with

$r = 3$, $\boldsymbol{\beta_1} = \big(1.5, 1.6, 0.9\big)'$, $\boldsymbol{\beta_2} = \big(0.8, 0.6, 0.7\big)'$, and $\boldsymbol{\beta_3} = \big(1.8, 2.1, 0.8\big)'$.

If several models are close to each other within the whole covariate domain, a high classification error (i.e., incorrect treatment assignment) can be expected due to the lack of functional separations. In general, the functional behavior of a multi covariate nonlinear model cannot be easily visualized. Type-I models used here have relatively substantial functional differences compared to some Type II models for each K. Tables 2.2 and 2.3 show the correct assignment frequencies for a representative set of multi-groups cases. Again, the results for all examined cases were very similar to the few presented here.

Examination of the results reveal high assignment accuracy for large sample sizes and low error variability. In general, we observed fairly high accuracies for low covariate dimensions. The presented simulation results are based on *sine* and *cosine* functions which are bounded in (-1, 1). Hence, an increment in σ by 0.1 adds a relatively large noise to a model. Consequently, as expected, we observed a decline in the correct assignment frequency as σ is increased. The results for the three groups case for both Type I and II models are somewhat comparable whereas the results for Type II models for four groups case were lower compared to those corresponding to Type I models. As indicated in the previous paragraph, we believe these differences are due to relative lack of separation in the model functions.

2.3 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. The severity of HIV progression is measured through a decline in CD4 counts. This trial periodically measured a patient's CD4 count as the clinical outcome. In our analysis, we considered the log transformed CD4 count of a patient after 20 weeks of treatment as the clinical response. As covariates, we used log-CD4 and log-CD8 counts at baseline, age, weight, and the number of months a patient received pre-antiviral therapy.

We applied the proposed treatment assignment strategy to the data from all four arms of the study. We also provide an illustration to compare with several existing two-treatment methods. In each situation, we randomly selected 200 patients from each arm as "training" data to estimate the SIMs. Remaining patients were considered as new (test) patients. After fitting SIMs to training data we estimated the scores for the test cases and estimated the corresponding p_i functions using Gaussian kernels at corresponding scores to assign each test patient to the best treatment suggested by the largest estimated p_i value.

We report the results for the two group comparisons first. When we used the proposed method, out of 651 test patients, only 3 were assigned to arm-0, suggesting that possibly a large number of patients would have experienced a more favorable outcome from arm-1. We also applied the two-group assignment methods proposed by Cai et al. (2011), Zhao et al. (2012), and Zhang et al. (2012), for the same training and test data. These methods also assigned lesser number of patients to arm 0, than the actual assignment by the randomized trial. We present these results

in Table 2.5. For example, in Table 2.5, the (1, 0) cell for the Proposed Method indicates that only 2 out of 319 patients who were actually treated in arm-1 would have been assigned to arm-0 had we used the proposed method.

In the multiple treatments assignment setting, we have a total of 1336 patients in the test set. Among them, we assigned the majority: 828 to arm-1 whereas 306 and 186 patients are assigned to arms 2 and 3, respectively. Similar to the two group assignment, the new method assigns only few patients to arm-0, seemingly suggesting that one of the other arms almost always dominate arm 0 with respect to our scoring mechanism. These results are summarized in Table 2.6. We noticed that, a large number of patients (1023) are proposed to be assigned to a different treatment arm than their actual assignment. Based on these allocations, it appears that the majority of patients in the study would have benefited from arm 1.

2.3.1 Examination of the survival aspect

The proposed treatment selection method above is an attempt to assign patients to receive the optimal outcome based on their score. Given that the above analysis shows that the optimal assignments based on patient characteristics are different from actual assignments towards a higher CD4 count, it might be the case that such an assignment rule could also improve the expected value of the related survival time conditional on the score. To explore whether such an implication might hold, we proceed as follows.

In the dataset, there are three types of events: (i) when an individual's CD4 count drops less than 50% of his/her pretreatment count, (ii) an event indicating progression to AIDS, (iii) death. Thus, the term "survival time" would denote an event time in the above sense. In addition, there was right censoring present in the data. Now, consider the i th subject in the test set with covariate value $\boldsymbol{X}_i$ who is

assigned to a particular arm by an assignment mechanism. Suppose the individuals estimated score is $\hat{u}_i = \left\{\hat{S}(\boldsymbol{X}_i), \hat{\delta}(\boldsymbol{X}_i)\right\}$. Let k_i^* be the group the procedure would assign this patient based on his/her estimated score $\hat{u}_i$ and let k_i be the treatment group he was assigned in the original trial. Conditional on $\hat{u}$, we estimated the difference in the survival times in the two groups, as

$$\Delta_i = E(t_{k_i^*}|\hat{u}_i) - E(t_{k_i}|\hat{u}_i).$$

For a fixed k, we consider a symmetric neighborhood of width $2h$ centered around $\hat{S}(\boldsymbol{X_i})$,

$$N_h = \left\{\hat{S}(\boldsymbol{X}_i) - h, \hat{S}(\boldsymbol{X}_i) + h\right\},$$

where h was the bandwidth chosen by the procedure given in Wand and Jones (1995) for scores for all patients. Next, we selected a subgroup of patients from the whole set (training and test), whose covariate values $\boldsymbol{X}$ satisfy (i) patient was originally treated in arm k and (ii) $\hat{S}(\boldsymbol{X}) \in N_h$ and (iii) the score satisfies $\hat{\delta}(\boldsymbol{X}) = \hat{\delta}(\boldsymbol{X_i})$. If the size (d, say) of the above subgroup is less than 30, we increased the width of the neighborhood N_h in multiples of h (i.e., $3h$, $4h$ etc.) to make $d \geq 30$. After that the Kaplan Meier estimator was calculated using the survival times of those individuals in N_h.

Our estimator of the expected survival time for each group, i.e., $E(t_{k_i}|\hat{u}_i)$, $k_i = 1, ..., K$, was the area covered under the corresponding Kaplan-Meier curve. For a given $\hat{u}_i$, we then find the estimated survival gain $\hat{\Delta}_i$ from the proposed selection as the difference between the two estimated expectations, $\hat{E}(t_{k_i^*}|\hat{u}_i) - \hat{E}(t_{k_i}|\hat{u}_i)$. Finally we estimate the overall treatment selection efficiency as the averaged $\hat{\Delta}_i$s for all test patients,

$$\rho = \frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i, \tag{2.10}$$

where N is the number in the test set. Note that a positive value for ρ indicates an overall effective treatment selection. Table 2.7 gives these ρ values for the proposed

procedure with two and multiple treatments cases along with the resulting estimated survival gains for methods proposed by Cai et al. (2011), Zhao et al. (2012), and Zhang et al. (2012), for the two-groups application. Additionally, we consider the marginal survival functions and define,

$$\Delta_i^{'} = E(t_{k_i^*}) - E(t_{k_i}),$$

where $E(t_{k_i^*})$ and $E(t_{k_i})$ are corresponding marginal expected survival times of new (k_i^*) and actual (k_i) arms. Again using the area under the marginal Kaplan Meier estimates, we calculate estimated values of $\Delta_i^{'}$, $i = 1, ..., N$. Similar to 2.10, we obtain ρ_m using these marginal estimates. Corresponding ρ_m's are also reported in Table 2.7. Since the proposed treatment selection is based on a scoring scheme, we argue that examining the score dependent survival outcome would be a more reliable approach. This is confirmed by the fact $\rho_m \leq \rho$ in all cases.

2.4 Discussion

In this Chapter we proposed a novel personalized treatment plan to select the optimal treatment from a set of multiple treatments. This method is a single step procedure which can be easily applied. The proposed method is based on semi parametric Single Index Models which, add great flexibility in modeling real life situations. Furthermore, this method can also be used for quantile regression SIMs providing additional model flexibility compared with existing methods based on conditional expectations. Our empirical studies show that the proposed method performs very satisfactorily in selecting the optimal treatment in a multiple treatment setting while outperforming existing methods for the two treatment case for non-linear models which are more realistic in practical situations. In addition, as our simulations showed, the method is rather robust against departures from SIMs. We show that the proposed method has desirable theoretical properties. Our anal-

ysis of a real clinical trials dataset which has the multiple treatment option reveals a possible connection between optimal treatment selection and a gain in patient survival.

This discussion 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. Our study is addressing the optimal treatment selection based on a single response. However, there are numerous circumstances where the optimality is desired with respect to multiple criteria. For example, a treatment may have to be selected to maximize the survival rates but minimize after effects and maximize the quality of life in terms of temporary side effects. In such cases we have a multi criteria optimization problem. This opens up another interesting future research avenue.

2.5 Tables

Models(regression function)	Error SD (σ)	Per Group Size (n)	Proposed Method	Cai's Method	Zhao's Method	Zhang's Method
(1) $(1.5X_1 - 0.1X_2 + 2X_3 + 2X_4 - 1.5X_5 - 1.6X_6)/\sqrt{15.07}$: Group 1 $(2X_1 + 1.6X_2 + 2.2X_3 + 3.5X_4 + 1.2X_5 + 1.5X_6)/\sqrt{27.34}$: Group 2	0.1	50	864	889	800*	899
		100	891	906	856*	902
	0.2	50	794	817	773*	801
		100	825	840	821*	844
(2) $\sin\{\pi(X_1 - 0.3X_2 - X_3)/\sqrt{2.09}\}$: Group 1 $\sin\{\pi/4 + \pi(X_1 + X_2 + X_3)/\sqrt{3}\}$: Group 2	0.1	50	900	683*	676*	683*
		100	891	722*	698*	728*
	0.2	50	860	670*	691*	693*
		100	853	708*	683*	718*
(3) $\sin\{\pi(0.8X_1 + 1.1X_2 + 0.9X_3 + X_4 + 0.9X_5 + 1.1X_6)/\sqrt{5.68}\}$: Group 1 $\sin\{\pi/2 + \pi(1.8X_1 - 1.3X_2 + 0.8X_3 + X_4 - 1.2X_5 - X_6)/\sqrt{9.01}\}$: Group 2	0.1	50	880	605*	633*	678*
		100	911	680*	671*	700*
	0.2	50	839	606*	603*	672*
		100	868	672*	652*	664*
(4) $\sin\{\pi(X_1 + X_2 + X_3)/\sqrt{3}\} + X_1^2$: Group 1 $\sin\{\pi/2 + \pi(X_1 + X_2 + X_3)/\sqrt{3}\} + 0.7X_1^2$: Group 2	0.1	50	920	743*	694*	804*
		100	935	794*	764*	842*
	0.2	50	908	741*	703*	796*
		100	924	774*	740*	834*

Table 2.1: Frequencies of correct treatment assignments in 1000 test cases by four competing algorithms in the two groups case. The regression models used in the simulations include linear and nonlinear SIM models, as well as models that are not SIM. Cases where the proposed method (ours) outperformed a competing method is denoted by *.

Number of Groups	Models (regression function)	Dimension r	Per Group Size	Normal Error SD (σ)			DE Error SD (σ)		
				0.1	0.3	0.5	0.1	0.3	0.5
Three	$\sin\Big\{\pi(\boldsymbol{C}'\boldsymbol{X})\Big\}$: Group 1 $\cos\Big\{\frac{\pi}{6}+\pi(\boldsymbol{C}'\boldsymbol{X})\Big\}$: Group 2 $\sin\Big\{\frac{7\pi}{5}+\pi(\boldsymbol{C}'\boldsymbol{X})\Big\}$: Group 3	3	50	937	857	733	929	804	649
			100	946	866	766	936	846	720
		5	50	934	843	761	922	818	682
			100	973	916	796	949	843	728
		8	50	897	816	707	883	726	563
			100	962	875	784	950	839	739
Four	$\sin\Big\{\pi(\boldsymbol{C}'\boldsymbol{X})\Big\}$: Group 1 $\sin\Big\{\frac{\pi}{2}+\pi(\boldsymbol{C}'\boldsymbol{X})\Big\}$: Group 2 $\sin\Big\{\frac{-\pi}{2}+\pi(\boldsymbol{C}'\boldsymbol{X})\Big\}$: Group 3 $\sin\Big\{\pi+\pi(\boldsymbol{C}'\boldsymbol{X})\Big\}$: Group 4	3	50	904	799	653	890	699	529
			100	939	822	714	909	773	625
		5	50	895	775	640	878	699	518
			100	946	827	687	910	752	608
		8	50	853	720	549	836	644	438
			100	926	809	689	903	717	577

Table 2.2: Frequencies of correct treatment assignments in 1000 test cases by the proposed method in multiple groups case ($K > 2$), using Type I nonlinear regression models, with $\boldsymbol{C'} = \big(1/\sqrt{r}, ...1/\sqrt{r}\big)_{1\times r}$.

Number of Groups	Models (regression function)	Dimension r	Group Size	Normal Error SD (σ)			DE Error SD (σ)		
				0.1	0.3	0.5	0.1	0.3	0.5
Three	$\sin\left\{\frac{\pi}{\|\beta_1\|}(\boldsymbol{\beta}_1'\boldsymbol{X})\right\}$: Group 1 $\cos\left\{\frac{\pi}{6}+\frac{\pi}{\|\beta_2\|}(\boldsymbol{\beta}_2'\boldsymbol{X})\right\}$: Group 2 $\sin\left\{\frac{7\pi}{5}+\frac{\pi}{\|\beta_3\|}(\boldsymbol{\beta}_3'\boldsymbol{X})\right\}$: Group 3	3	50	956	878	766	924	818	688
			100	970	896	796	948	844	737
		5	50	930	865	762	926	802	694
			100	947	897	802	942	838	741
		8	50	888	801	649	862	736	537
			100	947	881	791	942	812	724
Four	$\sin\left\{\frac{\pi}{\|\beta_1\|}(\boldsymbol{\beta}_1'\boldsymbol{X})\right\}$: Group 1 $\cos\left\{\frac{\pi}{8}+\frac{\pi}{\|\beta_2\|}(\boldsymbol{\beta}_2'\boldsymbol{X})\right\}$: Group 2 $\cos\left\{\frac{-\pi}{6}+\frac{\pi}{\|\beta_3\|}(\boldsymbol{\beta}_3'\boldsymbol{X})\right\}$: Group 3 $\sin\left\{\pi+\frac{\pi}{\|\beta_4\|}(\boldsymbol{\beta}_4'\boldsymbol{X})\right\}$: Group 4	3	50	830	652	562	772	602	464
			100	884	738	616	819	673	513
		5	50	727	595	508	731	558	410
			100	822	697	557	791	645	508
		8	50	725	586	447	685	486	381
			100	823	657	549	802	599	490

Table 2.3: Frequencies of correct treatment assignments in 1000 test cases by the proposed method in multiple groups case ($K > 2$), using Type II nonlinear regression models. Selected $\boldsymbol{\beta}$ vectors are shown in Table 2.4.

Treatment groups	Group	Number of covariates	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Three	1	3	1.5	1.6	0.9					
		5	1.5	1.6	0.9	1.2	1.4			
		8	1.5	1.6	0.9	1.2	1.4	-1.5	1.2	1.6
	2	3	1.0	1.4	0.8					
		5	1.0	1.4	0.8	0.8	0.6			
		8	1.0	1.4	0.8	0.8	0.6	-1.1	0.8	0.6
	3	3	1.3	1.7	0.7					
		5	1.3	1.7	0.7	0.9	1.1			
		8	1.3	1.7	0.7	0.9	1.1	-1.3	-0.1	0.9
Four	1	3	0.8	0.6	0.7					
		5	0.8	0.6	0.7	0.5	0.6			
		8	0.8	0.6	0.7	0.5	0.6	0.8	0.7	0.5
	2	3	1.2	1.4	0.9					
		5	1.2	1.4	0.9	1.5	0.9			
		8	1.2	1.4	0.9	1.5	0.9	1.1	1.4	1.2
	3	3	0.2	0.3	0.8					
		5	0.2	0.3	0.8	0.6	0.3			
		8	0.2	0.3	0.8	0.6	0.3	0.1	0.4	0.6
	4	3	1.8	2.1	0.8					
		5	1.8	2.1	0.8	0.7	0.9			
		8	1.8	2.1	0.8	0.7	0.9	1.3	1	1.3

Table 2.4: $\boldsymbol{\beta}$ vectors of Type II models, for model dimensions (r) 3, 5, and 8.

Orginal Assignment	New Assignment							
	Proposed Method		Cai's Method		Zhao's Method		Zhang's Method	
	Arm-0	Arm-1	Arm-0	Arm-1	Arm-0	Arm-1	Arm-0	Arm-1
Arm-0	1	331	13	319	0	332	28	304
Arm-1	2	317	11	308	0	319	25	294
Total	3	648	24	627	0	651	53	598

Table 2.5: Two groups treatment assignment summary for ACTG-175 trial, by four methods.

Original Assignment	Proposed Assignment			
	Arm-0	Arm-1	Arm-2	Arm-3
Arm-0	2	211	70	49
Arm-1	2	193	77	47
Arm-2	5	201	73	45
Arm-3	5	223	88	45
Total	14	828	308	186

Table 2.6: Four groups treatment assignment summary for ACTG-175 clinical trial, by the proposed method.

	Two Groups Assignments				Four Groups Assignments by Proposed Method
	Proposed Method	Cai's Method	Zhao's Method	Zhang's Method	
ρ	76.1	73.0	77.2	66.0	56.3
ρ_m	62.5	58.5	63.1	53.0	32.2

Table 2.7: Obseved ρ and ρ_m by four treatment selection methods, under different treatment possibilities.

CHAPTER 3
TREATMENT SELECTION FOR SURVIVAL OUTCOME

3.1 Treatment Selection

The goal of this project is to extend project 1 to handle censored responses. Here, the response is a survival time or an event time that is subject to right censoring.

Let (Y_i, X_i) be the survival response and covariate pair for the i'th treatment group, the covariate vector is r dimensional. Let C_i denote the right censoring time and let $T_i = Y_i \wedge C_i$ and $\delta_i^{'} = I[Y_i \leq C_i]$. We denote the survival function by $S^{'}(t) = E\{I[Y_i > t]\}$, with hazard function $\lambda(t)$ and cumulative hazard function $\Lambda(t)$. Suppose patients' covariate is used to obtain a score $U(X_i)$. Assuming all Y_i's are observed, in our previous work, we define a function $p_i(u)$ that provides the probability of dominance for the i th treatment over others, for a patients' with a covariate value X_k and a score $U(X_k) = u,$ for $k = 1, ..., K$:

$$p_i(u) = P[Y_i > \max_{i \neq j} Y_j | U(\boldsymbol{X_k}) = u; k = 1, ..., K].$$

For a given set of functions, $p_1(.)$, ..., $p_k(.)$, for a patient with a score of U_0, the best treatment is given by,

$$k^*(U_0) = arg\ max_{1 \leq i \leq K}\{p_i(U_0)\}.$$

As we described before, $p_i(u)$ is an alternative to measures based on conditional expectations which requires restrictive moment assumptions. Here, scores are ob-

tained via a set of single index models or partial linear models define for each group given by (2.3). We define a patient score as a composite function given by (2.5),

$$U(X) = (S(X), \delta(X)).$$

Since the score is estimated via a set of estimated single index models, for a given patient we have an estimated score of $\hat{U}(X) = (\hat{S}(X), \hat{\delta}(X))'$. For completely observed responses, the probability of ith treatment dominating the others for a given score value (u) is given by (2.7). However, once patients' response's are subject to right censoring, some Y_{ij} $'$s may be unobservable, which makes the application our of previous treatment selection method directly to available data impossible. In this study we introduce a modified method developed in parallel to the previous personalized treatment selection concept. This involves modification of the single index estimator and the estimator of $p_i(u)$. Use of data weighting schemes for the purpose of bias reduction are well known in statistical literature. One way to handle censored observations in the context of regression is to introduce a re-weighing scheme to the original estimator developed for complete data, in a way that the bias caused due to censoring fades away asymptotically. This idea was first introduced by Koul (1981), for the randomly right censored data in linear regression. Datta et al (2001) described estimating the marginal survival time in the presence of time dependent covariates, using a re-weighted Kaplan-Meier estimate, applying the IPCW calculated by Aalen's additive hazard model (1989). In a recent article, Lopez et al. (2013) described estimating the single index model incorporating IPCW. In their method, they used cox's proportional hazard model to obtain covariate dependent censoring probabilities. Similar to these concepts, in our study, we introduce a reweighing scheme to estimators of single index model and $p(u)$ function.

3.1.1 Estimation of the IPCW Weights

We applied Aalen's additive hazard model to estimate the weights in the case of covariate dependent censoring. Aalen's model is more flexible compared to Cox's proportional hazard model. However, it's important to note that estimating Aalen's model involves inverting a non full rank matrix. Although estimated hazards depend on the solution of the selected generalized inverse, it doesn't impact the proposed weighting scheme, since weights are uniquely defined. In this study, we employed Moore-Penrose generalized inverse. Also, in Aalen's model there is no strong criteria imposed to restrict the hazard to be positive. Aalen's additive hazard model can be written as,

$$\lambda_c[t|Z_i(t)] = \sum_{j=0}^{J} \beta_j(t) Z_{ij}(t).$$

where, $\beta_j(t)$ is an unknown function, that needs to be estimated. $Z_{ij}(t)$ is a predicable process and $Z_{ij}(t)$ is the corresponding value avialabe just before time t. Here, $Z_{i0}(t) \equiv 1$. Define $B_j(t)$ as,

$$B_j(t) = \int_0^t \beta_j(s) ds.$$

Aalen's model estimates, $B(t) = (B_0(t), ..., B_J(t))$ by,

$$\hat{B}(t) = \sum_{i=1}^{n} I(t_i \leq t)(1 - \delta_i) A^{-1}(t_i) Z_i(t_i),$$

where, $Z_i(t) = (Z_{i0}(t), ..., Z_{iJ}(t))$, and $A(t) = \sum_{i=1}^{n} I(t_i \geq t) Z_i(t_i) Z_i^T(t_i)$. Cumulative hazard at time t for covariate $Z_i(t)$, is given by,

$$\hat{\Lambda}_c[t|Z_i(t)] = \sum_{j=0}^{J} \int_0^t Z_{iJ}(t)\, d\hat{B}_j(s).$$

In the case of random censoring, we obtain inverse censoring probabilities by Kaplan-Meier estimator.

3.1.2 Re-weighted Single Index Estimator

Suppose variable Y is linked to a linear combination of set predictors (X) represented by $\boldsymbol{\beta' X}$ via an unknown link function g. SIM model is defined by,

$$Y = g(\beta' X) + \epsilon,$$

where,

$$E(\epsilon|X) = 0.$$

For the purpose of identifiability, we may replace β by a unit vector,

$$\theta = \beta||\beta||^{-1}.$$

where $||.||$ is the Euclidean norm. Thus, an equivalent model can be written as,

$$Y = g(\theta' X) + \epsilon.$$

Assuming all Y's are completely observed, Ichimura et al. (1993) proposed an estimator to estimate the above SIM model. Accordingly, the unknown function $g(.)$ is estimated at point u, by leave-one-out cross validation method, omitting the pair of (Y_i, X_i),

$$\hat{g}_{-i}(u|\theta) = \frac{\sum_{j \neq i} Y_j \omega_h(u - \theta' X_j)}{\sum_{j \neq i} \omega_h(u - \theta' X_j)}.$$

where, h is a smoothing parameter, $\omega_h(.) = \omega(./h)$, and ω, is a fixed kernel function with $\omega \geq 0$ and $\int \omega(t)dt = 1$. Ichimura et al. (1993) showed, estimates of θ and h can be achieved by simultaneously minimizing the following objective function with respect to θ and h.

$$\hat{S}(\theta,\, h) = \sum_i \{Y_i - \hat{g}_i(\theta' x|\theta)\}^2$$

Once observed data subject to right censored, above SIM estimator is no longer valid. We suggest an alternative SIM estimator, which is capable of handling right censored data in survival outcomes. This new estimator is primarily based on the

method proposed by Ichimura et al (1993), but it's re-weighted by a IPCW weighting scheme. We define leave-one-out re-weighed estimator of $g(.)$ as,

$$\hat{g}_{-i}(u|\theta) = \frac{\sum_{j\neq i} \frac{\delta_j' T_j}{K^c_{-i}(T_j-)} \omega_h(u - \theta' X_j)}{\sum_{j\neq i} \omega_h(u - \theta' X_j)}.$$

Here, $K^c(T-)$ is the survival probability of an individual not being censored just before time T, which can be estimated from either Aalen's linear model or Kaplan-Meier estimator, depending on the censoring mechanism. We estimate θ and h by minimizing the following weighted objective function denoted by $(\hat{S}'(\theta, h))$ simultaneously with respect to both θ and h.

$$\hat{S}'(\theta,\, h) = \sum_i \frac{\delta_i'}{K^c(T_i-)} \{T_i - \hat{g}_i(\theta' X_i|\theta)\}^2$$

For estimators $\hat{\theta}$ and $\hat{h}$, $g(.)$ function at a new point $u_0 = \theta' x_0$, can be estimated as,

$$\hat{g}_(u_0|\hat{\theta}) = \frac{\sum_i \frac{\delta_j' T_j}{K^c(T_j-)} \omega_h(u_0 - \hat{\theta}' X_j)}{\sum_i \omega_h(u_0 - \hat{\theta}' X_j)}. \tag{3.1}$$

In reality, we replace $K^c(.)$ by its corresponding estimator $\hat{K}^c(.)$. A simulation study to evaluate the properties of the re-weighted single index estimator showed reasonable performance under both random and covariate dependent censoring, seemingly suggesting that the alternative SIM estimator is a reasonable estimator of estimating patient scores with right censored responses. We explain detailed results later in the "Empirical Studies".

3.1.3 Re-weighted Estimator of Treatment Selection

Assuming all treatment responses (Y) are observed, in project-1 we provided a treatment selection rule that assigns a patient to the treatment that dominates all others for a given patient score. The corresponding probability that ith treatment dominating for a patent with a score of u is given by (2.7). In the same fashion

as the single index model was re-weighted, we introduce a new estimator of $p(u)$ capable of coping with right censored data by employing IPCW weighting scheme. Again, the new rule is structured on the same score criteria given by (2.6). The new estimator is defined as below. Let

$$\mathcal{J} = \{(j_1, \ldots, j_K) \,|j_i \in \{1, \ldots, n_i\}, i = 1, \ldots, K\}\,.$$

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

$$\hat{w}_J\,(s) = \prod_{i=1}^{K} \frac{1}{h_i} w\left(\frac{s - \hat{S}\,(\boldsymbol{X}_{ij_i})}{h_i}\right)$$

where w is a kernel function with $w \geq 0$ and $\int w(t)dt = 1$, and h_is are a set of smoothing parameters. Also, let

$$\begin{aligned}\hat{\eta}_J\,(d) &= \prod_{k=1}^{K} I\left[\hat{g}_d\left(\hat{\boldsymbol{\beta}}_d' \boldsymbol{X}_{kj_k}\right) = \max_m \left\{\hat{g}_m\left(\hat{\boldsymbol{\beta}}_m' \boldsymbol{X}_{kj_k}\right)\right\}\right] \\ &= \prod_{k=1}^{K} I\left(\hat{\delta}\,(\boldsymbol{X}_{kj_k}) = d\right).\end{aligned}$$

As indicated before, suppose, $K^c(T-)$ is the survival probability of an individual not being censored just before time T. We define $\hat{\kappa}_J$,

$$\hat{\kappa}_J = \prod_{k=1}^{K} \frac{\delta'_{kj_k}}{\hat{K}(T_{kj_k}-)}.$$

Estimator of $p(u)$, is given by,

$$\hat{p}_i\,(u) = \frac{\sum_{J\in\mathcal{J}} I\left[T_{ij_i} > \max_{k\neq i}\left\{T_{kj_k}\right\}\right] \hat{w}_J\,(s)\,\hat{\eta}_J\,(d)\,\hat{\kappa}_J}{\sum \hat{w}_J\,(s)\,\hat{\eta}_J\,(d)}. \tag{3.2}$$

For a given patient with a covariate vector x_0 and estimated score of $\hat{u}_0 = (\hat{S}(x_0), \hat{\delta}_0)$, the proposed treatment rule assign him/her into k^* th group if,

$$\hat{k^*} = \arg \max_{1\leq i\leq K} \{\hat{p}_i(\hat{u}_0)\}. \tag{3.3}$$

As mentioned in the treatment selection for complete data, the optimal bandwidths selection for this problem is challenging. Our empirical studies demonstrated reasonable performance using the bandwidth selection given by Wand and Jones (1995), even in the right censored case.

As an alternative approach to the above treatment selection plan that is based on $p_i(u)$'s, we also propose a method based on comparing conditional expected means using a re-weighted estimator for smooth means given $u = (s, d)$. The proposed estimator is given by,

$$\hat{\mu}_i(u) = \frac{\sum_{j=1}^{n_i} \frac{\Delta_{ij}}{\hat{K}(T_{ij}-)} T_{ij} w\left((s - \hat{S}(\boldsymbol{X}_{ij}))/h_i\right) I\left(\hat{\delta}(\boldsymbol{X}_{ij}) = d\right)}{\sum_{j=1}^{n_i} w\left((s - \hat{S}(\boldsymbol{X}_{ij}))/h_i\right) I\left(\hat{\delta}(\boldsymbol{X}_{ij}) = d\right)}. \tag{3.4}$$

The optimal treatment using $\hat{\mu}_i(u)$'s is defined for an individual with an estimated score $\hat{u}_0 = (\hat{S}(x_0), \hat{\delta}_0)$ as,

$$\hat{k}^* = \arg \max_{1 \le i \le K} \{\hat{\mu}_i(\hat{u}_0)\}. \tag{3.5}$$

In this work we used the Wand and Jones (1995) bandwidth selection concept for estimating $\mu_i(u)$'s for a comparative study with optimal treatment selection based on $p_i(u)$'s.

3.2 Empirical studies

In this section we provide detailed information on a simulation study performed to assess the performance of new treatment selection rule and single index model estimator for right-censored data.

3.2.1 Re-weighted Single index estimator

The following simulation study explores the performance of the IPCW re-weighted single index estimator in finite sample.

In this experiment, we generated survival time Y's using a three dimensional covariate vector X, in a highly non-linear model given by,

$$Y = a + sin\{\pi(\beta^{'}X)\} + \varepsilon, \quad \varepsilon \sim N(0,\, 0.1^2).$$

We chose, $\beta = (\frac{1}{\sqrt{3.5}}, \frac{1.5}{\sqrt{3.5}}, \frac{0.5}{\sqrt{3.5}})'$. and $a = 1.2$. Here each component of X was independently generated from $U(-1, 1)$. The study examines the performance of the SIM estimator under two censoring mechanisms; random and covariate dependent with combinations of various censoring percentages ranging approximately from 10% to 50%, and model training sizes (n) from 100 to 2000. Random censoring time was generated from single parameter scale Exponential distribution with a parameter ϕ selected from the set $\{0.1,\, 0.3,\, 0.4\}$, whereas in the covariate dependent censoring setting, we obtained censoring time using a function which is a mixture of two distinct Exponential distributions, defined for a threshold value of a linear combination of covariates given by,

$$C \sim I(\beta_c^{'}X > w)exp(\phi_1) + I(\beta_c^{'}X \leq w)exp(\phi_2).$$

We fixed $w = 0.4$ for $\beta_c = (0.2, 0.3, 0.4)$ and selected (ϕ_1, ϕ_2) from the set $\big\{(0.01, 0.10),$ $(0.15, 0.40),\ (0.30, 0.70)\big\}$. In each scenario, we first estimated the SIM model and then determined the average L_1 distance between predicted and true functions for a testing set of size 1000, that was generated similarly to the training set. Thus the average L_1 distance is given by,

$$\Delta = \frac{1}{1000}\sum_i |\hat{g}(\hat{\beta}^{'}X_i) - g(\beta' X_i)|.$$

Tables 3.1 and 3.2 show average bias, standard error of re-weighted SIM parameters and average Δ observed under each setting, using 1000 Mote-Carlo simulations.

Δ clearly decreased as n increased for both censoring mechanisms and rates, suggesting that the estimated function asymptotically converges to the true function. Also, we observed a reduction in bias and standard error for large n. As to be expected, comparably larger bias and standard error were reported under high censoring rates. Suppose,

$$\Delta \sim n^{-\gamma}.$$

So that,

$$\log(\Delta) \sim -\gamma \log(n).$$

In the simulated examples, we observed a linear trend in $\log(\Delta_{avg})$ vs $\log(n)$, suggesting that the above linear relationship is reasonable. These graphs are shown in Figures 3.1 and 3.2. Estimated γ's for random censoring with 10%, 30%, 50% are $0.38, 0.42, 0.43$ and for covariate dependent with censoring rates 10%, 30%, 50% cases are 0.32, 0.31, 0.35 respectively.

3.2.2 Treatment Selection

In this section we provide details of an extensive simulation study that investigates the properties of the proposed treatment selection method under various settings. The overall study design is comparable to the empirical study described in Project 1. Primarily, we focused on the accuracy of treatment assignment of a new (test) observation by using estimated values of the p_i functions from a set of training data. Illustrating the personalized treatment concept, we choose our model sets such that each model in a set dominates others for some combination of covariate values. In this simulation study, we address both two ($K = 2$) and multiple groups ($K \geq 2$) treatment selection, under random and covariate dependent censoring. We simulated $K, K \in \{2, 3, 4\}$ independent samples of size $n, n \in \{50, 100, 200\}$ per group. All components of r dimensional covariate vectors were generated indepen-

dently from $U(0,1)$, where r is ranged from 2 to 8. Using various functions and index vectors, we obtained the survival outcome from model (2.3), where σ was selected from the set $\{0.1, 0.2\}$. For the kth, $k = 1, .., K$ group, we generated random censoring time (C_k) using single parameter Exponential distribution with parameter ϕ_k, whereas covariate dependent censoring was obtained using a combination of two single parameter Exponential distributions, given by following expression.

$$C_k \sim I(\beta_c^{'} X > w_k) exp(\phi_{k1}) + I(\beta_c^{'} X \leq w_k) exp(\phi_{k2}), \tag{3.6}$$

Here, (ϕ_{k1}, ϕ_{k2}), β_c, and w_k were chosen specifically for the kth model given by (3.6), controlling the censoring percentage. We provide all selected censoring parameters corresponding to each scenario in Tables 3.5 - 3.8. Once K samples were generated, we estimated SIM for each group using the re-weighted SIM estimator applying a Gaussian kernel. The probability of not being censored ($K_c(.)$) was estimated using Kaplan-Meier and Aalen's estimators for random and covariate dependent censoring scenarios, respectively. A new covariate value X_0 was generated in the same manner as generating training set. After that, using the estimated re-weighted SIM models, scores were calculated at each covariate value including the estimated score at x_0 ($\hat{u}_0$). We estimated $\hat{p_k}(\hat{u}_0)$ and $\hat{\mu_k}(\hat{u}_0)$ for $k, k = 1, ..., K$, choosing a $U(-1,1)$ probability density function as the kernel. The bandwidth was chosen by Wand and Jones (1995) procedure for each $i, i = 1, ..., K$. The treatment group for the new patient was determined via (3.3) and (3.5). Using the treatment mean model $g_i(\beta_i \mathrm{X}_0)$ we then generated the response Y_i^*, for each treatment group as in model (2.3) with the same error distribution used to generate model sets. We defined the treatment assignment to be correct if,

$$\arg\max_i \{\hat{p}_i(\hat{u}_0)\} = \arg\max_i \{Y_i^*\}.$$

This procedure was then repeated for 1000 times for each combination of selected settings. We also considered an alternative approach based on the Cox method,

where we fit a Cox regression model for the survival outcome of each treatment. These models are used to find the conditional hazards given individuals' covariates, which allows one to estimate the expected survival outcome of a given individual. In this approach the best treatment can be selected as the treatment with the highest expected survival mean. In the case of two groups (Table 3.3), we chose both linear and nonlinear model functions. When linear models are used, the treatment selection based on Cox method had the best accuracy in terms of optimal treatment selection. However, in cases of non-linear model functions, the proposed methods using $p_i(u)$'s and $\mu_i(u)$'s clearly compete the Cox approach. In general, we observed a high selection accuracy for low censoring percentages in both random and covariate dependent censoring settings. Comparing accuracies observed for $p_i(u)$'s and $\mu_i(u)$'s, we noticed that use of $p_i(u)$'s has a relatively higher accuracy than $\mu_i(u)$'s when the censoring rate gets severe. As to be expected, results revealed comparably high accuracies for large model sizes and low error variances. The model set 4 in Table 3.3 violates the SIM condition. We chose this model to demonstrate the robustness of the proposed concepts under the departure from SIM structure. Observed high accuracies reflect the robustness of the proposed concepts against the departure from SIM assumptions.

We performed simulations for Multiple treatment groups ($K = 3, K = 4$) simulations using Type-1 nonlinear models described in "Empirical Studies" of project 1. We chose the common vector as, $\boldsymbol{C'} = \left(1/\sqrt{r}, ..., 1/\sqrt{r}\right)_{1\times r}$ and dimension (r) selected from the set of {3, 5, 8} with error standard deviation $\sigma = 0.1$ and $\sigma = 0.2$. Simulation results that are presented in Table 3.4 for the multiple group case show high accuracies for larger sample sizes and low censoring rates, the same trend that was observed in the two groups case. These results indicate better treatment selection accuracy with $p_i(u)$'s, than $\mu_i(u)$'s. We observed a relative decline in the accuracy as the dimension was increased. However, the observed

result seems to be highly acceptable in consideration of the complexity in treatment selection problem under a high degree of non-linear models and severities of censoring.

3.3 ACTG-175 HIV Clinical Trial

In this section we illustrate our proposed method using a dataset resulted from ACTG 175 clinical trial. A complete description of this trial and the data set is given in Section 2.3.

This trial measured the survival of HIV patients as one of its secondary outcomes, which had been severally subjected to right-censoring. A primary analysis of the survival outcome shows the rate of censoring is approximately 80%. In our analysis, we considered the log transformed number of survived days as the survival outcome. As covariates, we used log-CD4 counts and log-CD8 counts at baseline, age, weight, and the number of months a patient received pre-antiviral therapy.

We applied proposed treatment assignment strategies based on two proposed concepts: probabilities of dominances and smooth means, on the survival outcome using all four arms of the study. We randomly selected 200 patients from each arm as "training" data to estimate the SIM's, while considering the remaining patients as new (test) patients, which are considered for the treatment selection. After fitting SIMs to training data we estimated the scores for the test cases and estimated the corresponding $p_i(u)$'s and $\mu_i(u)$'s functions using Gaussian kernels at corresponding scores to assign each test patient to the best treatment suggested by the corresponding largest estimated $p_i(u)$ or $\mu_i(u)$ value.

We present two treatment assignment results in Table 3.9. For example, in Table 3.9, the (0, 0) cell indicates that 141 out of 322 patients who were actually treated in arm-0 would have been assigned to arm-0 had we used the proposed

method based on for probability of dominance. The overall result shows a notable difference between treatment assignments for proposed techniques. For example, the major proportion of test patients have been assigned to arm-1, using the probabilities of dominance. On the other hand, the smooth means approach has assigned a large number of patients to arm-0. Such disparity in treatment selection was observed throughout our simulation studies also, especially when the censoring rate was severe. Overall these simulation studies seem to indicate a better performance in treatment assignment with probabilities of dominances.

In the multiple treatments assignment setting, we have a total of 1336 patients in the test set. Using probabilities of dominances, we assigned the majority; 553 to arm-2, whereas 42, 254 and 477 patients are assigned to arms 0, 1 and 3, respectively. This result seemingly suggesting that one of the other arms almost dominate arm-0 with respect to our scoring mechanism when we use $p_i(u)$'s.These results are summarized in Table 3.11. Similar to the two treatment assignment we noticed a clear difference between two methods. A close inspection of the overall assignment indicates that, a large number of patients; 1105 and 976 are proposed to be assigned to a different treatment arm than their actual assignment in the original trial with selectionss based on $p_i(u)$'s and $\mu_i(u)$'s respectively.

3.4 Discussion

In this work, we developed a novel personalized treatment selection concept that addresses the multiple treatment selection for survival outcome which can be subjected to right-censoring. This method is an extension of the project-1 described in the previous chapter.

We introduced IPCW re-weighting schemes to Ichimura et al. (1993) SIM estimator and the treatment selection estimator introduced in project-1 to handle

the selection bias caused by the censoring mechanism. Similar to the method based on probabilities of dominances, we also introduce a treatment selection concept using conditional means. Based on an empirical study that evaluate the performance of the re-weighted SIM estimator, we observed the absolute error associated with the estimated function goes away asymptotically as sample size increases. The performance of the new treatment selection concept was investigated using a series of simulation studies to study the of accuracy in selecting the best treatment that maximizes the patients' survival outcome. Where we considered both two treatment and multiple treatments options under various model functions with various rates of random and covariate dependent censoring. The overall empirical results indicated a reasonable treatment selection accuracy. The proposed method seems to be robust under the deviation from SIM conditions. Between two proposed techniques; probabilities of dominances and smooth means, our empirical studies suggested a comparable performance with both methods under low rates of censoring. However when the censoring rate is severe, the method based on probability of dominances is performing relatively better than method based on smooth means. Demonstrating the applicability of our method in real data, an application of the new concept using ACTG 175 HIV trial data showed acceptable treatment allocations and its potential of maximizing the survival outcome.

Since the treatment selection method introduced in project-1 can be utilized with quantile regression models, for a greater applicability, our method can be further generalized by adjusting quantile regression SIM estimators using a suitable re-weighting concept in the same fashion. Often, outcomes of a treatment is not only restricted to a single response. For example, one may consider patients' CD4 counts and survival as paired responses for a HIV patient. A possible extension of our method could be the addressing of such complex treatment selection problems that deal with maximizing various outcomes which may include potentially right-

censoring.

3.5 Tables

Censoring Rate	Sample Size	Bias (10^{-3})			S.E (10^{-2})			$\Delta(10^{-1})$
		β_1	β_2	β_3	β_1	β_1	β_1	
50%	100	4.449	-0.875	0.712	9.478	6.975	5.749	2.554
	500	-0.226	0.078	0.578	1.574	1.129	2.196	1.217
	1000	-0.137	0.074	-0.138	1.151	0.672	1.302	0.944
	2000	0.016	-0.018	0.004	1.563	0.536	0.381	0.701
30%	100	-0.209	0.120	0.027	3.679	2.678	4.003	1.836
	500	-0.078	0.043	-0.047	1.221	0.713	1.397	0.945
	1000	-0.078	0.053	-0.024	0.614	0.394	0.760	0.713
	2000	-0.009	0.004	0.001	0.340	0.213	0.478	0.519
10%	100	0.217	-0.274	-1.451	1.887	1.179	2.227	1.152
	500	-0.297	0.384	-0.587	0.509	0.317	0.605	0.597
	1000	-0.264	0.149	0.020	0.350	0.216	0.383	0.488
	2000	-0.092	0.092	0.416	0.238	0.175	0.253	0.369

Table 3.1: Properties of re-weighted SIM model for randomly generated censoring times, evaluated with 1000 Monte-Carlo simulations.

a+sin$\{\pi + \pi(\boldsymbol{\beta'X})\}$ a+sin$\{\frac{3\pi}{2} + \pi(\boldsymbol{\beta'X})\}$

a+sin$\{\pi + \pi(\boldsymbol{\beta'X})\}$

Censoring Rate	Sample Size	Bias (10^{-3})			S.E. (10^{-2})			$\Delta(10^{-1})$
		β_1	β_2	β_3	β_1	β_1	β_1	
50%	100	-7.610	1.631	-37.814	9.400	6.648	10.493	3.243
	500	0.550	7.485	-28.060	2.571	1.785	3.339	1.667
	1000	2.244	6.336	-26.591	1.685	1.142	2.206	1.404
	2000	2.724	6.210	-26.281	1.092	0.782	1.609	1.272
30%	100	5.673	5.502	-18.270	4.831	3.231	5.694	2.109
	500	0.856	3.542	14.011	1.4381	1.010	1.938	1.192
	1000	0.712	3.291	-12.064	0.877	0.604	1.183	0.984
	2000	0.855	3.035	-11.305	0.556	0.400	0.883	0.858
10%	100	-0.918	1.518	4.343	1.675	1.132	2.092	1.124
	500	0.082	0.773	-0.265	0.528	0.361	0.655	0.619
	1000	0.108	0.642	2.213	0.3403	0.231	0.393	0.495
	2000	0.090	0.526	1.793	0.2190	0.148	0.266	0.399

Table 3.2: Properties of re-weighted SIM model for covariate dependent censoring times, evaluated with 1000 Monte-Carlo simulations.

Models	Sample Size	Error SD	No Censoring			Random Censoring									Covariate Dependent Censoring								
						10%			30%			50%			10%			30%			50%		
	(n)	(σ)	PD	SM	Cox	PD	SM	Cox	PD	SM	Cox	PD	SM	Cox	PD	SM	Cox	PD	SM	Cox	PD	SM	Cox
A1: Group 1 A2: Group 2	50	0.1	916	920	942	906	883	947	857	802	940	772	677	940	910	882	947	869	811	946	789	714	943
		0.2	878	878	902	877	867	899	828	774	899	755	703	899	879	860	906	838	798	902	765	697	891
	100	0.1	946	949	953	930	922	949	887	844	943	837	746	943	922	901	949	923	864	953	865	781	946
		0.2	899	905	909	886	880	904	869	817	902	802	719	902	890	864	910	893	838	914	851	770	900
	200	0.1	950	950	951	934	922	948	920	869	951	890	810	952	940	927	951	929	883	948	917	816	951
		0.2	905	904	909	897	891	909	889	836	906	865	797	909	909	897	908	895	856	904	873	797	908
A3: Group 1 A4: Group 2	50	0.1	899	884	594	846	838	602	785	760	619	739	709	603	865	842	603	810	770	610	737	701	606
		0.2	836	834	587	822	814	599	778	767	604	711	684	585	838	820	586	785	760	602	724	694	590
	100	0.1	902	895	605	899	887	629	856	835	623	792	776	621	895	888	619	888	846	623	825	749	626
		0.2	856	861	597	865	856	629	825	809	608	770	737	612	860	861	619	852	826	613	791	738	620
	200	0.1	922	917	657	893	891	646	892	863	631	850	823	638	913	914	661	892	864	631	860	808	643
		0.2	893	889	654	857	852	641	843	825	622	835	796	630	885	879	651	870	853	625	837	790	639
A5: Group 1 A6: Group 2	50	0.1	854	855	534	823	807	563	769	749	600	682	654	556	818	797	561	776	755	578	651	635	561
		0.2	817	814	530	797	793	558	743	715	567	662	650	557	804	797	575	743	725	562	649	614	541
	100	0.1	874	874	565	834	832	582	804	786	576	782	738	580	864	853	563	811	783	556	759	716	557
		0.2	845	843	553	802	793	586	785	755	579	739	707	574	842	832	568	803	780	568	732	682	548
	200	0.1	904	903	566	862	863	567	844	821	608	818	771	593	867	864	588	852	826	601	812	784	600
		0.2	862	868	580	839	850	564	825	804	600	792	755	580	843	845	575	819	799	601	800	763	595
A7: Group 1 A8: Group 2	50	0.1	798	784	532	757	743	537	673	664	547	539	550	529	736	725	533	603	606	526	505	478	527
		0.2	756	751	547	727	726	537	630	633	540	568	568	534	751	738	528	581	591	507	490	463	531
	100	0.1	843	838	557	844	836	554	759	743	545	627	604	542	824	821	521	741	725	556	588	564	525
		0.2	799	798	550	820	815	564	751	728	546	599	591	542	808	808	525	710	689	549	570	548	536
	200	0.1	892	885	564	877	864	549	835	796	563	754	712	540	872	858	526	820	804	528	658	628	547
		0.2	866	867	578	839	835	548	804	778	553	716	687	540	832	829	544	782	781	539	650	634	539

Table 3.3: Frequencies of correct treatment assignments in 1000 test cases for three methods: Probability of Dominance (PD), Smooth Means (SM), and Cox model approach, in two groups case ($K = 2$) using linear and nonlinear regression models that are listed in Table 3.6, under random and covariate dependent censoring.

Groups	Models	Dimension	Sample Size	Error SD	No		Random Censoring						Covariate Dependent Censoring					
					Censoring		10%		30%		50%		10%		30%		50%	
		(r)	(n)	(σ)	PD	SM	PD	SM	PD	SM	PD	SM	PD	SM	PD	SM	PD	SM
Three	B1: Group 1 B2: Group 2 B3: Group 3	3	50	0.1	928	926	886	858	844	760	628	582	905	892	856	792	737	667
				0.2	878	878	851	817	784	712	619	578	866	846	824	782	672	619
			100	0.1	935	935	907	881	876	821	785	681	924	919	897	847	853	762
				0.2	893	892	852	831	843	800	748	668	871	847	857	830	805	727
			200	0.1	938	938	935	917	891	833	855	740	935	931	921	880	892	785
				0.2	902	900	884	871	862	821	798	695	896	890	874	843	837	772
		5	50	0.1	923	921	875	847	768	722	579	580	865	855	805	752	600	588
				0.2	847	847	821	802	736	708	560	575	828	817	763	735	564	571
			100	0.1	925	925	880	854	853	793	748	680	913	892	881	840	766	714
				0.2	882	880	867	833	831	765	688	637	871	858	846	806	737	694
			200	0.1	933	933	911	889	879	821	828	735	937	912	903	849	847	774
				0.2	893	893	869	846	842	783	790	702	889	877	854	827	813	751
		8	50	0.1	929	927	891	877	835	806	683	698	897	892	863	832	624	612
				0.2	886	886	861	849	792	781	655	659	846	847	806	792	607	599
			100	0.1	943	939	929	909	897	864	796	757	940	928	898	870	801	762
				0.2	904	904	883	865	857	816	769	744	898	894	879	841	766	728
			200	0.1	950	949	943	925	914	876	877	805	940	939	933	894	892	836
				0.2	917	917	896	880	888	844	853	776	911	904	896	873	846	810
Four	B1: Group 1 B2: Group 2 B3: Group 3 B4: Group 4	3	50	0.1	926	924	866	835	826	754	702	628	903	880	857	790	748	664
				0.2	878	880	838	827	782	714	645	605	866	845	797	753	712	638
			100	0.1	939	939	902	874	865	792	801	700	918	904	894	830	846	724
				0.2	883	886	859	836	840	759	757	649	886	882	858	820	798	710
			200	0.1	945	945	910	903	893	846	838	741	929	921	904	866	884	790
				0.2	885	884	875	855	851	818	805	729	887	882	872	842	831	752
		5	50	0.1	914	918	857	834	781	703	603	587	865	847	790	749	638	610
				0.2	831	831	793	782	740	681	542	543	819	805	765	730	595	588
			100	0.1	922	922	892	873	862	797	748	664	911	889	871	815	779	705
				0.2	874	875	842	823	792	729	717	674	857	842	818	775	744	680
			200	0.1	938	937	911	886	882	832	845	741	924	908	907	866	872	790
				0.2	872	872	848	835	838	798	786	736	862	852	847	811	821	751
		8	50	0.1	887	887	837	820	744	725	585	603	857	853	732	717	580	591
				0.2	833	834	791	766	714	692	542	568	798	789	709	700	548	560
			100	0.1	918	917	871	855	830	799	690	693	914	909	847	801	719	716
				0.2	874	875	834	828	773	746	702	671	860	852	788	759	708	685
			200	0.1	928	929	908	895	877	827	823	751	915	900	896	871	829	775
				0.2	890	889	869	853	830	793	782	710	877	863	849	803	785	751

Table 3.4: Frequencies of correct treatment assignments in 1000 test cases by two methods: Probabilities of Dominances (PD) and Smooth Means (SM), in multiple groups case ($K > 2$), using nonlinear regression models that are shown in Table 3.7.

Model Set	Group	θ			$\binom{\phi_1}{\phi_2}$			w	$\boldsymbol{\beta}_c'$
		20%	30%	50%	20%	30%	50%		
1	1st	0.10	0.20	0.40	$\binom{0.05}{0.30}$	$\binom{0.10}{0.40}$	$\binom{0.20}{0.70}$	0.6	(1,1,0)
	2nd	0.10	0.20	0.40	$\binom{0.05}{0.30}$	$\binom{0.10}{0.40}$	$\binom{0.20}{0.70}$	0.6	(1,1,0)
2	1st	0.15	0.25	0.40	$\binom{0.05}{0.30}$	$\binom{0.10}{0.40}$	$\binom{0.20}{0.70}$	0.6	(1,1,0)
	2nd	0.25	0.35	0.70	$\binom{0.05}{0.30}$	$\binom{0.10}{0.40}$	$\binom{0.20}{0.70}$	0.6	(1,1,0)
3	1st	0.15	0.25	0.40	$\binom{0.05}{0.15}$	$\binom{0.20}{0.35}$	$\binom{0.20}{0.50}$	0.3	(1,1,1,1,0,0,0,0)
	2nd	0.25	0.55	0.80	$\binom{0.20}{0.30}$	$\binom{0.30}{0.70}$	$\binom{0.60}{1.00}$	0.3	(1,1,1,1,0,0,0,0)
4	1st	0.10	0.20	0.40	$\binom{0.05}{0.20}$	$\binom{0.10}{0.40}$	$\binom{0.40}{0.60}$	0.4	(1,0,0)
	2nd	0.10	0.20	0.40	$\binom{0.20}{0.30}$	$\binom{0.30}{0.70}$	$\binom{0.60}{1.00}$	0.4	(1,0,0)

Table 3.5: The used sets of parameters to generate censoring times for two groups assignment cases.

Regression Functions
A1: $2.0 + \frac{1}{\sqrt{3.19}}(1.9X_1 + 2X_2 - 1.6X_3)$
A2: $1.8 + \frac{1}{\sqrt{6.52}}(4.8X_1 - 3.5X_2 + 2.7X_3)$
A3: $1.2+\sin\{\frac{\pi}{\sqrt{3.19}}(0.7X_1 + 1.2X_2 - 1.1X_3)\}$
A4: $1.2+\cos\{\frac{\pi}{\sqrt{6.62}}(0.6X_1 - 0.1X_2 + 2.5X_3)\}$
A5: $1.2+\sin\{\frac{\pi}{\sqrt{4.28}}(X_1 - 0.3X_2 + 0.7X_3 + 0.1X_4 + 0.8X_5 - 0.6X_6 + 0.5X_7 + 1.2X_8)\}$
A6: $1.2+\sin\{\pi + \frac{\pi}{\sqrt{11.65}}(2X_1 + X_2 + 0.5X_3 + 0.5X_4 - 0.7X_5 + 0.1X_6 + 2.3X_7 - 0.6X_8)\}$
A7: $1.2+\cos\{\frac{\pi}{\sqrt{1.26}}(X_1 - 0.5X_2 + 0.1X_3)\} + 0.4X_1^2$
A8: $1.2+\cos\{\frac{3\pi}{2} + \frac{\pi}{\sqrt{7.13}}(1.6X_1 + 2.1X_2 + 0.4X_3)\} + 0.2X_1^2$

Table 3.6: Model functions that are used for two group treatment selection simulations. Table 3.5 shows the used censoring parameters.

Regression Functions
B1: a+sin$\{\pi(\boldsymbol{\beta'X})\}$
B2: a+sin$\{\frac{3\pi}{2}+\pi(\boldsymbol{\beta'X})\}$
B3: a+sin$\{\pi(\boldsymbol{\beta'X})\}$
B4: a+sin$\{\frac{-\pi}{2}+\pi(\boldsymbol{\beta'X})\}$

Table 3.7: Model functions that are used for multiple treatment selection simulations. Here, $\boldsymbol{\beta'} = \left(1/\sqrt{r}, ...1/\sqrt{r}\right)_{1\times r}$. Table 3.8 shows the used censoring parameters.

Number of Groups	Dimension (r)	Groups	θ			$\binom{\phi_1}{\phi_2}$			w	$\boldsymbol{\beta}_c'$
			20%	30%	50%	20%	30%	50%		
3	3	1st	0.15	0.25	0.50	$\binom{0.10}{0.20}$	$\binom{0.10}{0.40}$	$\binom{0.30}{0.70}$	1.00	(1,1,0)
		2nd	0.10	0.20	0.40	$\binom{0.10}{0.20}$	$\binom{0.10}{0.40}$	$\binom{0.30}{0.70}$	1.00	(1,1,0)
		3rd	0.20	0.40	0.80	$\binom{0.10}{0.30}$	$\binom{0.20}{0.60}$	$\binom{0.30}{1.00}$	1.20	(1,1,0)
	5	1st	0.20	0.40	0.70	$\binom{0.10}{0.40}$	$\binom{0.30}{0.60}$	$\binom{0.50}{1.00}$	1.00	(1,1,1,0,0)
		2nd	0.10	0.20	0.40	$\binom{0.10}{0.20}$	$\binom{0.20}{0.50}$	$\binom{0.30}{1.00}$	1.00	(1,1,1,0,0)
		3rd	0.20	0.30	0.60	$\binom{0.10}{0.30}$	$\binom{0.20}{0.60}$	$\binom{0.40}{1.00}$	1.20	(1,1,1,0,0)
	8	1st	0.30	0.60	1.20	$\binom{0.40}{0.60}$	$\binom{0.50}{1.00}$	$\binom{0.80}{1.50}$	1.00	(1,1,1,1,1,0,0,0)
		2nd	0.20	0.30	0.50	$\binom{0.20}{0.40}$	$\binom{0.30}{0.50}$	$\binom{0.50}{1.00}$	1.00	(1,1,1,1,1,0,0,0)
		3rd	0.10	0.20	0.40	$\binom{0.10}{0.40}$	$\binom{0.20}{0.40}$	$\binom{0.50}{1.00}$	1.00	(1,1,1,1,1,0,0,0)
4	3	1st	0.15	0.25	0.50	$\binom{0.15}{0.30}$	$\binom{0.20}{0.50}$	$\binom{0.50}{1.00}$	0.40	(1,1,0)
		2nd	0.30	0.60	1.20	$\binom{0.10}{0.70}$	$\binom{0.30}{0.90}$	$\binom{0.50}{1.20}$	1.00	(1,1,0)
		3rd	0.15	0.20	0.40	$\binom{0.10}{0.40}$	$\binom{0.20}{0.50}$	$\binom{0.20}{0.70}$	0.60	(1,1,0)
		4th	0.25	0.40	0.80	$\binom{0.20}{0.40}$	$\binom{0.40}{0.60}$	$\binom{0.70}{1.30}$	0.50	(1,1,0)
	5	1st	0.15	0.25	0.50	$\binom{0.20}{0.60}$	$\binom{0.50}{1.00}$	$\binom{0.80}{1.50}$	0.40	(1,1,1,0,0)
		2nd	0.30	0.60	1.20	$\binom{0.10}{0.70}$	$\binom{0.30}{0.90}$	$\binom{0.50}{1.20}$	1.00	(1,1,1,0,0)
		3rd	0.10	0.20	0.40	$\binom{0.10}{0.40}$	$\binom{0.20}{0.50}$	$\binom{0.20}{0.70}$	0.60	(1,1,1,0,0)
		4th	0.25	0.40	1.00	$\binom{0.20}{0.40}$	$\binom{0.40}{0.60}$	$\binom{0.70}{1.30}$	0.50	(1,1,1,0,0)
	8	1st	0.15	0.25	0.50	$\binom{0.30}{0.50}$	$\binom{0.40}{0.80}$	$\binom{0.60}{1.50}$	0.60	(1,1,1,1,1,0,0,0)
		2nd	0.30	0.60	1.20	$\binom{0.10}{0.70}$	$\binom{0.30}{0.90}$	$\binom{0.60}{1.30}$	1.50	(1,1,1,1,1,0,0,0)
		3rd	0.10	0.20	0.40	$\binom{0.10}{0.40}$	$\binom{0.20}{0.50}$	$\binom{0.50}{1.30}$	0.90	(1,1,1,1,1,0,0,0)
		4th	0.25	0.40	1.00	$\binom{0.10}{0.30}$	$\binom{0.20}{0.40}$	$\binom{0.40}{1.00}$	0.75	(1,1,1,1,1,0,0,0)

Table 3.8: The used sets of parameters to generate censoring times for multiple groups ($K > 2$) assignment cases.

	PD		SM	
	Arm-0	Arm-1	Arm-0	Arm-1
Arm-0	141	191	225	107
Arm-1	119	200	195	124
Total	260	391	420	231

Table 3.9: Two groups treatment assignment summary for ACTG-175 trial, by two proposed techniques: Probability of Dominance (PD) and Smooth Means (SM).

Table 3.10: My caption

	PD				SM			
	Arm-0	Arm-1	Arm-2	Arm-3	Arm-0	Arm-1	Arm-2	Arm-3
Arm-0	6	62	137	127	114	82	69	67
Arm-1	7	64	138	110	104	94	63	58
Arm-2	12	61	136	115	117	65	64	78
Arm-3	17	67	152	125	133	87	53	88
Total	42	254	553	477	468	328	249	291

Table 3.11: Four groups treatment assignment summary for ACTG-175 clinical trial, by two proposed techniques: Probability of Dominance (PD) and Smooth Means (SM).

3.6 Figures

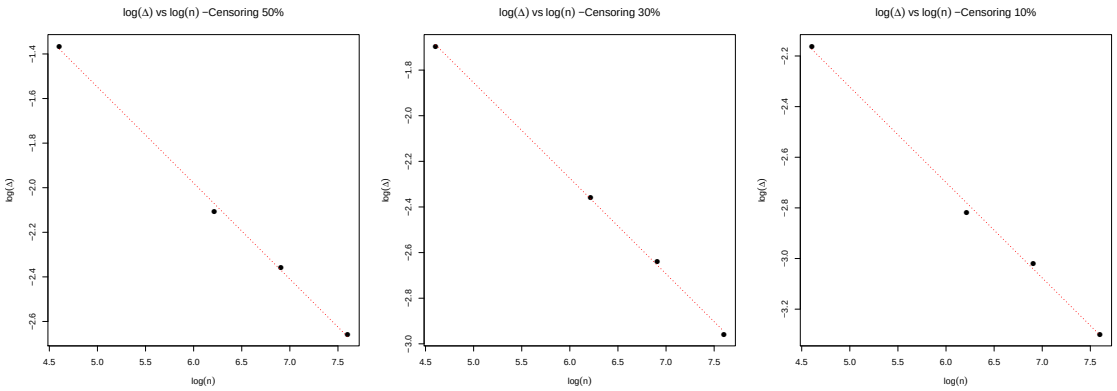

Figure 3.1: Graphs of $\log(\Delta) vs \log(n)$ by re-weighted SIM model, for randomly censored cases, with censoring percentages from left to the right: 50%, 30%, 10%. Dotted line represent the fitted liner line for the data.

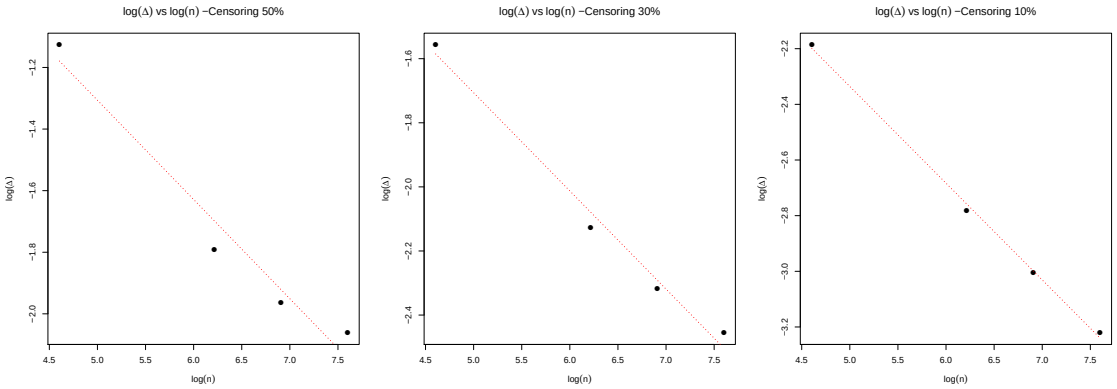

Figure 3.2: Graphs of $\log(\Delta) vs \log(n)$ by re-weighted SIM model, for covariate dependent censored cases, with censoring percentages from left to the right: 50%, 30%, 10%. Dotted line represent the fitted liner line for the data.

CHAPTER 4
PERSONALIZED TREATMENT PLANS WITH MULTIVARIATE OUTCOME MEASURES

4.1 Treatment Selection

In this section we describe the proposed procedure. Let $(\mathbf{Y}_i^*, \boldsymbol{X})$ be the hypothetical (counterfactual) response vector and covariate pair for treatment i, $i = 1, \ldots, K$ where without loss of generality larger values of the each component of the q dimensional response vectors $\mathbf{Y}_i^* = (Y_{1i}^*, ..., Y_{qi}^*)'$ are indicative of better outcomes and $\boldsymbol{X}$ is a vector of r covariates. Assume further that a patient's covariate value $\boldsymbol{X}$ is used to obtain a lower dimensional composite patient score $U(\boldsymbol{X})$. In practice one cannot observe the whole composite vector $(\mathbf{Y}_1^*, ..., \mathbf{Y}_K^*)'$ for a single patient.

In the single response case ($q = 1$) using iid observations of type $(\tilde{Y}_{1i}, \boldsymbol{X}_i, A_i)$, $i = 1, ..., n$ where A_i is the binary treatment indicator for two treatments and $\tilde{Y}_{1i}$ is the observed single response for the ith 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

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

$k = 1, ..., K$ for a suitable score function U as the best treatment. In contrast to this approach, in the 1st project we consider pairs of independent observations $(Y_{1k}, \boldsymbol{X}_k)$ from the marginal distribution of $(Y^*_{1k}, \boldsymbol{X}), k = 1, ..., K$ for the treatment selection for K treatments. In that work they proposed a method based on estimators of a set of probabilities defined as

$$p_k(u) = P[Y_{1k} > \max_{k \neq j} Y_{1j} | U(\boldsymbol{X_l}) = u; l = 1, ..., K]; k = 1, ..., K \tag{4.1}$$

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

$$\mu_k = E\left[Y_{1k} | U(\boldsymbol{X_k}) = u\right] \tag{4.2}$$

$k = 1, ..., K$ as the best treatment. Note that in (4.1) and (4.2), 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 $\boldsymbol{X}$). Although the function $p_i(u)$ above does not use the joint distribution of $(Y^*_{11}, ..., Y^*_{1K}, \boldsymbol{X})$ for a patient with covariate value $\boldsymbol{X}$, they argue that p_i above nevertheless gives a measure of dominance for the ith treatment over the others and hence can be used in selecting the best treatment. This was suggested as an alternative to measures based on conditional expectations $(\mu_k, k = 1, ..., K)$ which require restrictive moment assumptions on the error distribution for inferential aspects in a regression context. The method based on the p_is was very competitive against methods based on μ_is for a variety of models as shown from their empirical studies.

In dealing with multiple responses, we consider pairs of independent observations $(\mathbf{Y}_k, \boldsymbol{X}_k)$ from the marginal distribution of $(\mathbf{Y}^*_k, \boldsymbol{X}), k = 1, ..., K$ 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 (4.1)

above. For example, in generalizing the approach of using the conditional means for the response vector $\mathbf{Y}_k = (Y_{1k}, ..., Y_{qk})'$ we define

$$\mu_{ik}(u_i) = E[Y_{ik}|U_i(\boldsymbol{X}_k) = u_i]; i = 1, ..., q; k = 1, ..., K \tag{4.3}$$

and vectors $\boldsymbol{\mu}_k(\mathbf{u}) = (\mu_{1k}(u_1), ..., \mu_{qk}(u_q))'$ 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 to get size K vectors $\mathbf{v}_i(\mathbf{u}) = (v_{i1}(\mathbf{u}), ..., v_{iK}(\mathbf{u}))'$ where $v_{ik}(\mathbf{u})$ is the rank of μ_{ik} among $\mu_{ik}, k = 1, ..., K$ for each i with the largest μ_{ik} 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 work 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 $\omega_i; i = 1, ..., q$ and a distance measure γ (Pihur et al. 2007), we minimize a quantity

$$\psi(\mathbf{v}) = \sum_{i=1}^{q} \omega_i \gamma(\mathbf{v}, \mathbf{v}_i(\mathbf{U}_0)) \tag{4.4}$$

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

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

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

$$k^*(\mathbf{U}_0) = \arg \min_{1 \le k \le K} \{v_k^*\} \tag{4.5}$$

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 $\mu_{ik}, i = 1, ..., 4; k = 1, ..., 3$ and their corresponding ranks in the weighted Spearman's Footrule distance γ combined with $\omega_i = 1, i = 1, ..., 4$ 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 project-1 to rank the treatments, we consider

$$p_{ik}(u_i) = P[Y_{ik} > \max_{k \neq j} Y_{ij} | U_i(\boldsymbol{X_l}) = u_i; l = 1, ..., K]; i = 1, ..., q; k = 1, ..., K \tag{4.6}$$

and use the same aggregation method above to vectors of ranks corresponding to $\boldsymbol{p}_k(\mathbf{u}) = (p_{1k}(u_1), ..., p_{qk}(u_q))'$, $k = 1, ..., K$ in a similar fashion. We base our discussion on a set of Single Index Models relating the ith component Y_{ik} of the response vector $\mathbf{Y}_k$ for the kth treatment and covariates $\boldsymbol{X}_k$ via

$$Y_{ik} = g_{ik}\left(\boldsymbol{\beta}_{ik}' \boldsymbol{X}_k\right) + \epsilon_{ik} \tag{4.7}$$

for $i = 1, \ldots, q$ and $k = 1, ..., K$ where each $\boldsymbol{\beta}_{ik}$ is a r-vector of parameters, g_{ik}s are unknown link functions for which we assume some reasonable smoothness conditions to hold, and ϵ_{ik} are error terms with $E[\epsilon_{ik}|\boldsymbol{X}] = 0$. Furthermore we assume independence of ϵ_{ik}s across $k = 1, ..., K$ for a fixed i where these terms are corre-

lated across is for any given k. The Single Index formulation provides flexibility and reasonable efficiency in modeling many types of data.

Our observations are of the following form. Let Y_{ikj} indicate the ith component of the jth response from a group of n_k individuals under treatment k with associated covariate values $\boldsymbol{X}_{kj}, j = 1, \ldots, n_k$. The 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 (4.7) is written as

$$Y_{ikj} = g_{ik}(\boldsymbol{\beta}'_{ik}\boldsymbol{X}_{kj}) + \epsilon_{ikj}, j = 1, \ldots, n_k. \tag{4.8}$$

Following Siriwardhana et al. (2015) we define an appropriate overall score vector $\mathbf{U}$ as follows. First define

$$S_{ik}\left(\boldsymbol{X}\right) = g_{ik}\left(\boldsymbol{\beta}'_{ik}\boldsymbol{X}\right) - \max_{l \neq k}\left\{g_{il}\left(\boldsymbol{\beta}'_{il}\boldsymbol{X}\right)\right\}.$$

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

$$\begin{aligned} S_i\left(\boldsymbol{X}\right) &= \max_k\left\{S_{ik}\right\} \\ \delta_i\left(\boldsymbol{X}\right) &= \arg\max_k\left\{S_{ik}\right\}. \end{aligned} \tag{4.9}$$

Then, for a patient with covariate value $\boldsymbol{X}$ we define the patient score as $\mathbf{U}(\boldsymbol{X}) = (U_1, ..., U_q)'$ where $U_i = (S_i, \delta_i)'$ for $i = 1, ..., q$. In practice one does not know the error distributions and model functions for models defined in (4.7) and therefore we cannot directly calculate either the $\boldsymbol{\mu}_k$s or $\boldsymbol{p}_k$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_{ikj} values as well as observed $U_i, i = 1, ..., q$ values corresponding to those responses. However, $\mathbf{U}$s defined above are hypothetical scores for a covariate value $\boldsymbol{X}$ as we do not know link functions g_{ik}s and index vectors $\boldsymbol{\beta}_{ik}$s. Hence, in estimating p_{ik}s and μ_{ik}s, we propose to use "estimated" $\mathbf{U}(\boldsymbol{X}_{kj})$ values , $\hat{\mathbf{U}}(\boldsymbol{X}_{kj})$, say, corresponding

to responses $Y_{ikj}, i = 1, ..., q; j = 1, ..., n_k; k = 1, ..., K$. Now, to obtain $\hat{\mathbf{U}}(\boldsymbol{X}_{ij})$ values, suitable estimators of link functions g_{ik}s and index vectors $\boldsymbol{\beta}_{ik}$s can be used to construct estimators $\hat{S}_i(\boldsymbol{X})$ and $\hat{\delta}_i(\boldsymbol{X})$ of $S_i(\boldsymbol{X})$ and $\delta_i(\boldsymbol{X})$, respectively. Estimators of the link functions and index vectors can be obtained using responses for each i,$i = 1, ..., q$ coupled with the corresponding covariate observations for any given k, $k = 1, ..., K$, 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 Ichimura et al. (1993) in our simulations and data analysis in the sequel. In the sequel these estimators will be generically denoted by $\hat{g}_{ik}$ and $\hat{\boldsymbol{\beta}}_{ik}$, respectively, for $i = 1, \ldots, q$ and $k = 1, \ldots, K$. In particular, for any given vector $\boldsymbol{x}$, let

$$\hat{S}_{ik}(\boldsymbol{x}) = \hat{g}_{ik}\left(\hat{\boldsymbol{\beta}}'_{ik}\boldsymbol{x}\right) - \max_{l \neq k}\left\{\hat{g}_{il}\left(\hat{\boldsymbol{\beta}}'_{il}\boldsymbol{x}\right)\right\}$$

$$\hat{S}_i(\boldsymbol{x}) = \max_k\left\{\hat{S}_{ik}(\boldsymbol{x})\right\}$$

$$\hat{\delta}_i(\boldsymbol{x}) = \arg\max_k\left\{\hat{S}_{ik}(\boldsymbol{x})\right\}$$

and

$$\hat{U}_i(\boldsymbol{x}) = (\hat{S}_i(\boldsymbol{x}), \hat{\delta}_i(\boldsymbol{x}))' \tag{4.10}$$

We randomly select an index $\hat{\delta}_i$ in the unlikely event that multiple treatments produce the same $\hat{S}_{ik}$.

Now, for a given i, $i = 1, ..., q$, we construct estimators for $\mu_{ik}(u)$ and $p_{ik}(u)$, $k = 1, ..., K$ at a given $u = (s, d)'$ as follows. For estimating μ_{ik} for a given i and k, we first let w be a kernel function with $w \geq 0$ and $\int w(t)dt = 1$, and let $h_l, l = 1, ..., K$ be a set of smoothing parameters. We define an estimator of $\mu_{ik}(u), k = 1, ..., K$ as

$$\hat{\mu}_{ik}(u) = \frac{\sum_{j=1}^{n_k} Y_{ikj} w\left((s - \hat{S}_i(\boldsymbol{X}_{kj}))/h_k\right) I\left(\hat{\delta}_i(\boldsymbol{X}_{kj}) = d\right)}{\sum_{j=1}^{n_k} w\left((s - \hat{S}_i(\boldsymbol{X}_{kj}))/h_k\right) I\left(\hat{\delta}_i(\boldsymbol{X}_{kj}) = d\right)} \tag{4.11}$$

where $I(A)$ is the indicator of A. The estimator $\hat{p}_{ik}(u)$ of $p_{ik}(u)$ for a given u is obtained by 2.7.

For a realization $\boldsymbol{x_0}$ of the covariate $\boldsymbol{X}$, if we knew the corresponding realizations of the scores, $u_{i0} = (S_i(\boldsymbol{x_0}), \delta_i(\boldsymbol{x_0}))'$, we can estimate $\mu_{ik}(u_{i0})$ and $p_{ik}(u_{i0})$ by $\hat{\mu}_{ik}(u_{i0})$ and $\hat{p}_{ik}(u_{i0})$ respectively. However, due to the aforementioned reasons, we can only find an estimate $\hat{u}_{i0}$ of u_{i0} using (4.10) above. Thus, we use $\hat{\mu}_{ik}(\hat{u}_{i0})$ and $\hat{p}_{ik}(\hat{u}_{i0})$ as our estimates of $\mu_{ik}(u_{i0})$ and $p_{ik}(u_{i0})$ respectively for $i = 1, ..., q; k = 1, ..., K$. The estimators $\hat{\mu}_{ik}(\hat{u}_0)$, $i = 1, ..., q; k = 1, ..., K$ and $\hat{p}_{ik}(\hat{u}_0)$ are consistent for $\mu_{ik}(u_{i0})$ and $p_{ik}(u_{i0})$ follows from arguments similar to those given in project-1.

Finally, for either using means or the probabilities, for a given estimated score vector $\hat{\boldsymbol{u}}_0 = (\hat{u}_{10}, ..., \hat{u}_{q0})'$, the estimated best treatment for a patient with covariate value $\boldsymbol{x_0}$ is defined via the minimization of

$$\psi(\boldsymbol{v}) = \sum_{i=1}^{q} \omega_i \gamma(\boldsymbol{v}, \boldsymbol{v_i}(\hat{\boldsymbol{u}}_0)) \tag{4.12}$$

over P_K and defining a

$$\hat{k}^* = \arg \min_{1 \le k \le K} \{\hat{v^*}_k\} \tag{4.13}$$

where $\hat{v^*}_k, k = 1, ..., K$ are the ranks obtained by the minimization of the distance function (4.12) for the corresponding procedure.

Bandwidth selection for estimating the link functions, p_{iK}s and μ_{ik}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.

4.2 Empirical Studies

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

We conducted a series of simulations with the proposed procedure under various settings. Primarily, we focused on the accuracy of treatment assignment of a new (test) observation using estimated values of μ_{ik} and p_{ik} functions from a set of training data. This simulation study was performed for treatment groups cases $K = 2$ 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 $\{0.1, 0.5, 0.9\}$. We examined $r = 3, 8$ and 10 cases. Using various link functions and index vectors, we obtained the treatment responses from model (4.7) for each k. Here, for a given k, $k = 1, ..., K$, the errors were generated from either a q dimensional multivariate normal distribution or a multivariate t distribution with zero mean and a correlation matrix with the ijth element given by $x_i^{|i-j|}$ where x_is 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 $\{0.1, 0.3, 0.5\}$. The degrees of freedom for each marginal was set at 3 and 8 for t variables. We examined the performance of the proposed method-

ology under a variety of both linear and nonlinear regression models with the 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 $\hat{\boldsymbol{u}}_0$, we calculated $\hat{\mu}_{ik}(\hat{u}_{i0})$ and $\hat{p}_{ik}(\hat{u}_{i0})$ for $i = 1, ..., q; k = 1, ..., K$ and the corresponding $\hat{k}^*$ values for equal weights ($\omega_i = 1$ for all i) 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, $\boldsymbol{Y_k^*}, k = 1, ..., K$, each with mean vector $(g_{1k}(\boldsymbol{\beta}'_{1k}\boldsymbol{X_0}), ..., g_{qk}(\boldsymbol{\beta}'_{qk}\boldsymbol{X_0}))'$ for $k = 1, ..., K$, corresponding to this $\boldsymbol{X_0}$ using model (4.7) 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 $\tilde{\boldsymbol{v}}_i$,$i = 1, ..., q$, say, for each row of the data matrix $(\boldsymbol{Y_1^*}, ..., \boldsymbol{Y_K^*})$, and minimize

$$\psi(\boldsymbol{v}) = \sum_{i=1}^{q} \omega_i \gamma(\boldsymbol{v}, \tilde{\boldsymbol{v}}_i) \tag{4.14}$$

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

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

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

We repeated this procedure 1000 times for each model and error distribution combination. The frequency of correct treatment assignment for a representative set of cases are given in the Tables 4.2 - 4.9. The results presented below are for model

functions and index vectors given in Table 4.1 below and for covariate dimension $r = 10$.

An examination of these tables reveal that the selection accuracy drops when the error distribution has a high variability. Both methods, based on smoothed means and the method based on p_{ik} have very comparable selection frequencies in all 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 very marginal in most cases. When the number of responses was 4, selection frequencies appear to get lower as the correlation among the responses increase. We also observed a slight increment in the selection frequency when the number of responses are increased for all group sizes. This is perhaps due to the performance of the rank aggregation method which seems to perform better when aggregating a larger number of ranked lists compared with just two lists.

4.3 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. This trial periodically measured both these cell counts for each patient as the clinical outcome. 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.

We applied the proposed treatment assignment strategy to the data from all four arms of the study. In each situation, we randomly selected 200 patients from each arm as "training" data to estimate the SIMs. Remaining patients were considered as new (test) patients. After fitting SIMs to training data we estimated the scores for test cases and estimated the corresponding p_{ik} and μ_{ik} functions at those scores to assign each test patient to the best treatment group suggested by the rank aggregation method.

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). Therefore, when we applied the proposed method with two responses (both CD4 and CD8), we weighted the importance of two responses differently ($\omega_{CD4} = 0.6$, $\omega_{CD8} = 0.4$), by giving more priority to the CD4.

We report the results for the joint response case (Table 4.10) and for cases when CD4 and CD8 were considered as single responses (Tables 4.11-4.12). For example, in Table 4.10, the (0, 0) cell indicates that only 6 out of 332 patients

who were actually treated in arm-0 would have been assigned to arm 0 had we used the proposed method based on p_{ik}'s. When we applied the new method based on p_{ik}'s for the joint response, the majority: 773, out of 1336 test patients were assigned to arm 1. Similarly, a great number: 668 were assigned to arm 1 using the assignment based on μ_{ik}'s. As shown in Tables 4.11 and 4.12, if we are to only consider CD4 (or CD8) as the response, 828 (or 421) patients were assigned to arm 1 using p_{ik}'s, whereas 640 (or 400) individuals were correspondingly assigned to arm 1 using the method given by μ_{ik}'s. We observed a notable difference between the number of individuals assigned to arm 0 using single responses. For instance, if the CD4 was used as the response, only 14 individuals were assigned to arm 0 using p_{ik}'s. However, comparably more number of individuals: 367 were assigned to arm 0 using the same approach, if the CD8 had been used. Comparing Tables 4.10 and 4.11, we noticed a clear agreement between the resulted overall assignment by the weighted joint response and the single assignment by CD4, which is reasonable since we used a larger weight for the CD4.

4.4 Discussion

In this project we proposed a novel personalized treatment plan to select the optimal treatment from a set of multiple treatments when the outcome measures are multivariate. This method is a single step procedure which can be easily applied. The proposed method is based on semi parametric Single Index Models which add great flexibility in modeling real life situations. Furthermore, this method can also be used for quantile regression SIMs providing additional model flexibility compared with existing methods based on conditional expectations. 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 a possible changes if one were to use multiple outcome measures as opposed to a single measure.

This project 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.

4.5 Tables

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

Table 4.1: Sets of mean functions used to generate treatment responses, with $\boldsymbol{C'} = \big(1/\sqrt{10}, ...1/\sqrt{10}\big)_{1\times 10}$.

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	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	877	844	858	817	833	800
			$\rho_e = 0.5$	901	863	878	852	829	816
			$\rho_e = 0.9$	870	827	871	837	836	829
		$\sigma = 0.5$	$\rho_e = 0.1$	678	680	648	650	678	682
			$\rho_e = 0.5$	626	614	637	633	622	622
			$\rho_e = 0.9$	630	634	629	623	569	569
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	919	862	892	862	913	867
			$\rho_e = 0.5$	911	880	909	871	892	846
			$\rho_e = 0.9$	914	860	911	866	900	859
		$\sigma = 0.5$	$\rho_e = 0.1$	667	671	690	687	679	677
			$\rho_e = 0.5$	626	632	634	636	676	670
			$\rho_e = 0.9$	629	623	619	626	627	630
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
T	$n = 100$	$DF = 3$	$\rho_e = 0.1$	626	626	643	630	603	610
			$\rho_e = 0.5$	597	600	591	593	591	591
			$\rho_e = 0.9$	549	550	541	542	508	499
		$DF = 8$	$\rho_e = 0.1$	640	639	616	626	595	595
			$\rho_e = 0.5$	566	564	600	595	590	581
			$\rho_e = 0.9$	569	568	558	571	540	541
	$n = 200$	$DF = 3$	$\rho_e = 0.1$	646	647	606	605	625	624
			$\rho_e = 0.5$	600	604	580	582	589	596
			$\rho_e = 0.9$	528	530	562	564	535	525
		$DF = 8$	$\rho_e = 0.1$	640	640	640	649	627	631
			$\rho_e = 0.5$	606	605	554	561	589	591
			$\rho_e = 0.9$	553	556	549	555	538	537

Table 4.2: Frequencies of correct treatment assignments in 1000 test cases by the proposed method. Two treatments with two responses.

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	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	864	860	842	841	792	781
			$\rho_e = 0.5$	870	856	816	826	805	809
			$\rho_e = 0.9$	885	875	826	815	799	792
		$\sigma = 0.5$	$\rho_e = 0.1$	535	548	545	544	597	591
			$\rho_e = 0.5$	558	538	521	528	535	524
			$\rho_e = 0.9$	560	556	543	552	506	498
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	913	921	891	898	876	876
			$\rho_e = 0.5$	897	899	878	871	873	870
			$\rho_e = 0.9$	862	860	889	879	864	858
		$\sigma = 0.5$	$\rho_e = 0.1$	587	581	572	581	579	559
			$\rho_e = 0.5$	578	574	552	551	572	576
			$\rho_e = 0.9$	576	556	531	535	536	538
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
T	$n = 100$	$DF = 3$	$\rho_e = 0.1$	507	507	511	516	493	487
			$\rho_e = 0.5$	496	496	526	506	511	519
			$\rho_e = 0.9$	542	542	501	502	504	501
		$DF = 8$	$\rho_e = 0.1$	516	516	491	491	524	527
			$\rho_e = 0.5$	527	527	507	511	484	487
			$\rho_e = 0.9$	513	513	523	510	505	516
	$n = 200$	$DF = 3$	$\rho_e = 0.1$	506	506	525	534	498	499
			$\rho_e = 0.5$	490	490	495	501	518	516
			$\rho_e = 0.9$	499	499	526	517	532	524
		$DF = 8$	$\rho_e = 0.1$	520	520	538	536	516	516
			$\rho_e = 0.5$	510	510	523	521	538	532
			$\rho_e = 0.9$	478	478	519	523	497	491

Table 4.3: Frequencies of correct treatment assignments in 1000 test cases by the proposed method. Two treatments with three responses.

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	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	869	825	812	751	786	754
			$\rho_e = 0.5$	855	795	806	757	811	779
			$\rho_e = 0.9$	886	834	828	775	773	749
		$\sigma = 0.5$	$\rho_e = 0.1$	617	613	640	640	573	576
			$\rho_e = 0.5$	598	593	555	556	575	575
			$\rho_e = 0.9$	557	555	541	538	521	519
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	870	824	863	832	853	811
			$\rho_e = 0.5$	887	861	861	797	858	805
			$\rho_e = 0.9$	875	807	874	819	851	789
		$\sigma = 0.5$	$\rho_e = 0.1$	650	652	626	626	625	621
			$\rho_e = 0.5$	610	609	569	576	602	607
			$\rho_e = 0.9$	594	598	549	555	571	569
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
T	$n = 100$	$DF = 3$	$\rho_e = 0.1$	575	566	559	547	555	560
			$\rho_e = 0.5$	519	520	555	553	554	559
			$\rho_e = 0.9$	541	528	532	537	518	518
		$DF = 8$	$\rho_e = 0.1$	596	609	603	604	576	577
			$\rho_e = 0.5$	572	577	562	565	544	551
			$\rho_e = 0.9$	515	519	508	501	522	513
	$n = 200$	$DF = 3$	$\rho_e = 0.1$	594	593	577	584	558	564
			$\rho_e = 0.5$	577	571	545	545	551	547
			$\rho_e = 0.9$	561	555	539	543	530	526
		$DF = 8$	$\rho_e = 0.1$	611	615	581	594	607	612
			$\rho_e = 0.5$	560	562	540	537	554	549
			$\rho_e = 0.9$	509	505	532	533	503	514

Table 4.4: Frequencies of correct treatment assignments in 1000 test cases by the proposed method. Two treatments with four responses.

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	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	622	577	587	563	559	547
			$\rho_e = 0.5$	606	588	600	581	551	550
			$\rho_e = 0.9$	629	580	596	563	584	558
		$\sigma = 0.5$	$\rho_e = 0.1$	661	634	654	644	620	601
			$\rho_e = 0.5$	650	614	658	632	630	625
			$\rho_e = 0.9$	669	635	653	631	637	634
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	377	382	368	359	357	349
			$\rho_e = 0.5$	372	394	370	368	379	391
			$\rho_e = 0.9$	360	360	349	341	362	368
		$\sigma = 0.5$	$\rho_e = 0.1$	379	389	372	355	367	370
			$\rho_e = 0.5$	372	382	357	343	371	374
			$\rho_e = 0.9$	359	368	347	353	360	358
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
T	$n = 100$	$DF = 3$	$\rho_e = 0.1$	403	406	398	407	417	416
			$\rho_e = 0.5$	403	404	395	390	403	409
			$\rho_e = 0.9$	393	406	415	407	378	366
		$DF = 8$	$\rho_e = 0.1$	422	426	442	447	448	430
			$\rho_e = 0.5$	421	414	437	448	415	417
			$\rho_e = 0.9$	414	421	421	419	421	429
	$n = 200$	$DF = 3$	$\rho_e = 0.1$	367	375	384	382	355	356
			$\rho_e = 0.5$	352	349	359	341	370	375
			$\rho_e = 0.9$	365	362	367	363	326	311
		$DF = 8$	$\rho_e = 0.1$	359	362	377	380	372	389
			$\rho_e = 0.5$	376	386	363	366	372	375
			$\rho_e = 0.9$	393	381	330	323	357	352

Table 4.5: Frequencies of correct treatment assignments in 1000 test cases by the proposed method. Three treatment groups with two responses.

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	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	707	650	705	671	669	640
			$\rho_e = 0.5$	720	671	682	652	672	635
			$\rho_e = 0.9$	698	660	687	645	648	645
		$\sigma = 0.5$	$\rho_e = 0.1$	380	400	368	369	385	407
			$\rho_e = 0.5$	408	390	374	384	391	379
			$\rho_e = 0.9$	401	408	382	384	386	386
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	806	753	790	718	729	690
			$\rho_e = 0.5$	772	709	793	720	743	685
			$\rho_e = 0.9$	783	737	800	725	738	704
		$\sigma = 0.5$	$\rho_e = 0.1$	438	434	401	413	395	392
			$\rho_e = 0.5$	385	391	393	399	396	395
			$\rho_e = 0.9$	403	413	400	403	381	388
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
T	$n = 100$	$DF = 3$	$\rho_e = 0.1$	350	327	370	362	369	360
			$\rho_e = 0.5$	343	336	372	369	337	336
			$\rho_e = 0.9$	339	343	339	328	349	348
		$DF = 8$	$\rho_e = 0.1$	378	370	359	360	343	351
			$\rho_e = 0.5$	336	333	343	345	338	341
			$\rho_e = 0.9$	346	347	312	309	374	365
	$n = 200$	$DF = 3$	$\rho_e = 0.1$	355	373	355	358	353	355
			$\rho_e = 0.5$	357	363	345	353	337	335
			$\rho_e = 0.9$	332	333	353	357	354	358
		$DF = 8$	$\rho_e = 0.1$	382	385	339	341	373	373
			$\rho_e = 0.5$	350	371	358	357	353	344
			$\rho_e = 0.9$	343	338	368	373	354	344

Table 4.6: Frequencies of correct treatment assignments in 1000 test cases by the proposed method. Three treatment groups with three responses.

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	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	741	682	685	637	648	643
			$\rho_e = 0.5$	701	665	671	627	643	607
			$\rho_e = 0.9$	709	640	703	649	636	610
		$\sigma = 0.5$	$\rho_e = 0.1$	410	411	402	410	373	378
			$\rho_e = 0.5$	375	385	382	376	384	378
			$\rho_e = 0.9$	408	388	376	383	377	379
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	781	716	747	715	745	712
			$\rho_e = 0.5$	737	684	722	690	684	659
			$\rho_e = 0.9$	755	700	734	687	690	653
		$\sigma = 0.5$	$\rho_e = 0.1$	438	425	437	439	408	416
			$\rho_e = 0.5$	410	422	395	401	415	425
			$\rho_e = 0.9$	419	413	394	393	410	397
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
T	$n = 100$	$DF = 3$	$\rho_e = 0.1$	363	341	377	377	322	313
			$\rho_e = 0.5$	339	351	337	333	354	352
			$\rho_e = 0.9$	338	329	360	346	328	316
		$DF = 8$	$\rho_e = 0.1$	341	351	320	316	346	336
			$\rho_e = 0.5$	337	341	348	347	353	349
			$\rho_e = 0.9$	352	342	319	320	363	368
	$n = 200$	$DF = 3$	$\rho_e = 0.1$	365	352	377	379	328	320
			$\rho_e = 0.5$	369	378	341	351	352	334
			$\rho_e = 0.9$	357	359	335	317	354	352
		$DF = 8$	$\rho_e = 0.1$	364	349	336	336	378	383
			$\rho_e = 0.5$	368	375	351	355	340	345
			$\rho_e = 0.9$	337	342	377	368	330	326

Table 4.7: Frequencies of correct treatment assignments in 1000 test cases by the proposed method. Three treatment groups with four responses.

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	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	690	660	656	642	621	593
			$\rho_e = 0.5$	699	658	652	627	612	606
			$\rho_e = 0.9$	675	653	647	601	590	589
		$\sigma = 0.5$	$\rho_e = 0.1$	397	414	368	374	401	401
			$\rho_e = 0.5$	414	426	399	412	401	399
			$\rho_e = 0.9$	407	419	374	391	376	369
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	734	702	731	688	677	664
			$\rho_e = 0.5$	736	686	729	685	706	647
			$\rho_e = 0.9$	742	709	728	681	704	685
		$\sigma = 0.5$	$\rho_e = 0.1$	442	441	440	450	404	411
			$\rho_e = 0.5$	398	391	387	396	408	411
			$\rho_e = 0.9$	420	424	398	398	390	393
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
T	$n = 100$	$DF = 3$	$\rho_e = 0.1$	368	353	336	336	363	357
			$\rho_e = 0.5$	363	353	364	358	329	334
			$\rho_e = 0.9$	330	342	362	353	338	335
		$DF = 8$	$\rho_e = 0.1$	369	371	357	373	342	339
			$\rho_e = 0.5$	362	368	349	350	353	362
			$\rho_e = 0.9$	362	352	335	320	364	357
	$n = 200$	$DF = 3$	$\rho_e = 0.1$	356	365	350	357	345	346
			$\rho_e = 0.5$	370	374	378	377	341	349
			$\rho_e = 0.9$	340	345	344	346	353	344
		$DF = 8$	$\rho_e = 0.1$	366	372	362	363	380	364
			$\rho_e = 0.5$	351	367	348	345	333	319
			$\rho_e = 0.9$	354	353	364	364	358	358

Table 4.8: Frequencies of correct treatment assignments in 1000 test cases by the proposed method, for three treatment groups with three responses, using weights, $\omega_1 = 0.5$, $\omega_2 = 0.3$, and $\omega_3 = 0.2$, for responses 1, 2, and 3, respectively.

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	$n = 100$	$\sigma = 0.1$	$\rho_e = 0.1$	701	678	690	655	634	611
			$\rho_e = 0.5$	735	680	683	646	613	584
			$\rho_e = 0.9$	704	667	665	624	608	597
		$\sigma = 0.5$	$\rho_e = 0.1$	399	401	401	413	380	386
			$\rho_e = 0.5$	372	371	391	386	377	371
			$\rho_e = 0.9$	392	381	396	382	393	394
	$n = 200$	$\sigma = 0.1$	$\rho_e = 0.1$	796	765	738	709	726	712
			$\rho_e = 0.5$	771	737	748	710	697	670
			$\rho_e = 0.9$	778	737	753	683	698	670
		$\sigma = 0.5$	$\rho_e = 0.1$	404	407	430	439	412	417
			$\rho_e = 0.5$	405	414	411	402	415	420
			$\rho_e = 0.9$	400	409	394	384	396	388
				$\rho_c = 0.1$		$\rho_c = 0.5$		$\rho_c = 0.9$	
T	$n = 100$	$DF = 3$	$\rho_e = 0.1$	368	348	358	358	332	335
			$\rho_e = 0.5$	324	342	333	337	344	326
			$\rho_e = 0.9$	326	312	358	357	321	326
		$DF = 8$	$\rho_e = 0.1$	355	352	325	340	366	358
			$\rho_e = 0.5$	370	370	317	325	359	362
			$\rho_e = 0.9$	339	334	337	324	355	354
	$n = 200$	$DF = 3$	$\rho_e = 0.1$	368	367	349	359	344	339
			$\rho_e = 0.5$	353	370	349	336	362	350
			$\rho_e = 0.9$	365	354	349	326	343	339
		$DF = 8$	$\rho_e = 0.1$	335	335	359	367	357	354
			$\rho_e = 0.5$	360	374	365	373	363	360
			$\rho_e = 0.9$	347	346	349	362	334	326

Table 4.9: Frequencies of correct treatment assignments in 1000 test cases by the proposed method, for three treatment groups 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.

Orginal Assignment	Proposed Assignment							
	Arm-0		Arm-1		Arm-2		Arm-3	
	Prob. of dominance	Smooth means	Prob. of dominance	Smooth means	Prob. of dominance	Smooth means	Prob. of dominance	Smooth means
Arm-0	6	18	201	159	60	84	65	71
Arm-1	7	20	178	162	60	67	74	70
Arm-2	13	25	187	158	63	66	61	75
Arm-3	11	26	207	189	76	76	67	70
Total	37	89	773	668	259	293	267	286

Table 4.10: Treatment assignment summary for ACTG-175 clinical trial data, by the proposed method selecting both CD4 and CD8 counts as clinical response using weights $\omega_{CD4} = 0.6$, $\omega_{CD8} = 0.4$ for CD4 and CD8 counts, respectively.

Orginal Assignment	Proposed Assignment							
	Arm-0		Arm-1		Arm-2		Arm-3	
	Prob. of dominance	Smooth means	Prob. of dominance	Smooth means	Prob. of dominance	Smooth means	Prob. of dominance	Smooth means
Arm-0	2	27	211	152	70	93	49	60
Arm-1	2	20	193	161	77	76	47	62
Arm-2	5	31	201	151	73	82	45	60
Arm-3	5	38	223	176	88	84	45	63
Total	14	116	828	640	308	335	186	245

Table 4.11: Treatment assignment summary for ACTG-175 clinical trial data, by selecting CD4 counts as the clinical response.

Orginal Assignment	Proposed Assignment							
	Arm-0		Arm-1		Arm-2		Arm-3	
	Prob. of dominance	Smooth means	Prob. of dominance	Smooth means	Prob. of dominance	Smooth means	Prob. of dominance	Smooth means
Arm-0	89	89	125	93	59	42	59	108
Arm-1	76	85	94	96	75	42	74	96
Arm-2	98	78	94	106	73	33	58	107
Arm-3	104	97	108	105	68	37	81	122
Total	367	349	421	400	275	154	272	433

Table 4.12: Treatment assignment summary for ACTG-175 clinical trial data, by selecting CD8 counts as the clinical response.

CHAPTER 5

FLEXIBLE SEMI-PARAMETRIC REGRESSION OF STATE OCCUPATIONAL PROBABILITIES IN A MULTISTATE MODEL WITH RIGHT-CENSORED DATA

5.1 The Proposed Methodology

5.1.1 Data Structure and Notations

In this section we briefly describe the outline of a multistate model and notations used. Suppose that starting from an initial state 0, a set of n individuals move along an acyclic interconnected networked system of J number of states, 0, 1, ..., $(J-1)$ in a time continuous multistate model, where an individual enters the j th state at most once. Such a model can be graphically represented as a directed graph using arrows to represent transitions from one node to another. For example, Figure 5.6 presents the progressive illness death model that contains three states.

Consider the i th, $i = 1, ..., n$, individual in the multistate system. Let $S_i(t)$ be the state occupied by the individual at time t. Suppose, the time taken to reach the final state $(J-1)$ to be T_i^*, which can be subjected to right-censoring. Let C_i be the right-censoring time and $T_i = \min\{T_i^*, C_i\}$ is the observed final transition time. Define, $\delta_i = I(C_i \geq T_i^*)$ indicates observing the final transition. For any two states j and $j^{'}$, in case the ith individual moves from j to $j^{'}$, we define $U_{i,jj'}$ to be the transition time from j to $j^{'}$, which is considered to be ∞ if the transition is not made at all. Let $K_i(t)$ be the conditional survival function of the censoring distribution

for the individual. Suppose $X_i = (X_{i1}, X_{i2}, ..., X_{ip})$ is the individual's p dimensional covariate vector that contains subject specific baseline information. We assume full data consists of independent and identically distributed copies of $\{S(t), X(t), \delta(t)\}$ given by $\{S_i(t), X_i(t), \delta_i(t)\}$, corresponding to $1 \leq i \leq n$, n individuals, which are observed at points t in a contentious time frame ($t \in [0, \infty)$). Our goal is to estimate the conditional state occupational probabilities given baseline covariate vector (x): $p_j(t) = Pr\{S(t) = j | X = x\}$, $j = 0, 1, ..., J-1$, based on the observed data.

5.1.2 Binary Choice Single Index Model for the Right-censored Data

Use of data weighting schemes for the purpose of bias reduction are well known in statistical literature. One way to handle censored observation in the context of regression is to introduce a re-weighing scheme to the original estimator developed for complete data, in a way that the bias caused due to censoring fades away asymptotically. This idea was first introduced by Koul, Susarla, and Van Ryzin (1981) for the randomly right censored data in linear regression.

The main body of this work is based on the IPCW re-weighted Klein and Spady (1993) binary choice single index model, which is used to estimate conditional transitions and risk processes to formulate a conditional transition hazard matrix given a covariate value at a specific time point. Generally speaking, conditional risk processes estimated for a multistate model provide natural approximations for the conditional state occupational probabilities. However such approximates could have a noisy form in finite samples. Thus, we consider obtaining fairly stable estimates using a product limit function of transitions hazards, which we will discuss in a sequel.

Let $N_{jj'}$ be the counting process for transitions from state j to j', with jumps at t corresponding to $\Delta N_{jj'}(t) = \sum_{i=1}^{n} I(S_i(t-) = j, S_i(t) = j') = \sum_{i=1}^{n} \Delta N_{i,jj'}(t)$

and let $Y_j(t) = \sum_{i=1}^{n} I(S_i(t-) = j)$ be the risk process of individuals occupying state j just prior to t. These two process may not be completely observable due to the right-censoring.

We assume the means of conditional at-risk process in jth state $\mu_{i,j}^{Y}(t|X)$ and event process between states j to j' $\mu_{i,jj'}^{N}(t|X)$ of an individual i at t are in following forms,

$$\mu_{i,j}^{Y}(t|X) = E(Y_{i,j}(t)|X_i) = f_{t,j}(\beta_{t,j}^{T} X_i), \tag{5.1}$$

$$\mu_{i,jj'}^{N}(t|X) = E(N_{i,jj'}(t)|X_i) = f_{t,jj'}(\beta_{t,jj'}^{T} X_i), \tag{5.2}$$

where $f_{t,j}$, $f_{t,jj'}$ are unknown smooth functions and $\beta_{t,j}$, $\beta_{t,jj'}$ are p dimensional unknown parameter vectors. Here we present IPCW modified versions of the Kevin and Spardy (1993) binary choice single index model to estimate $\mu_{u,j}^{Y}(t|X = x)$ and $\mu_{u,jj'}^{N}(t|X = x)$, for a hypothetical individual (u) with a specified covariate vector $X = x$. Since these functions are estimated in the same fashion, we describe the proposed algorithm in terms of a generic subject specific process denoted by $H_i(t)$.

Suppose we have the independent observations $H_1(t), H_2(t), ..., H_n(t)$ from n individuals on a right-censored stochastic process $H(t)$ with corresponding censoring indicators $\delta_1(t), \delta_2(t), ..., \delta_n(t)$ and respective survival probabilities of censoring $K_1(t), K_2(t), ..., K_n(t)$. Using the triplet $\Omega_i(t) = \{H_i(t), \delta_i(t), K_i(t)\}, i = 1, ..n$, we proceed to estimate $E(H(t)|X = x)$ in the following manner. Suppose,

$$P(H_i(t) = 1|X = x_i) = P_i(t) = F_t(\beta_t^{'} x_i); i = 1, ..., n.$$

The IPCW re-weighted log-likelihood function $L(t)$ at t is given by,

$$L(t) = \sum_{i=1}^{n} \frac{\delta_i(t)}{K_i(t)} \Big\{ H_i(t) ln\big[P_i(t)\big] + \big[1 - H_i(t)\big] ln\big[1 - P_i(t)\big] \Big\},$$

where $K_i(t) = \prod_{s \geq t}[1 - \lambda_C(t|\bar{Z}_i(s))ds]$ and $\bar{Z}_i(t)$ is a generalized covariate defined for the ith individual, which we will explain in detail in the sequel, when estimating $K_i(t)$. Generally speaking $K_i(t)$ does not have the survival function interpretation,

unless $\bar{Z}_i(t)$ is formed with non-time-varying covariates. We will later describe a flexible model to estimate $K_i(t)$.

Equivalently, we represent $L(t)$ as,

$$L(t) = \sum_{i=1}^{n} \frac{\delta_i(t)}{K_i(t)} \Big\{ H_i(t) ln\big[F_t(\beta_t' x_i)\big] + \big[1 - H_i(t)\big] ln\big[1 - F_t(\beta_t' x_i)\big] \Big\}.$$

For the case of completely observed data, Klein and Spady (1993) introduced a semi-parametric likelihood, by approximating $F_t(\beta' x_i)$ using a non parametric estimator, which is similar to the leave-one-out estimator described by Ichimura, Hall, and Hardle (1993). Thus, for the right-censored data, we obtain a similar estimator by introducing IPCW criteria,

$$\hat{F}_{t_{-i}}(\beta_t' x_i) = \frac{\sum_{l \neq i}^{n} \phi\big(\frac{\beta_t' x_i - \beta_t' x_l}{h}\big) \frac{H_l(t)\delta_l(t)}{K_l(t)}}{\sum_{l \neq i}^{n} \phi\big(\frac{\beta_t' x_i - \beta_t' x_l}{h}\big)},$$

where $\phi(.)$ is kernel function with $\phi(.) \geq 0$ and $\int \phi(u)du = 1$, and h is a smoothing parameter. This leads to find a quasi likelihood function given by,

$$L_q(t) = \sum_{i=1}^{n} \frac{\delta_i(t)}{K_i(t)} \Big\{ H_i(t) ln\big[\hat{F}_{t_{-i}}(\beta_t' x_i)\big] + \big[1 - H_i(t)\big] ln\big[1 - \hat{F}_{t_{-i}}(\beta_t' x_i)\big] \Big\} \hat{\tau}_i,$$

where $\hat{\tau}_i$ is a trimming sequence that is introduced by Klein and Spady (1993) for feasible likelihood criterion, which down-weight observations for which the corresponding densities are small. The solution of β_t is the maximizer of the $L_q(t)$. For the finite sample problem, we obtain estimates $(\hat{\beta}_t, h_0)$ of β_t, h by maximizing the quasi log-likelihood function $(L_q(t))$ with respect to both β_t and h simultaneously. Finally, $F_t(\beta_t' x)$ for $X = x$, is given by,

$$\hat{F}_t(\hat{\beta}_t' x) = \frac{\sum_{i=1}^{n} \phi\big(\frac{\hat{\beta}_t' x - \hat{\beta}_t' x_i}{h_0}\big) \frac{H_i(t)\delta_i(t)}{K_i(t)}}{\sum_{i=1}^{n} \phi\big(\frac{\hat{\beta}_t' x - \hat{\beta}_t' x_i}{h_0}\big)}.$$

In the proposed method, we estimate conditional means of at-risk process and event process between all possible pairs of nodes in the multistate model at t given a covariate $X = x$, using the above procedure. Thus, we find $\hat{\mu}_j^Y(t|X = x)$ and

$\hat{\mu}^{N}_{jj'}(t|X = x)$ by fixing $\{H_i(t), K_i(t), \delta_i(t)\}$ by their respective processes given by $\{Y_{i,j}(t), \hat{K}_i(t^{'}-), \delta_i(t^{'})\}$ and $\{N_{i,jj'}(t), \hat{K}_i(t^{'}-), \delta_i(t^{'})\}$, where $t^{'} = \min\{t, \min_{j' \neq j}\{U_{i,jj'}\}\}$, and $\delta_i(a) = I(C_i \geq a)$.

$$\hat{E}(Y_j(t)|X = x) = n\hat{\mu}^{Y}_{u,j}(t|X = x)$$

$$\hat{E}(N_{jj'}(t)|X = x) = n\hat{\mu}^{N}_{u,jj'}(t|X = x)$$

In the estimation process, one can conveniently choose elements of t as the union of all event times of the multistate model. Since state-to-state conditional transitions are supposed to be non-decreasing function of time, estimated means of the conditional transitions are monotonized via isotonic regression with the generalized pooled adjacent violators algorithm (Barlow et al., 1972; Leeuw, Hornik, and Mair, 2009).

5.1.3 Conditional Transition Hazard Rates and State Occupation Probabilities

In an uncensored experiment, the conditional hazards of transitions from state j to $j^{'}$ $(j \neq j')$ given a specific value of covariate vector $X = x$ is given by,

$$\alpha_{jj'}(t|X = x) = \lim_{dt \to 0} Pr\big\{S(s) = j^{'}, \text{for some } s \in \big[t, t + dt\big)|S(t-) = j, X = x\big\}/dt,$$

where $S(t-)$ is the state occupied just before time t. Note that this conditional transition hazard only condition on the current state for a given $X = x$. Hence this also known as the partially conditional transition hazards (Pepe and Cai, 1993).

The cumulative (integrated) conditional state-to-state transition hazard matrix $(A(t|X = x))$ for the multistate model can be obtained as,

$$A_{jj'}(t|X = x) = \begin{cases} \int_0^t \alpha_{jj'}(u|x)du, & \text{if } j \neq j^{'} \\ -\sum_{j \neq j'} A_{jj'}(t|X = x), & \text{otherwise.} \end{cases}$$

Now we obtain an estimator for $A_{jj'}(t|X=x)$ as described in Datta and Satten (2002), which is given by,

$$\hat{A}_{jj'}(t|X=x) = \begin{cases} \int_0^t d\hat{E}(N_{jj'}(u|X=x))/\hat{E}(Y_j(u|X=x)), & \text{if } j \neq j' \\ -\sum_{j\neq j'} \hat{A}_{jj'}(t|X=x), & \text{otherwise.} \end{cases}$$

Note that, components of the $\hat{A}(t|X=x)$ have the Nelson-Aalen form. Thus, $\hat{A}(t|X=x)$ is referred as the generalized Nelson-Aalen estimator for a multistate model. The process given by $d\hat{E}(N_{jj'}(t|X=x))$ can be obtained from the corresponding jumps of estimated conditional state-to-state transitions between $[t-,t)$. In this calculation we interpret the division by zero to be zero. The estimator of state occupation probabilities follows from the above results via the product limit (integral). This estimator reduce to the Aalen-Johansen estimator (Aalen and Johansen, 1978) under the independent censoring and it's valid even when the multistate model is not hold the Markovian property (Datta and Satten, 2001; 2002). The state conditional occupation probabilities on a given value of $X=x$, $p_j(t|X=x) = Pr(S(t)=j|X=x)$, is given by,

$$\hat{p}_j(t|X=x) = \sum_{k=0}^{J-1} \frac{\hat{Y}_k(0+|X=x)}{n} \hat{p}_{kj}(0,t|X=x)$$

where $\hat{p}_{kj}(0,t|X=x)$ is the (j,k)-th element of the matrix $\hat{P}(s,t|X=x) = \prod_{(s,t]}(I + d\hat{A}(u|X=x))$.

5.1.4 Censoring Hazards and Estimation of the Weights $K_i(t)$

Let $Z_i(t)$ is a generalized covariate defined for individual i, $1 \leq i \leq n$, at time t, which may contains both baseline and additional covariates (could be time varying) than the covariates of primary interest (X), that are affect the censoring hazards. For example, current state occupation indicator at the given time t may affect to the censoring hazard in addition to the baseline covariates. Suppose $\bar{Z}_i(t) =$

$\sigma\{Z_i(s) : 0 \leq s < t\}$ is the observed covariate history prior to t. We assume that the censoring mechanism satisfies, $\lambda_C^i(\bar{Z}_i(t), S_i(.)) = \lambda_C^i(t|\bar{Z}_i(t))$, where $\lambda_C^i(t|.) = \lim_{dt\to 0} Pr(C_i \in [t, t+dt)|T_i \geq t, .)/dt$. We use the Aalen's nonparametric additive model (Aalen, 1980 and 1989) to calculate IPCW weights, which provides a flexible structure to estimate the censoring hazards by allowing covariates to be varied over the time. The censoring hazard of ith individual at time t is given by the following the linear form,

$$\lambda_C^i(t|\bar{Z}_i(t)) = \sum_{k=0}^{m} \beta_k(t) W_{ik}(t),$$

where, $W_{i0}(t) \equiv 1$ and $W_{ik}(t) = f_k(\bar{Z}_i(t)), k = 1, ..., m$, are possibly time-dependent function of the past history of the covariate process for subject i. $\beta_k(t)$ are unknown regression functions that measure the effect of corresponding covariate function on the censoring hazard. Define $W_i(t) = (W_{i1}(t), .., W_{ip}(t))$. Then the Aalen's estimator of cumulative censoring hazard for the ith individual is given by,

$$\Lambda_C^i(t|\bar{Z}_i(t)) = \int_0^t \hat{\lambda}_C^i(u|\bar{Z}_i(t))du = \sum_{j=1}^{n} I(T_j \leq t)(1-\delta_j) W_i(T_j) R^{-1}(T_j) W_j(T_j),$$

where, $R(t) = \sum_{i=1}^{n} I(T_i \geq t) W_i(t) W_i^{'}(t)$. The estimated IPCW weight for ith individual can be expressed as, $\hat{K}_i(t) = exp(-\hat{\Lambda}_C^i(t|\bar{Z}_i(t)))$, where, $\Lambda_C^i(t|\bar{Z}_i(t)) = \int_0^t \lambda_C^i(u|\bar{Z}_i(t))du$.

5.2 Simulation Study

5.2.1 Study Design

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

We conducted the simulation study using the progressive illness death model given in Figure 5.6. In this model, an individual starting from the initial healthy

state (state-0) at time 0, moves to the absorbing state (state-2) denoted as death, by following either of two possible paths, where the path is controlled by an independent Bernoulli random variable. An individual at state-0 has a 60% chance of moving through the intermediate disease state (state-1) to the final state and 40% chance of reaching the final state directly. We simulated n independent individuals starting from state-0 at time 0, whereas n is chosen from the set $\{100, 500, 1000\}$. For each individual, we then generated a three dimensional covariate vector ($X = (X_1, X_2, X_3)$) that contains subject specific information, where its first component (X_1) is generated from the bernoulli distribution with the probability of 0.5 and the other two components (X_2, X_3) are obtained from multivariate normal distribution with a mean vector of $(0, 0)'$ and a dispersion matrix (D) given by,

$$D = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}.$$

We assumed that individuals' transition times follow Log-Normal distributions, in such a way that the log-mean parameters depend upon their baseline covariate vectors. Likewise, transition times T_{01}, T_{12}, and T_{02} are drawn using log-normal distribution ($lnN(\mu_{jj'}, 0.5^2)$), with log-mean parameters $\mu_{01} = (\beta' X)^2 + 0.1$, $\mu_{12} = (\beta' X)^2$, and $\mu_{02} = (\beta' X)^2 + 0.5$ respectively. We chose β to be a normalized vector given by $\beta = (0.40, 0.79, 0.46)'$. In the uncensored experiment, the time (T^*) required for an individual to move from state 0 to 2 directly or through the intermediate state are T_{02} or $T_{01} + T_{12}$, respectively. However, the observed time is considered to be $T = \min\{T^*, C\}$, when the right-censoring is present. In this simulation study, we considered both random and covariate dependent right-censoring cases with 25% and 50% censoring rates. The random censoring times are generated from the Exponential distributions with scale parameters 3×10^{-2} and 1.5×10^{-1} for 25% and 50% cases, respectively. For the covariate dependent censoring setting, we obtained censoring time using an Exponential distribution that is specified by

an indicator function for a threshold value of a linear combination of covariates, which is given by,

$$C \sim I(\beta_c^{'} X > w) exp(\phi_1) + I(\beta_c^{'} X \leq w) exp(\phi_2).$$

We fixed $w = 0.1$ for $\beta_c = (0.3, 0.2, 0.5)'$ and select pairs (ϕ_1, ϕ_2) from the set of $\{(6 \times 10^{-2}, 1 \times 10^{-2}), (1.8 \times 10^{-1}, 9 \times 10^{-2})\}$, for respective 25% and 50% censoring cases.

We applied the proposed procedure to estimate state occupation probabilities for an arbitrarily selected covariate vector, that was given by $x_0 = (1.0, 0.2, -0.1)$. In this work, we used uniform kernels for the IPCW re-weighted binary choice single index models.

5.2.2 Absolute Error of Estimated State Occupation Probability

We examined the absolute error $(\Delta_{L_1}(t|x_0))$, measured between estimated and "true" state occupying probabilities at 25^{th}, 50^{th}, and 75^{th} quantile points correspond to state reaching (states - 1, 2) and leaving (state - 0) times, for the covariate vector $X = x_0$. $\Delta_{L_1}(t|x_0)$ is defined as, $\Delta_{L_1}(t|x_0) = E|\hat{\theta}(t|x_0) - \theta(t|x_0)|$, where $\theta(x_0)$ is the "true" conditional state occupation probability for $X = x_0$ and $\hat{\theta}(x_0)$ is the corresponding estimator by the proposed method. We approximated $\theta(x_0)$ by the proportions of state occupation using an uncensored experiment with large number (10,000) of instances that are generated for the covariate vector $X = x_0$. In this study, we used Cox's model (Cox, 1972) based alternative procedure as a benchmark estimate, in which the state-to-state transitions hazards in the multi-state model are estimated using Cox's regression model fitted marginally for each transition, where the corresponding baseline hazard function is estimated by Breslow's method (Breslow, 1972). Thus, the conditionally estimated transition hazards of transitions between states j to j' $(j \neq j'; j, j' = 0, 1, 2)$ at a given time t can be obtained as,

$\alpha_{jj'}^{Cox}(t|X) = \lambda_0^{jj'}(t)exp(X^T\gamma)$, where $\lambda_0^{jj'}(t)$ is the baseline hazard of transition from state j to j' at time t, $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ is the three dimensional vector that contains respective regression coefficients of baseline covariates $X = (X_1, X_2, X_3)$. Next, we obtained the conditional transition hazards matrices based on $\hat{\alpha}_{jj'}^{Cox}(t|X)$ to derive estimates for the state occupation probabilities $\hat{p}_j^{Cox}(t|X)$ at a specific covariate value $X = x_0$, by following the Aalen-Johansen formulation as discussed in Section 5.3.

We calculated $\Delta_{L_1}(t|x)$ via averaging 500 Monte-Carlo repetitions. Table 5.1 presents the resulted $\Delta_{L_1}(t|x)$ for selected settings under random and covariate dependent censoring cases. Demonstrating reasonable performance, results for the proposed method show a clear decline in the $\Delta_{L_1}^{New}(t|x)$, as the sample size is increased for all types of censoring and rates. As to be expected, $\Delta_{L_1}^{New}(t|x)$ is higher for scenarios with a larger censoring rate. $\Delta_{L_1}^{New}(t|x)$ at 50^{th} and 75^{th} quantile times are relatively larger than corresponding values at 25^{th} quantiles.

Suppose $\Delta_{L_1}^{New}(t|x) \sim n^{-\kappa}$, and $log\{\Delta_{L_1}^{New}(t|x)\} \sim -\kappa log(n)$. Complying with this relationship, we observed a linear trend between $log(n)$ and $log\{\Delta_{L_1}^{New}(t|x)\}$ in every case. For example, Figure 5.6 shows linear relationships in plots of $log(n)$ vs $log\{\Delta_{L_1}^{New}(t|x)\}$ that are developed for the random censoring at a rate of 25%. This infers that the proposed estimator is reaching the true conditional state occupying probability, in an asymptomatic fashion. Clearly, throughout the whole experiment $\Delta_{L_1}^{Cox}(t|x)$ remained almost unchanged with sample sizes. This result shows considerably large L_1 errors, especially at 50% and 75% quantile points for the Cox approach compared to the proposed method, which suggests the Cox approach is incapable of providing reliable state occupation probability estimates under the selected transition model functions.

5.2.3 Coverage of Confidence Interval Developed for Estimated State Occupation Probability

Next, we examined the empirical coverage probability of 95% bootstrap based confidence bands developed for the estimated state occupation probability.

Let $\hat{\theta}_h$ be the estimator of θ, which is calculated using a sample of size n with a bandwidth (h) that results from a cross-validation process. Suppose $\hat{\theta}^*$ is the respective bootstrap estimator which results from a bootstrap sample of size n that obtained by re-sampling data with a replacement. Li and Datta (2001) described the distributional relationship given by $D(\hat{\theta}_{\tilde{h}} - \theta) \approx D(\hat{\theta}^* - \hat{\theta}_{\tilde{g}})$, where $\hat{\theta}_g$ is the corresponding over-smoothed estimate that involves a larger smoothing parameter g $(g > h)$, satisfies $\lim_{n\to\infty} \frac{h}{g} = 0$. Denote, $\upsilon^{*_b} = \hat{\theta}^{*_b} - \hat{\theta}_g$, $b = 1, ..., B$, where B is the number of bootstrap samples drawn from the given data. Suppose $(Q_{\frac{\alpha}{2}}, Q_{1-\frac{\alpha}{2}})$ are the respective $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ percentile points for the data given by $\upsilon^{*_1}, .., \upsilon^{*_B}$. Thus, $(\hat{\theta}_h - \theta)$ lies in the interval of $[Q_{\alpha/2}, Q_{1-\alpha/2}]$ with a probability of $(1-\alpha)$, which yields the $(1-\alpha)\%$ confidence interval for θ as $[\hat{\theta}_h - Q_{1-\frac{\alpha}{2}}, \hat{\theta}_h - Q_{\frac{\alpha}{2}}]$. Suppose $\hat{\theta}_{\tilde{h}}$ is the conditional state occupation probability estimated by the proposed method using a set of bandwidths $\tilde{h} = (h_1, ., h_m, ., h_M)$, $M \in \mathbb{Z}^+$. The over-smoothed estimator $\hat{\theta}_{\tilde{g}}$ can be determined as $\tilde{g} = (g_1, ., g_m, ., g_M)$, such that $g_m > h_m$, $1 \leq m \leq M$. For $0 < \varphi < 1$; one can select g_m as,

$$g_m = \begin{cases} h_m^{\varphi}, & \text{if } h < 1 \\ h_m^{1/\varphi}, & \text{otherwise.} \end{cases}$$

In this simulation, we first calculated $\hat{\theta}_{\tilde{h}}$, $\hat{\theta}_{\tilde{g}}$ from a sample of n individuals. Then we drew n individuals from the original data with replacement to obtain $\hat{\theta}^{b^*}_{\tilde{h}_b}$, $b = 1, .., B$ for B number of bootstrap samples. We chose φ to be 0.9 and B to be 100. For a randomly generated dataset, we determined the indicator of $I\big[\theta \in [\hat{\theta}_{\tilde{h}} - Q_{1-\frac{\alpha}{2}}, \hat{\theta}_{\tilde{h}} - Q_{\frac{\alpha}{2}}]\big]$, which shows the presence of empirically approximated "true"

state occupation probability is inside the constructed confidence interval. Finally, the coverage probability is determined by averaging indicators of 500 Monte-Carlo repetitions. For simplicity, we focused on coverages at the 25^{th}, 50^{th}, and 75^{th} quantile time points corresponding to state reaching (states - 1, 2) and leaving (state - 0) times, described earlier.

Due to the computational burden, this study was limited to random censoring setting with sample sizes $n = 100$ and $n = 500$ cases. Resulted coverages that are summarized in Table 5.2 show reasonable coverages close to the nominal level of 95% almost in every case, which seemingly suggest that the proposed method potentially holds a reliable precision.

5.2.4 Power and Size of Regression Parameters

In this section, we study the power and size properties of regression parameters in the IPCW re-weighted binary choice SIM model. For this purpose, we focused on the parameters estimated for risk processes, which are natural approximations for state occupation probabilities in a multi-state model. Following the normality property of the binary choice SIM parameter estimates described by Klein and Spady (1993), we evaluated the hypothesis represented by $H_0 : \beta_p(t, s) = 0$ vs $H_1 : \beta_p(t, s) \neq 0$, $p = 1, .., P$, using the bootstrap method based standard error. For illustrative purposes, this study is conducted for the 2nd component of β parameter vector; β_2, using the three state progressive illness death model with the transition settings as described in section 3.1, but ranging the value of β_2 in a sequence from 0 to 0.95, under the constraint that satisfies $|\beta^{'}| = 1$. This experiment is initiated by estimating $\hat{\beta}$ from a sample of n individuals at a fixed time. After that, we drew n individuals from the original data with replacement to obtain the corresponding bootstrap estimates: $\hat{\beta}^{*b}$, $b = 1, .., B$ for B number of bootstrap samples. For each

fixed value of β_2 the corresponding estimate $\hat{\beta}_2$ is tested for rejecting the above null hypothesis. Likewise the power is computed by the average rejections in 500 Mote-Carlo simulations, using 100 bootstrap per each. This power study is conducted under random censoring settings with 25% rate for $n = 100$ and $n = 500$ cases.

We also compare our results with a method based on pseudo-values described by Anderson and Klein (2007). In this approach, pseudo-values of the state occupation probabilities are calculated using the Jackknife method which yields $\hat{p}_i^{ps}(t) = n \cdot \hat{p}(t) - (n-1) \cdot \hat{p}^{-i}(t)$, $i = 1, ..., n$, where $\hat{p}^{-i}(t)$ is the Aalen-Johansen (Aalen and Johansen, 1978) state occupation probability estimate obtained from a sample of size $n - 1$ by eliminating the ith individual from the data, and $\hat{p}(t)$ is the corresponding estimate calculated using the whole sample. For the ith individual, the most probable state occupation at a fixed time is determined using the observed pseudo-values, which allows one to estimate the parametric Logistic regression model that relates state occupational indicators with the baseline covariates. In such a way, we conducted the power study for the 2nd parametric component; β_2^{ps} in the above Logistic model.

In Figures 5.6 and 5.6 we present powers and sizes evaluated at 25^{th}, 50^{th}, and 75^{th} quantile time points for a proposed model and pseudo-value approach, respectively. Examining the properties of power functions derived for the IPCW re-weighted binary choice SIM model, we observed a fair agreement between the observed size with the nominal value of 0.05 for most scenarios that covered in the experiment. Clearly, the power is steadily increased as the sample size increases from $n = 100$ to $n = 500$. The overall behavior of power function provides key evidence for the reliability of the new technique, showing its ability to detect crucial covariates upon the state occupation at a given time. The pseudo-value technique has illustrated poor performance in terms of the power, whereas the plots show an irregular behavior in the power function, which can possibly be caused due to

the incapability of handling subjects' nonlinear transition mechanism with baseline covariates by the logistic model that assumes a pre-specified linear model structure.

5.2.5 Robustness of the Proposed Method

The proposed method for estimating conditional state-to-state transitions counts and at-risk processes is established on a strict condition that assumes the underlying functions have the SIM form. However, in practice this may not often hold. Thus, we performed a simulation study to investigate the performance of the proposed method under perturbed SIM functions. Similar to Section 5.2, we examined the absolute error denoted by $\Delta_{L_1}(t)$ at quintile time points, using the same three-state progressive illness-death model described in our previous simulations, but using a new set of functions that have perturbed SIM structure. Likewise, in this experiment, T_{01},T_{12}, and T_{02} transition times are generated from Log-Normal distribution ($lnN(\mu_{jj'}, 0.5^2)$) with log-mean parameters $\mu_{01} = (\beta^T X)^2+(\vartheta X_2)^2+0.1$, $\mu_{12} = (\beta^T X)^2 + (\vartheta X_2)^2$, and $\mu_{02} = (\beta^T X)^2 + (\vartheta X_2)^2 + 0.5$ respectively, where X_2 is the 2nd component in the covariate vector $X = (X_1, X_2, X_3)$. Consequently, this new set of models violate the SIM conditions in temporal processes of multi-state model.

In this study, we chose β and ϑ to be $\beta = (0.40, 0.79, 0.46)'$ and $\vartheta = 0.4$. The robustness of the proposed method was examined under the random censoring cases. Table 5.3 summarizes both $\Delta_{L_1}^{New}(t|x)$ and $\Delta_{L_1}^{Cox}(t|x)$ values that are obtained by averaging 500 Monte-Carlo simulations. Similar to the outcome observed in section 3.2, $\Delta_{L_1}^{New}(t|x)$ clearly declines with the sample size. Demonstrating the consistency of our estimate, further investigation of the results show a linearity between $log(n)$ and $log\{\Delta_{L_1}^{New}(t|x)\}$ values (refer to Figure 5.6), that suggests the proposed method is rather robust under perturbed SIM conditions. As we expected, the Cox approach

demonstrated poor performance with the underlying simulation conditions. The overall result has a high correspondence to the outcome we described in Section 5.2.

5.3 Applications

In this section, we illustrate two applications of the proposed method in real data using bone marrow transplant study (Copelan et al., 1991), and spinal code injury data (Harkema et al., 2012).

5.3.1 Bone Marrow Transplant study

The bone marrow transplant study had been conducted during years 1984 to 1989 for acute leukemia patients in four worldwide centers. This transplant surgery is considered one of the standard therapies for acute leukemia condition. Usually, a subject experiences various clinical conditions during the recovery process, after receiving the transplant from a donor, which can be represented as states of a multi-state model. In this study, 137 acute leukemia subjects underwent bone marrow transplantation. After that, these individuals were followed up to a maximum of 7 years. In a multi-state representation, starting from the primary state of receiving the bone marrow transplant, there are five intermediate states that an individual may reach before he/she reaches the final (absorbing) state, which is relapse or death by leukemia. Five intermediate states are represented by conditions of developing Acute Graft Versus Host disease (GVHD); returning of platelet levels to normal levels; returning of platelet levels to normal levels after developing acute GVHD; Developing acute GVHD after platelet recovery; and developing chronic GVHD. A schematic representation of the multi-state model is given in Figure 5.6. We summarize the transition counts for the study in Table 5.4. For additional details of this study and the dataset, we refer readers to Klein and Moeschberger (1997).

The bone marrow transplant data described here has been analyzed by many authors for various aspects. In this work, our primary goal is to estimate the state occupation probability for a hypothetical individual represented by a specific covariate vector. Among several covariates available, we use ages of patient and donor for our work. We apply the proposed procedure by choosing the covariate vector $x = (28, 28)'$ corresponding to patient's age and donor's age to estimate the state occupation probability $(\hat{p}_j(t|x), j = 0, .., 6)$. After that, we construct 95% bootstrap method based confidence intervals by following the procedure described in the simulation section. Datta and Satten (2001) showed that the censoring depends on the currently occupied state. Hence we consider patient age and the time varying covariate given by current state for calculating IPCW weights.

Figure 5.6 provides sets of plots that are show the estimated conditional state occupation probabilities for an individual of age 28 who receives a bone marrow transplant from a donor of the same age of 28. The plot of state-0 shows probability of staying at this state dramatically declines within a short period, which means the individual possibly moves to another clinical state soon after he/she undergoes surgery. As shown in the plot of state-2, the most probable second state for such an individual must be the 'Platelet Recovery' state, whereas the probability of staying in this state at the very beginning appears to be as high as 0.7. However, the state occupation probability for this state quickly decreases below 0.3 and reaches an almost steady level within first 500 days of the surgery. Plots of states 1 and 3 clearly indicate that the chances of staying in 'Acute GVHD' and one of its subsequent states given by 'Platelet recovery after acute GVHD' are very low for the particular individual. As shown in the plot 4, occupying 'Acute GVHD After Platelet Recovery' at the beginning seems to be close to 0.1, which then declines gradually. The occupation probability at state-5 that is 'Chronic GVHD', quickly increases to approximately 0.25 and remains almost unchanged for a long span. As

to be expected, the chance of moving to the absorbing state (Relapse/Death) increases with time, then it reaches to a constant level after nearly 1000 days after the surgery. Based on 95% bootstrap confidence intervals (with $B = 500$) developed for state occupation probabilities, we observe reasonable precision with our estimates.

We provide another illustration (refer Figure 5.6) of conditional state occupation probabilities for two covariate vectors: $x = (20, 20)'$ and $x = (40, 40)'$. Visual inspection of these plots shows some noticeable differences for two baseline vectors. For example, the overall state occupation probability at 'Platelet recovery after acute GVHD' state for an individual with $x = (20, 20)'$ is larger compared to an individual with $x = (40, 40)'$. Conversely, occupying 'Chronic GVHD' state is lower for $x = (20, 20)'$ than $x = (40, 40)'$ case. To determine if these differences are statistically significant, we developed a test statistic and computed its corresponding p-value using a re-sampling scheme. We calculated the mean absolute difference $D = \int |\hat{\theta}(t|X = x_1) - \hat{\theta}(t|X = x_2)| dE_n(t)$, where $\hat{\theta}(.|X = x)$ is the proposed estimator of state occupation probability given $X = x$ to quantify the overall difference for above covariate vectors. We drew a bootstrap sample of size n by re-sampling labels from the original data $\{1, .., n\}$, using simple random sampling. Then we re-sampled X by drawing samples from the corresponding original data, which was performed independently from the previous re-sampling step. After that we calculated D using the bootstrap sample, denoted D^*. This procedure was repeated for B times to obtain $D_1^*, D_2^*, ..D_B^*$. Then the p-value was computed by $p = \frac{1}{B} \sum_{b=1}^{B} I(D_b^* \geq D)$. In Table 5.5, we present the result based on $B = 500$ bootstrap samples. As shown in the table, the difference at 'Platelet Recovery After Acute GVHD' is shown to be significant at 5% significant level. Other differences, such as in 'Acute GVHD After Platelet Recovery' can be considered as borderline significant.

5.3.2 Spinal Cord Injury Study

Spinal code injury (SCI) data contains information on 296 subjects with incomplete spinal code injury from a national activity-based rehabilitation program (Harkema et al., 2012). The program primarily focused on individuals with clinically incomplete SCI after discharge from inpatient rehabilitation. During the program, patients underwent with sessions of locomotor training, based on a screening assessment at enrollment. The locomotor training program is an activity-based therapy for functions relating to standing and walking. Functional progress of these individual has been thoroughly evaluated periodically in terms of mobility, standing, and stepping, after receiving therapeutic sessions. For instance, this study assessed 6-minute walk and 10-meter walk tests. There are several clinical benchmarks defined based on the walking speed of an individual. For example, 0.44 m/s represents the minimum walking speed associated with the ability to walk in the community, 0.7 m/s separates those who require assistive walking devices from those who do not, and 1.2 m/s approximately defines the speed required to cross a street at a stoplight (van Hedel and Dietz, 2010). Thus, we represent the progress of a SCI subject though the rehabilitation program as states of a multi-state model as follows: (1) nonambulatory, (2) able to walk but slower than 0.44 m/s, (3) able to walk but in between 0.44 to 0.7 m/s, (4) able to walk but in between 0.7 m/s to 1.2 m/s, (5) able to walk faster than 1.2 m/s. A graphical representation of the multi-state model with 5 states is given in Figure 5.9. We also provide a summary of transition counts in Table 5.5. It is important to note that individuals in this study have been entered to the multi-state model from various states, which is one of the primary difference between SCI and bone marrow transplant studies. In a previous anlsysis of this dataset, Lorenz and Datta (2015) estimated an individuals' waiting time to reach state-1 in the model, based on a linear hazards model approach.

Among several subject specific measures given, we used (1) initial speed at enrollment, (2) patient age at enrollment, (3) time from spinal cord injury to enrollment, and (4) lower motor score from the International Standards for Neurological Classification to estimate the state occupation probability for a given individual with a specified baseline covariate vector. We noticed that SCI subjects with severe conditions, such as nonambulatory or patients with slower walking speed than 0.44 m/s very rarely regain a walking ability than 1.2 m/s. Thus, most of these patients have been subjected to right-censoring before they were discharged from the program upon recovery, resulting approximately 82% censoring rate in terms of reaching the final state. We believe the censoring hazards may have an effect by the cumulative number of training sessions the individual received prior to censoring, which can be considered as a time varying covariate. Thus, we use this information in addition to baseline covariates to calculate the IPCW weights. Figure 5.6 presents conditional state occupation probabilities estimated using the proposed method along with 95% bootstrap confidence intervals ($B = 150$), for an individual with a covariate vector $x = (0.08, 38.0, 0.92, 33.0)'$, which corresponds to median baseline covariates of study participants. For this individual, represent by the selected covariate vector, the chance of occupying state-0 at the early period is close to 0.5, which then sharply declines. There is an approximately 0.25 chance of occupying state-1 at the beginning, which relatively increases over time. Occupying at states 2 gradually declines, while 3, and 4 considerably elevates when reaching to 500 days of enrollment. We believe lack of transitions along with high rate of censoring may have caused relatively large confidence bands for our estimates.

5.4 Discussion

In this work we proposed a novel method to estimate the conditional state

occupational probability of a multi-state model, given a covariate, in the presence of right-censoring. The proposed method has broad potential advantages in complex multi-state problems, where the transition mechanism follows a dynamically varying high nonlinearity with baseline covariates. We proposed IPCW re-weighted semi-parametric binary choice SIM model to estimate state-to-state transitions and at-risk processes, which allows one to estimate transition hazards between pairs of states in the multi-state model. The integrated IPCW re-weighting scheme handles the bias caused by censoring, ensuring theoretical properties of the model.

We present a series of simulation studies that investigates properties of the proposed method in finite sample. We primarily focused on the L_1 distance between the estimated and true conditional state occupational probabilities (Δ) given a vector of covariates, at a selected set of time points. Clearly, L_1 distance seems to be decreasing with the sample size for both random and covariate dependent censoring mechanisms, under low and high censoring rates. These results suggest a linear relationship between $log(n)$ and $log(\Delta)$. Thus, the proposed estimator seemingly converges to the true conditional state occupation probability asymptomatically. A study that investigates the coverage of bootstrap confidence intervals demonstrated a reasonable agreement between the empirical coverage and the nominal level. We also demonstrated the performance of the proposed model in examining the covariate effects using a power study, whereas this study infers the poor performance of pseudo-value based parametric method under complex transition mechanisms. Although the proposed method is based on SIM conditions, evidence of our study showed that the proposed estimator is rather robust under the model functions' departure from SIM conditions, signifying the potential capability of the proposed method for handling many real applications. Two applications of the proposed method in real data were illustrated using bone marrow transplantation and SCI data, which contain several possible clinical states.

A possible extension for the current methodology could be generalizing for high dimensional case that contain excessive number of covariates, which may be achievable by imposing a strong dimension reduction criteria. Such an approach may have wide range applications with sophisticated multi-state models under high dimensional covariates, such as genomic biomarkers.

5.5 Tables

State	Quantile Time Point	Sample Size (n)	Δ (10^{-2})									
			No Censoring		Random Censoring				Covariate Dependent Censoring			
					25% Rate		50% Rate		25% Rate		50% Rate	
			Proposed Method	Cox Method	Proposed Method	Cox Method	Proposed Method	Cox Method	Proposed Method	Cox Method	Proposed Method	Cox Method
State-0	25%	100	6.673	8.463	6.605	8.342	6.629	8.277	6.921	8.049	7.472	8.211
		500	3.610	8.503	3.378	8.608	3.549	8.491	3.710	8.249	3.884	7.947
		1000	2.705	8.446	2.578	8.527	2.581	8.448	2.839	8.216	3.652	8.216
	50%	100	7.719	18.358	7.399	18.355	8.120	18.360	7.927	18.210	8.158	17.916
		500	4.569	18.210	4.317	18.681	4.397	18.459	4.543	17.990	4.898	18.105
		1000	3.346	18.135	3.271	18.432	3.357	18.285	3.432	17.802	3.707	17.802
	75%	100	7.189	26.345	8.032	26.281	8.154	26.236	7.631	26.001	7.670	25.447
		500	4.312	25.956	4.441	26.751	4.720	26.397	4.156	25.567	4.485	25.896
		1000	3.222	25.869	3.228	26.389	3.501	26.208	2.954	25.391	3.711	25.391
State-1	25%	100	5.202	5.153	5.521	5.441	5.872	5.377	5.464	5.149	5.902	4.903
		500	2.966	4.970	2.867	5.113	2.871	5.064	2.798	4.936	3.087	5.139
		1000	2.325	4.915	2.247	5.011	2.284	4.981	2.315	4.843	3.677	4.843
	50%	100	6.560	11.149	6.246	11.446	6.332	11.505	6.866	11.269	7.477	11.025
		500	4.189	11.048	3.776	11.235	3.601	11.165	3.580	11.070	4.144	11.262
		1000	2.859	10.977	2.964	11.065	3.079	11.055	3.035	10.807	3.892	10.807
	75%	100	7.484	14.094	7.508	14.637	8.115	14.661	7.934	14.455	8.279	13.916
		500	4.369	14.055	4.448	14.272	5.042	14.146	4.420	13.920	5.108	14.375
		1000	3.055	13.902	3.083	14.050	3.362	14.015	2.993	13.869	4.207	13.869
State-2	25%	100	7.206	7.584	6.398	7.566	6.792	7.310	7.136	7.131	8.475	6.853
		500	3.902	7.186	3.506	7.519	3.718	7.392	3.883	6.954	4.209	7.321
		1000	2.735	7.231	2.920	7.463	3.167	7.369	2.951	6.942	4.375	6.942
	50%	100	7.871	15.720	9.151	16.071	8.930	15.756	8.907	15.074	10.095	15.081
		500	4.777	15.636	4.245	16.217	4.743	16.062	4.698	15.305	5.509	15.165
		1000	3.561	15.610	3.824	16.112	3.830	15.932	3.672	14.894	4.859	14.894
	75%	100	7.229	22.660	7.317	23.761	8.286	23.327	7.611	22.088	8.091	22.218
		500	4.587	23.096	4.326	24.026	4.896	23.722	4.621	22.565	6.082	22.022
		1000	3.101	23.055	3.569	23.834	4.042	23.542	3.355	22.113	5.282	22.113

Table 5.1: L_1 Distances of estimated state conditional occupying probabilities at 25th, 50th, and 75th quantiles of state reaching (states - 1, 2) and leaving (state - 0) times, with 500 Monte-Carlo simulations, for two different censoring mechanisms and various rates, using the proposed method and Cox-Regression approach.

Censoring Mechanism	Size (n)	95% Coverage Probability								
		State-0			State-1			State-2		
		25%	50%	75%	25%1	50%	75%	25%	50%	75%
No Censoring	100	0.93	0.94	0.97	0.91	0.95	0.98	0.95	0.97	0.98
	500	0.91	0.93	0.94	0.91	0.93	0.94	0.94	0.92	0.93
Random Censoring	100	0.95	0.95	0.96	0.94	0.95	0.97	0.96	0.96	0.96
25% Rate	500	0.94	0.93	0.96	0.94	0.91	0.95	0.93	0.92	0.92
Random Censoring	100	0.93	0.96	0.97	0.91	0.95	0.97	0.95	0.96	0.99
50% Rate	500	0.98	0.97	0.97	0.98	0.96	0.95	0.97	0.96	0.91

Table 5.2: The coverage probability of 95% bootstrap based confidence intervals for estimated state conditional occupation probabilities using the proposed method, at 25th, 50th, and 75th quantiles of state reaching (states - 1, 2) and leaving (state - 0) times, under random censoring with 0 to 50% rates, averaging 500 Monte-Carlo's with 100 bootstraps per each simulation.

State	Quantile Time Point	Sample Size (n)	Δ (10^{-2})					
			No Censoring		Random Censoring			
					25% Rate		50% Rate	
			Proposed Method	Cox Method	Proposed Method	Cox Method	Proposed Method	Cox Method
State-0	25%	100	6.516	9.988	6.621	9.791	6.661	9.719
		500	4.032	10.088	3.882	10.046	3.882	10.157
		1000	3.050	10.043	3.082	10.018	3.086	10.102
	50%	100	8.380	20.899	8.052	21.001	8.570	20.898
		500	5.127	20.793	4.994	20.965	5.091	21.166
		1000	3.932	20.754	4.045	20.806	4.274	20.932
	75%	100	9.053	29.593	8.798	29.669	9.367	29.577
		500	5.177	29.318	5.437	29.715	5.920	30.034
		1000	4.065	29.288	4.032	29.505	4.427	29.709
State-1	25%	100	5.062	5.945	5.162	6.114	5.297	6.101
		500	3.151	5.920	2.903	5.974	3.007	6.031
		1000	2.356	5.872	2.415	5.938	2.507	5.984
	50%	100	6.646	12.466	6.844	12.757	6.619	12.752
		500	4.377	12.303	4.230	12.411	4.343	12.461
		1000	3.311	12.253	3.362	12.303	3.469	12.353
	75%	100	8.481	15.046	8.197	15.509	8.627	15.468
		500	4.269	15.009	4.656	15.008	4.912	15.124
		1000	3.382	14.859	3.285	14.894	3.587	14.945
State-2	25%	100	6.842	9.260	6.853	8.851	7.332	8.886
		500	3.943	9.093	4.012	9.279	4.218	9.363
		1000	2.974	9.138	3.052	9.206	3.359	9.289
	50%	100	8.651	18.879	8.868	18.816	9.059	18.997
		500	5.401	18.801	5.377	19.221	5.469	19.367
		1000	4.096	18.842	4.251	19.054	4.626	19.193
	75%	100	8.221	26.701	8.303	27.254	9.422	27.510
		500	5.373	27.222	5.831	27.755	6.594	28.026
		1000	3.902	27.212	4.323	27.575	5.133	27.836

Table 5.3: L_1 Distances of estimated state conditional occupying probabilities at 25th, 50th, and 75th quantiles of state reaching (states - 1, 2) and leaving (state - 0) times, with perturbed SIM models, under random censoring, based on averaging 500 Monte-Carlo simulations, using the proposed method and Cox-Regression approach.

	To						
From	0	1	2	3	4	5	6
0	0	7	117	0	0	1	12
1		0	0	3	0	2	2
2			20	0	19	44	34
3				0	0	1	2
4					2	11	6
5						32	27
6							83

Table 5.4: Matrix showing the state-to-state transition counts for the Bone Marrow Transplant data

State	D (10^{-2})
0: Bone Marrow Transplant	0.533 (0.65)
1: Acute GVHD	0.854 (0.15)
2: Platelet Recovery	3.457 (0.13)
3: Platelet Recovery After Acute GVHD	1.710 (0.03)
4: Acute GVHD After Platelet Recovery	2.152 (0.08)
5: Chronic GVHD	3.506 (0.15)
6: Relapse/Death	1.843 (0.42)

Table 5.5: Absolute mean difference between conditional state occupation probabilities of two cases: $x = (20, 20)'$ and $x = (40, 40)'$, using the proposed method. The corresponding p-values are in parenthesis.

	To				
From	0	1	2	3	4
0	79	47	10	0	0
1		68	38	16	0
2			39	33	7
3				55	27
4					32

Table 5.6: Matrix showing the state-to-state transition counts for the Spinal Code Injury data

5.6 Figures

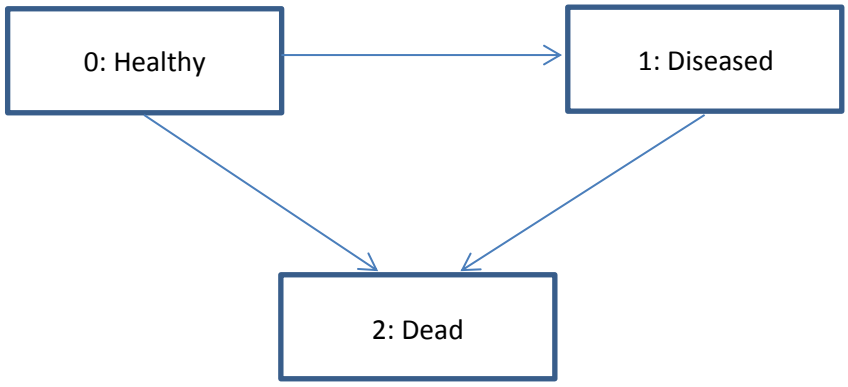

Figure 5.1: Graphical representation of the three state progressive illness-death model

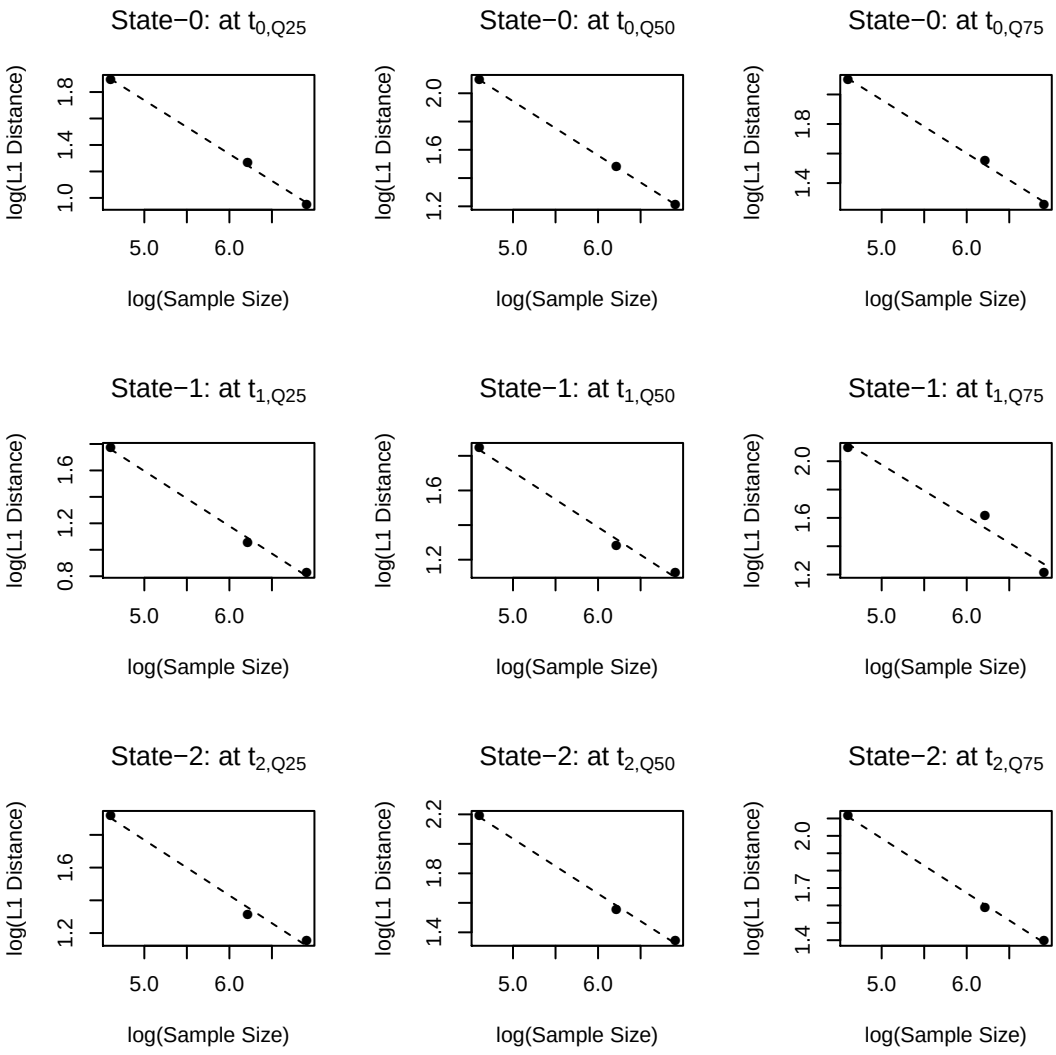

Figure 5.2: Plots of $log(n)$ vs $log(\Delta_{L_1}^{New}(t|x))$ generated at 25th, 50th, and 75th quantiles of state reaching (states - 1, 2) and leaving (state - 0) times, under random censoring with a rate of 50%. Dotted lines represent the corresponding liner regression line fitted to the data.

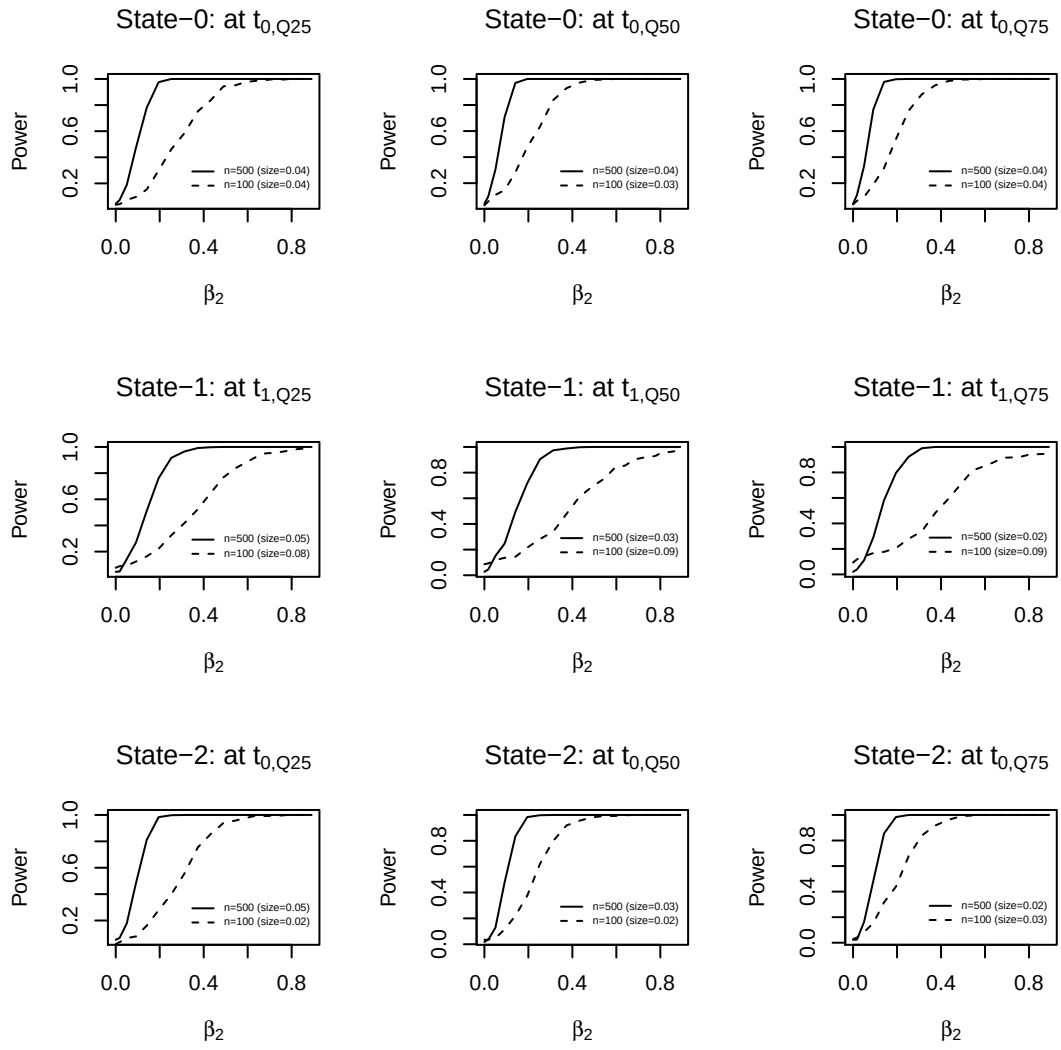

Figure 5.3: Plots showing power and size properties of β_2 parameter in the re-weighted binary choice SIM models, which are estimated for risk processes of the three state progressive illness death model. Plots are generated at 25th, 50th, and 75th quantiles of state reaching (states - 1, 2) and leaving (state - 0) times, under random censoring with 25% rate, for $n = 100$ and $n = 500$ cases.

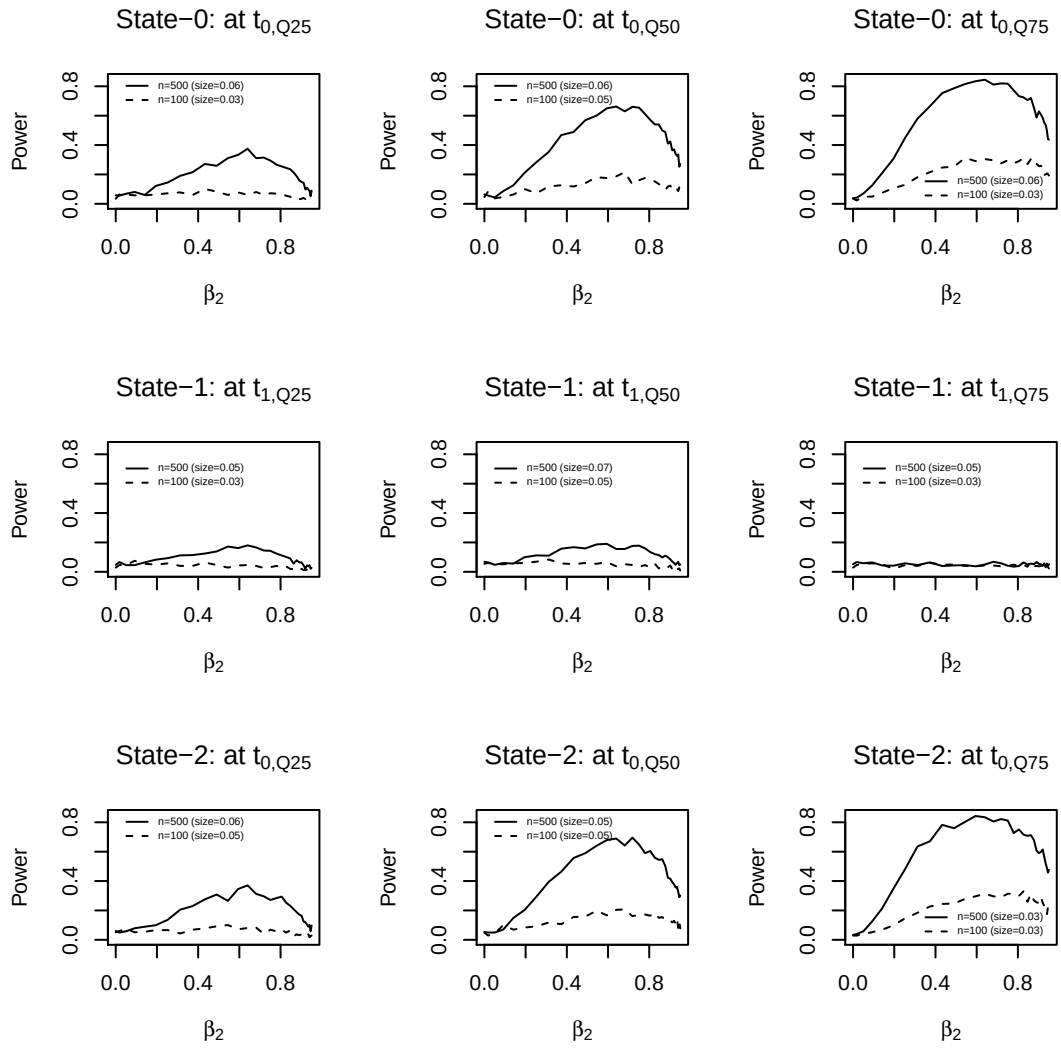

Figure 5.4: Plots showing power and size properties of β_2 parameter in the pseudo-value based Logistic regression models, which are estimated for state occupations in the three state progressive illness death model. Plots are generated at 25th, 50th, and 75th quantiles of state reaching (states - 1, 2) and leaving (state - 0) times, under random censoring with 25% rate, for $n = 100$ and $n = 500$ cases.

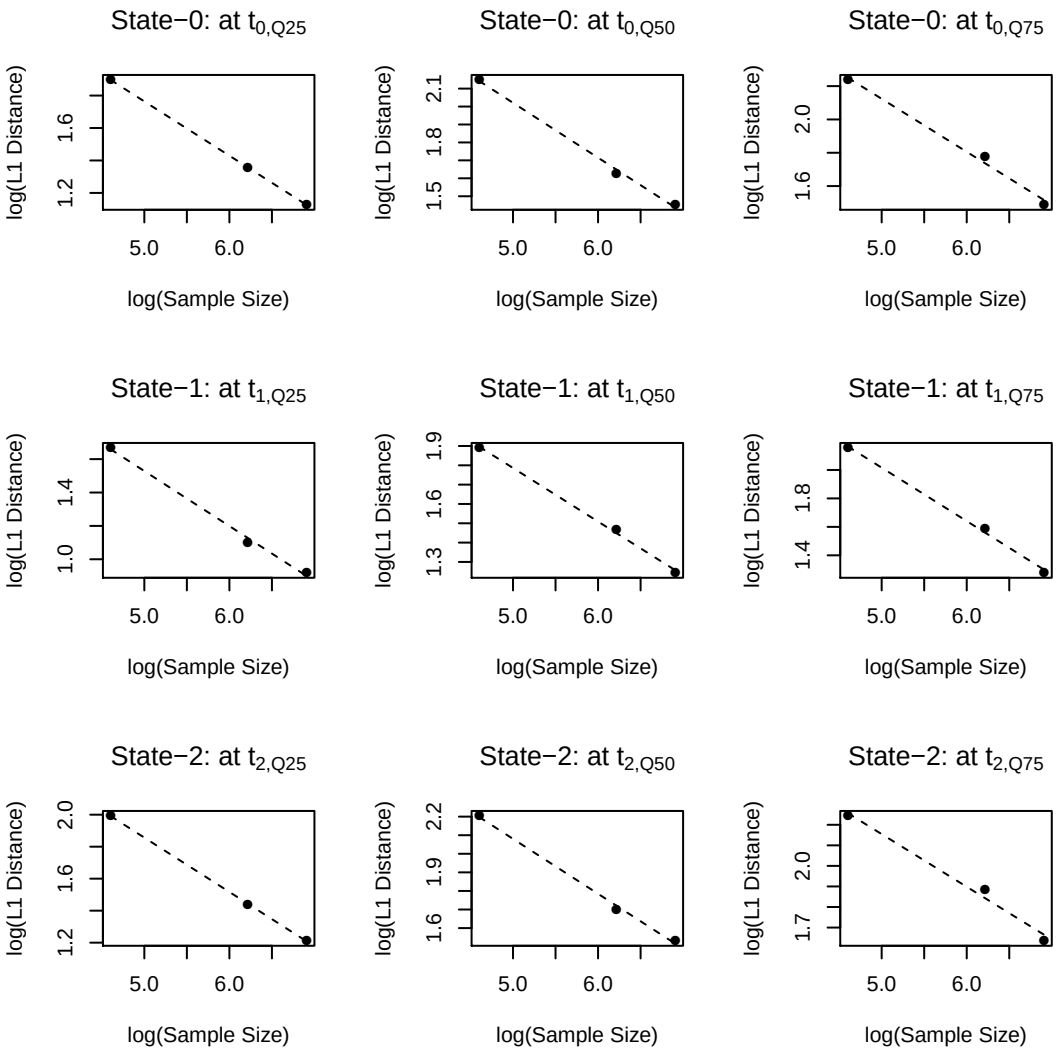

Figure 5.5: Plots of $log(n)$ vs $log(\Delta_{L_1}^{New}(t|x))$ generated at 25th, 50th, and 75th quantiles of state reaching (states - 1, 2) and leaving (state - 0) times, under random censoring with a rate of 50% with perturbed SIM models. Dotted lines represent the corresponding liner regression line fitted to the data.

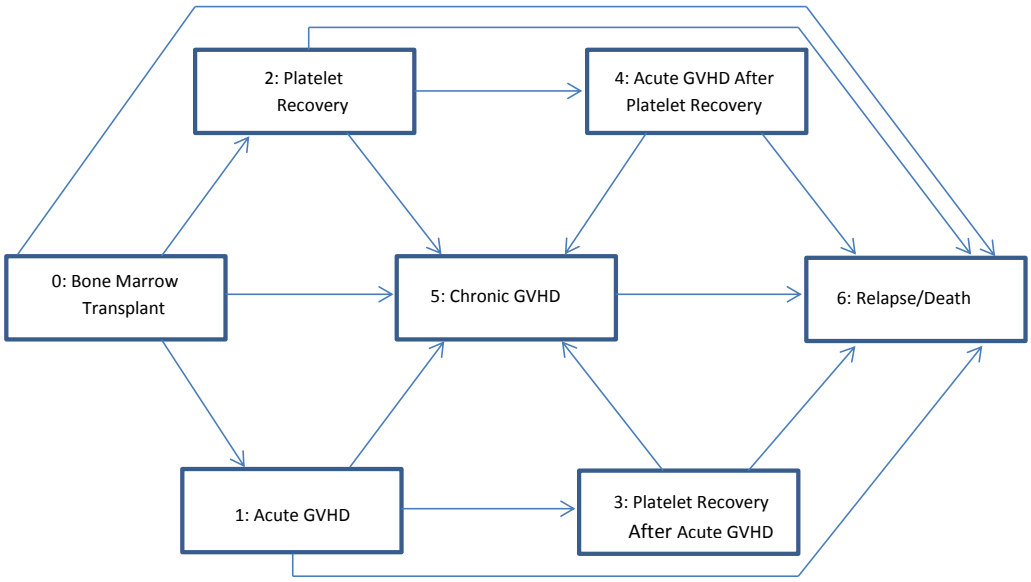

Figure 5.6: Graphical representation of Bone Marrow transplant multi-state model

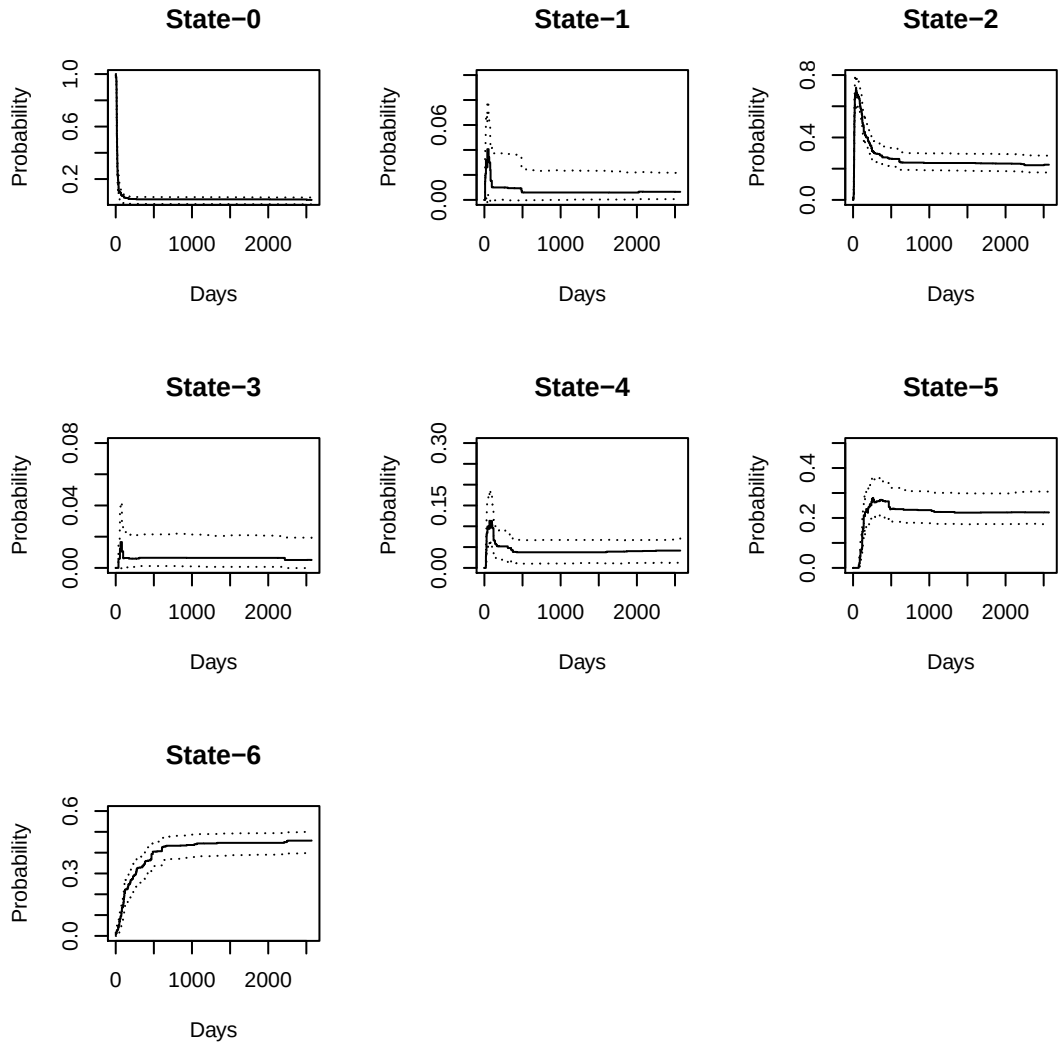

Figure 5.7: Plots of estimated conditional state occupational probabilities of 7 clinical states in the bone marrow transplant data, by the proposed method for a covariate vector of $x = (28, 28)'$, along with 95% bootstrap based confidence intervals (represented by the dotted lines).

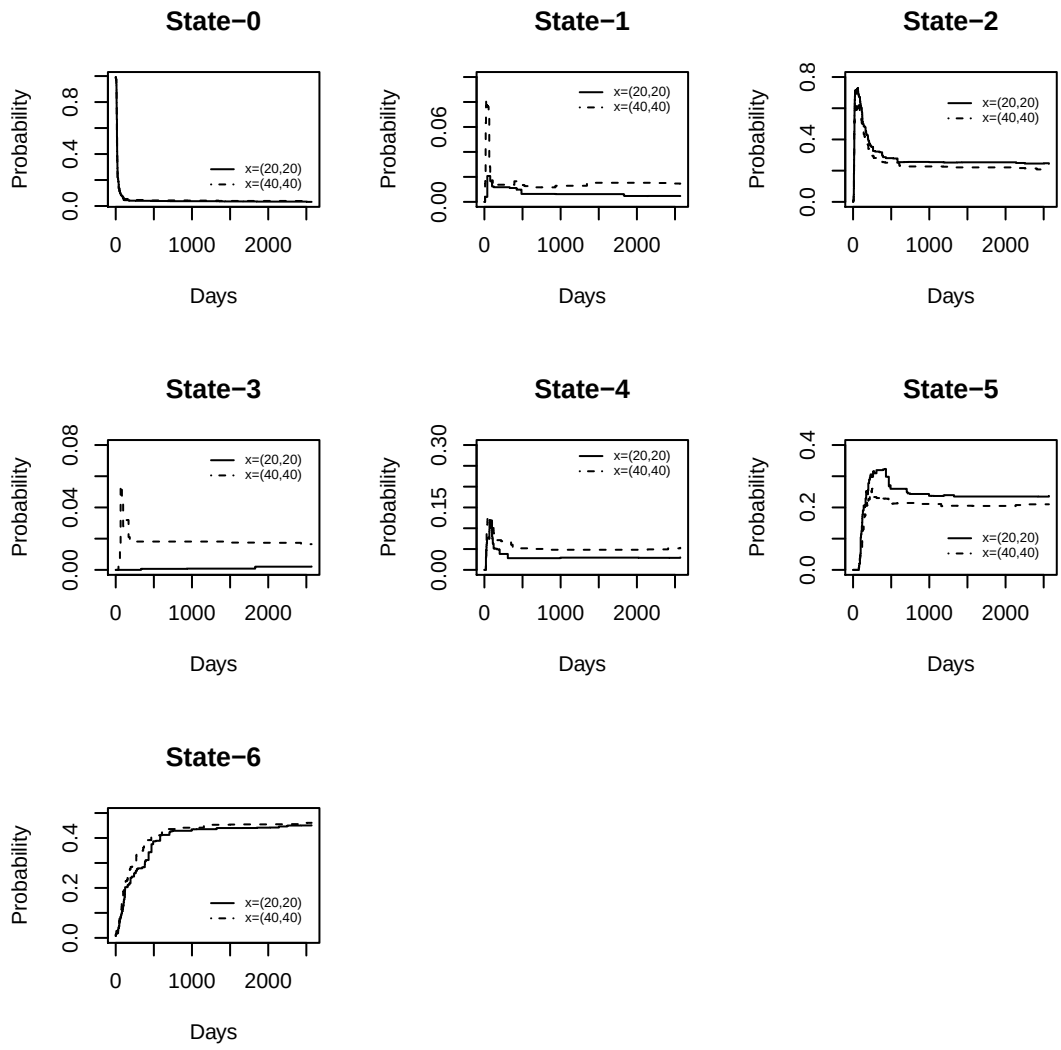

Figure 5.8: Plots of estimated conditional state occupational probabilities of 7 clinical states in the bone marrow transplant data, by the proposed method for a two covariate vectors: $x = (20, 20)$ and $x = (40, 40)$.

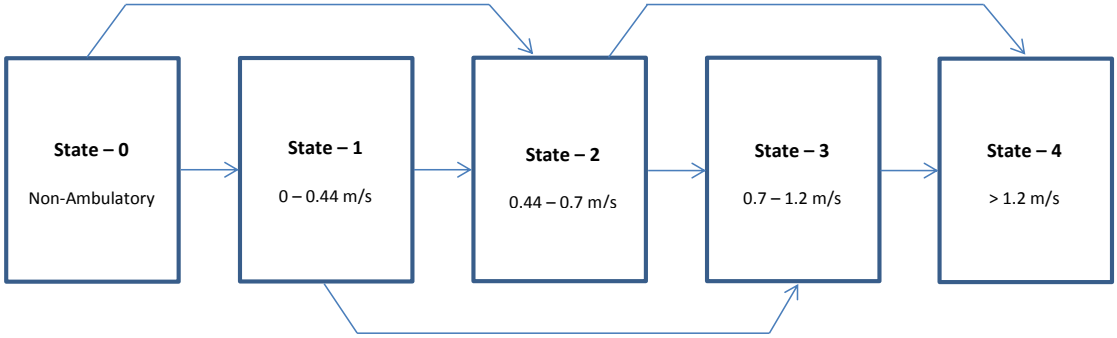

Figure 5.9: Graphical representation of spinal code injury multi-state model

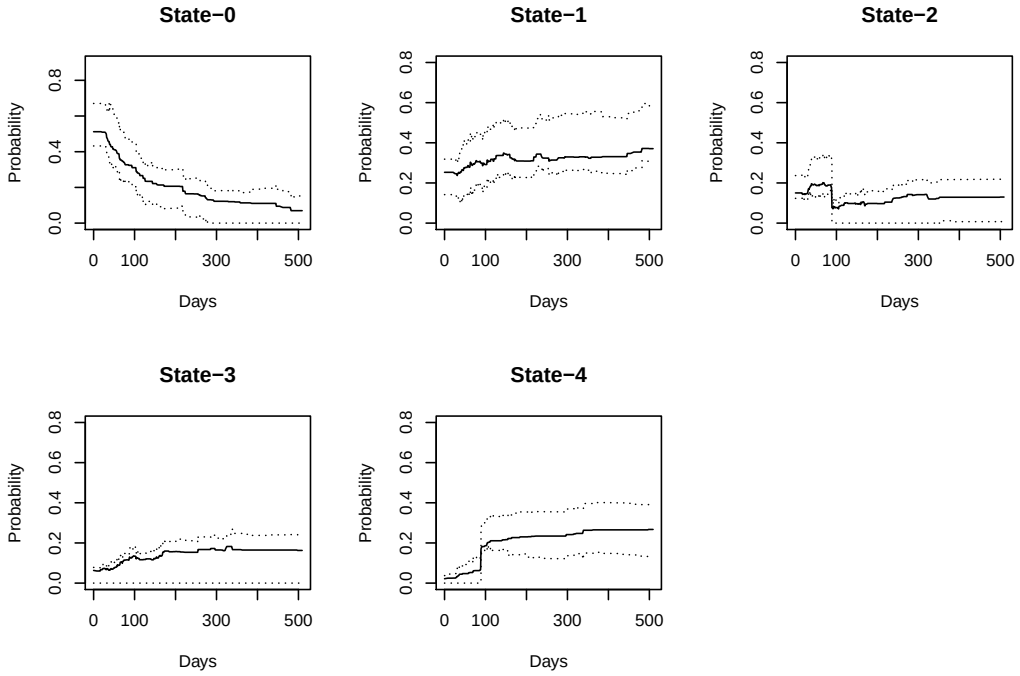

Figure 5.10: Plots of estimated conditional state occupational probabilities of 5 clinical states in the spinal Code Injury data, by the proposed method for a covariate vector of $x = (0.08, 38.0, 0.92, 33.0)'$, along with 95% bootstrap based confidence intervals (represented by the dotted lines).

REFERENCES

[1] Cai, T., Tian, L., Wong, P. H., and Wei, L. J. (2011), Analysis of Randomized Comparative Clinical Trial Data for Personalized Treatment Selections, *Biostatistics*, 12:270–282.

[2] 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*, 335:1081–1090.

[3] Hristache, M., Juditsky, A., Polzehl, J., and Spokoiny, V. (2001), Structure Adaptive Approach for Dimension Reduction, *The Annals of Statistics*, 29:1537–1566.

[4] 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.

[5] Liang, H., Liu, X., Li, R., amd Tsai, C-H. (2010), Estimation and Testing for Partially Linear Single Index Models, *Annals of Statistics*, 38:3811–3836.

[6] Polzehl, J. (2013),EDR: Estimation of the effective dimension reduction (EDR) space, R package version 0.6–5.1.

[7] Qian, M., and Murphy, S. A. (2011), Performance Guarantees for Individualized Treatment Rules. *The Annals of Statistics*, 39:1180–1210.

[8] Schulte, P.J., Tsiatis, A.A., Laber, E.B., and Davidian, M. (2014), Q- and A-learning Methods for Estimating Optimal Dynamic Treatment Regimes, *Statistical Science*, 29(4):640-661.

[9] Robins, J. (2004), Optimal structured nested models for optimal sequential decision, In: Lin DY, Heagerty PJ, editors. *Proceedings of the Second Seattle Symposium on Biostatistics. New York: Springer*, 179:189–326.

[10] Robins, J., Orellana, L., and Rotnitzky, A. (2008), Estimation and extrapolation of optimal treatment and testing strategies, *Statist. Med.*, 27(23):4678–4721

[11] Stute, W. (1991), Conditional U-Statistics, *Annals of Probability*, 19:812–825.

[12] van't Veer, L.J. and Bernards, R. (2008), Enabling Personalized Cancer Medicine Through Analysis of Gene-Expression Patterns, *Nature*, 452:564–570.

[13] Vazquez, A. (2013), Optimization of Personalized Therapies for Anticancer Treatment, *BMC Systems Biology*, 7–31

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

[15] Wang, L. and Yang, L. (2009), Spline Estimation of Single Index Models, *Statistica Sinica*, 19:765–783

[16] Yu, Y. and Ruppert, D. (2002), Penalized Spline Estimation for Partially Linear Single-Index Models, *Journal of the American Statistical Association*, 97:1042–1054.

[17] Zhang, B., Tsiatis A. A., Laber, E. B., and Davidian, M. (2012), A Robust Method for Estimating Optimal Treatment Regimes, *Biometrics*, 68(4):1010–1018.

[18] 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*, 107(449):1106–1118.

[19] Cai, T., Tian, L., Wong, P. H., and Wei, L. J. (2011), Analysis of Randomized Comparative Clinical Trial Data for Personalized Treatment Selections, *Biostatistics*, 97:270–282.

[20] 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.

[21] Ichimura, H., Hall, P., Hardle, W. (1993), Optimal smoothing in single index models, *The Annals of Statistics*, 21(1):157–178.

[22] 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.*, 4–16.

[23] 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.

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

[25] Siriwardhana, C., Zhao, M., Datta, S., and Kulasekera, K.B. (2015), Personalized Plans with Multiple Treatments, Technical Report TR2015_6_cs,mz,sd,kk,

http : *//www.clemson.edu/ces/math/dept_publications.html*, Department of Mathematical Sciences, Clemson University.

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

[27] Aalen O. O. (1976). Non-parametric inference in connection with multiple decrement models. *Scandinavian Journal of Statistics* 3:15–27.

[28] Aalen O. O. (1978). Non-parametric inference for a family of counting processes. *The Annals of Statistics* 6:701–726.

[29] Aalen, O. O. (1980). A model for non-parametric regression analysis of counting processes, In: Klonecki, W., Kozek, A., Rosiski, J. (Eds.), *Lecture Notes on Mathematical Statistics and Probability* 2, Springer, New York, 1–25.

[30] Aalen O. O. (1989). A linear regression model for the analysis of lifetimes. *Statistics in Medicine* 8:907–925.

[31] Aalen O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneouMarkov chaisn based censored observations. *Scandinavian Journal of Statistics* 5:141–150.

[32] Anderson, P. K. and Kieding, N. (2002). Multistate models for event history analysis, *Statistical Methods in Medical Research* 11:91–115.

[33] Anderson, P. K. and Klein, J. P. (2007). Regression analysis for multistate models based on a pseudo-value approach, with application to bone-marrow transplant studies. *Scandinavian Journal of Statistics* 34:3–16.

[34] Barlow, R. E., Bartholowmew, J. M., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. New York: Willey.

[35] Breslow, N. E. (1972). Discussion of the paper by D. R. Cox. *Journal of Royal Statistical Society, Series B* 34:216–217.

[36] Chakraborty, S., Datta, S., and Datta, S. (2015). Nonparametric regression of state occupation probabilities in a multistate Model. *Manuscript submitted for publication.*

[37] Copelan, E.A., Biggs, J. C., Thompson, J. M., Crilley, P., Szer, J., Klein, J. P., Kapoor, N., Avalos, B. R., Cunningham, I., Atkinson, K., Downs, K., Harmons, G. S., Daly, M. B., Brodsky, I., Bulova, S. I., and Tutschka, P. J. (1991). Treatment for Acute Myelocytic Leukemia With Allogeneic Bone Marrow Transplantation Following Preparation With BuCy2. *Blood* 78:838–843.

[38] Cox, D. R. (1972). Regression model and life-tables. *Journal of Royal Statistical Society Series B* 34:187–220.

[39] Datta, S. and Satten, G. A. (2001). Validity of the Aalen-Johansen estimators of state occupation probabilities and integrated transition hazards for non-Markov models. *Statistics and Probability Letters* 55:403–411.

[40] Datta, S. and Satten, G. A (2002). Estimation of integrated transition hazards and stage occupation probabilities for non-Markov system under dependent censoring. *Biometrics* 58:792-802.

[41] Harkema, S.J., Schmidt-Read, M., Behrman, A.L., Bratta, A., Sisto, S.A., Edgerton, V.R. (2912). Establishing the NeuroRecovery Network: Multisite rehabilitation centers that provide activity-based therapies and assessments for neurologic disorders. *Archives of Physical Medicine and Rehabilitation* 93:1498–1507.

[42] Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Monographs on Statistics and Applied Probability 43, Boca Raton: Chapman and Hall/CRC.

[43] Leeuw, J., Hornik, K., Mair, P. (2009). Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods. *Journal of Statistical Software* 32:1–24.

[44] Li, G. and Datta, S. (2001). A Bootstrap Approach to Nonparametric Regression for Right Censored Data. *Annals of the Institute of Statistical Mathematics* 53:708–729.

[45] Lorenz D. J. and Datta S. (2015). A nonparametric analysis of waiting times from a multistate model using a novel linear hazards model approach. *Electronic Journal of Statistics* 9:419–443.

[46] Ichimura, H., Hall, P., Hardle, W. (1993). Optimal smoothing in single index models. *Annals of Statistics* 21:157–178.

[47] Klein J. P. and Moeschberger, M. L. (1997), *Survival Anlsysis: Techniques for Censored and Truncated data*. New York: Springer-Verlag.

[48] Klein R. W. and Spady R. H. (1993). An efficient estimator for binary response models. *Econometrica* 66:387–421

[49] Koul, H., Susarla, V. Van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Annals of Statistics* 9:1276–1288.

[50] Mostjabai, F. and Datta, S. (2013). Nonparametric regression of state occupation, entry, exit, and waiting times with multistate right-censored data. *Statistics in Medicine* 32:3006–3019.

[51] Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analysis for current failure times and time dependent covariates. *Journal of American Statistical Association* 88: 811–820.

[52] Satten, G. A. and Datta, S. (2002). Marginal estimation for multi-state models: waiting time distribution and competing risks analyses. *Statistics in Medicine* 21:3–19.

[53] van Hedel, H.J. and Dietz, V. (2010). Rehabilitation of locomotion after spinal cord injury. *Restorative Neurology and Neuroscience* 28:123–134.

APPENDIX

In this Appendix we provide outlines of the proofs of the technical results of project 1.

Proof. Proof of Lemma 1

First, we prove (ii), the continuity of $p_i(u)$ in s under regularity conditions. Let

$$f_\epsilon(\varepsilon) = \prod_{i=1}^{K} f_\epsilon(\varepsilon_i), \tag{5.3}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_K)'$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_K)'$. For a K-vector $\mathbf{a} = (a_1, \ldots, a_K)'$ and a subset I of $\{1, \ldots, K\}$, let $\mathbf{a}_{(I)}$ denote the $(K - \|I\|)$- vector obtained from $\mathbf{a}$ by removing the ith element from $\mathbf{a}$ for an $i \in I$, and, let $\mathbf{a}_{\{I\}}$ denote the $\|I\|$-vector consisting of a_i for $i \in I$. Without loss of generality, assume $d = 1$. Let $\mathbf{t}_i = (t_{i1}, \ldots, t_{iK})'$ be vectors in $\mathbb{R}^K$ and $\mathbf{t}_j(s) = (t_{j1}(s), t_{j2}(s), \ldots, t_{jK}(s))'$ where

$$t_{ij}(s) = \begin{cases} \max_{k>1} t_{ik} + s, \text{ if } j = 1, \\ t_{ij}, \text{ if } j \neq 1. \end{cases}$$

It can be shown that

$$p_i(u) = \frac{\int \ldots \int \int_{A_s^{(i)}} f_\epsilon(\varepsilon) \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s)) \, d\varepsilon d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)}}{\int \ldots \int \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s)) \, d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)}} \tag{5.4}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_K)'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_K)'$, and
$A_s^{(i)} = \left\{\varepsilon \in \mathbb{R}^K | \varepsilon_i + t_{ii}(s) > \max_{k \neq 1,i} \{\varepsilon_k + t_{kk}(s)\}\right\}$. Let $s' = s + \Delta s$, $u' = (s', 1)$

and $\mathbf{e}_i = \left(0, \ldots, \underset{i\text{th}}{1}, \ldots, 0\right)$. We then have

$$\begin{aligned}
&\int \ldots \int \int_{A_{s'}^{(i)}} f_\epsilon(\varepsilon) \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s')) \, d\varepsilon d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)} \\
&\quad - \int \ldots \int \int_{A_s^{(i)}} f_\epsilon(\varepsilon) \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s)) \, d\varepsilon d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)} \\
&= \int \ldots \int \int_{A_s^{(i)}} f_\epsilon(\varepsilon - \Delta s \mathbf{e}_1) \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s) + \Delta s \mathbf{e}_1) \, d\varepsilon d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)} \\
&= \int \ldots \int \int_{A_s^{(i)}} \left(f_\epsilon(\varepsilon - \Delta s \mathbf{e}_1) \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s) + \Delta s \mathbf{e}_1) - f_\epsilon(\varepsilon) \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s)) \right) \\
&\qquad d\varepsilon d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)},
\end{aligned} \tag{5.5}$$

and

$$\begin{aligned}
&\int \ldots \int \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s')) \, d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)} - \int \ldots \int \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s)) \, d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)} \\
&= \int \ldots \int \left(\prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s) + \Delta s \mathbf{e}_1) - \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s)) \right) d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)}
\end{aligned} \tag{5.6}$$

By Assumption 3, we have

$$\lim_{\Delta s \to 0} p_i(u') - p_i(u) = 0$$

proving the continuity of p_i.

Now we show that $p_1(u) > p_k(u)$ for any $k > 1$. Since the denominator of the right hand side of 5.5 is not affected by i, we only need to show the inequality for the numerator. In the following discussion, we use the assumption that $\epsilon_1, \ldots, \epsilon_K$ are iid random variables with the common pdf f_ϵ. Consider we have

$$\begin{aligned}
I_1 &= \int \ldots \int \int_{A_s^{(1)}} f_\epsilon(\varepsilon) \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s)) \, d\varepsilon d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)} \\
&= \int \ldots \int P\left(\epsilon_1 + t_{11}(s) \geq \max_{j>1} \{\epsilon_j + t_{jj}(s)\} \right) \prod_{j=1}^{K} f_{\mathbf{T}}(\mathbf{t}_j(s)) \, d\varepsilon d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)}
\end{aligned}$$

Now, we have

$$
\begin{aligned}
&P\left(\epsilon_1 + t_{11}(s) \geq \max_{j>1}\left\{\epsilon_j + t_{jj}(s)\right\}\right) \\
&= P\left(\epsilon_1 \geq \epsilon_k + t_{kk}(s) - t_{11}(s) \text{ and } \epsilon_1 \geq \max_{j>1, j\neq k}\left\{\epsilon_j + t_{jj}(s) - t_{11}(s)\right\}\right) \\
&= P\left(\epsilon_k \geq \epsilon_1 + t_{kk}(s) - t_{11}(s) \text{ and } \epsilon_k \geq \max_{j>1, j\neq k}\left\{\epsilon_j + t_{jj}(s) - t_{11}(s)\right\}\right) \\
&\geq P\left(\epsilon_k \geq \epsilon_1 + t_{k1}(s) - t_{1k}(s) \text{ and } \epsilon_k \geq \max_{j>1, j\neq k}\left\{\epsilon_j + t_{jj}(s) - t_{1k}(s)\right\}\right) \\
&= P\left(\epsilon_k \geq \epsilon_1 + t'_{11}(s) - t'_{kk}(s) \text{ and } \epsilon_k \geq \max_{j>1, j\neq k}\left\{\epsilon_j + t'_{jj}(s) - t'_{kk}(s)\right\}\right) \\
&= P\left(\epsilon_k + t'_{kk}(s) \geq \max_{j\neq k}\left\{\epsilon_j + t'_{jj}(s)\right\}\right),
\end{aligned}
$$

where $t'_{ij} = t_{ij}$ if $i \neq 1, k$ and $t'_{1j} = t_{kj}$, $t'_{kj} = t_{1j}$. Thus, we have

$$
\begin{aligned}
I_1 &\geq \int \ldots \int P\left(\epsilon_k + t'_{kk}(s) \geq \max_{j\neq k}\left\{\epsilon_j + t'_{jj}(s)\right\}\right) \prod_{j=1}^{K} f_{\mathbf{T}}\left(\mathbf{t}'_j(s)\right) d\varepsilon d\mathbf{t}'_{1(1)} \ldots d\mathbf{t}'_{K(1)} \\
&= \int \ldots \int \int_{A_s^{(k)}} f_\epsilon(\varepsilon) \prod_{j=1}^{K} f_{\mathbf{T}}\left(\mathbf{t}_j(s)\right) d\varepsilon d\mathbf{t}_{1(1)} \ldots d\mathbf{t}_{K(1)}
\end{aligned}
$$

Considering the assumption of $f_\epsilon(0) > 0$, it can be shown that the above inequality is strict. This results in

$$
p_1(u) > p_k(u).
$$

□

Now we show an intermediate result that would be used in proving Theorem 1. First, for any given $u = (s, d)'$ we define

$$
\tilde{\eta}_J(d) = \prod_{k=1}^{K} I\left[g_d(\boldsymbol{\beta}'_d \boldsymbol{X}_{kj_k}) = \max_m\left\{g_m(\boldsymbol{\beta}'_m \boldsymbol{X}_{kj_k})\right\}\right]
$$

and

$$
\tilde{w}_J(s) = \prod_{i=1}^{K} \frac{1}{h_i} w\left(\frac{s - S(\boldsymbol{X}_{ij_i})}{h_i}\right).
$$

Now, let

$$\tilde{p}_i(u) = \frac{\sum_{J\in\mathcal{J}} I\left[Y_{ij_i} > \max_{k\neq i}\{Y_{kj_k}\}\right]\tilde{w}_J(s)\,\tilde{\eta}_J(d)}{\sum_{J\in\mathcal{J}}\tilde{w}_J(s)\,\tilde{\eta}_J(d)}, \tag{5.7}$$

The following lemma shows that $\tilde{p}_i$ above behaves almost as p_i for any u in large samples.

Lemma 2. *Under Assumptions 1–4, for $u = (s,d)$ such that $f(s,d) > 0$ for s in an open interval containing s, we have $\tilde{p}_i(u) \xrightarrow{P} p_i(u)$ if $h_i \to 0$ and $Nh_i \to \infty$, $i = 1,\ldots,K$.*

The following lemmas are needed for the proof of Lemma 2.

Lemma 3. *Let U and V be positive random variables, defined on a probability space $(\Omega_1, \mathcal{F}_1, P_1)$, and A, B be a subsets of Ω_1. We have (i) $Var(UI_A) \leq Var(U) + E^2(U)$, (ii) $|Cov(UI_A, VI_B)| \leq |Cov(U,V)| + E(U)E(V)$*

Proof.

$$\begin{aligned} Var(UI_A) &= E\left(U^2 I_A\right) - E^2(UI_A) \\ &\leq E\left(U^2\right) - E^2(U) + E^2(U) - E^2(UI_A) \\ &= Var(U) + E^2(U) - E^2(UI_A) \\ &\leq Var(U) + E^2(U) \end{aligned}$$

This proves (i).

$$\begin{aligned} Cov(UI_A, VI_B) &= E(UVI_AI_B) - E(UI_A)E(VI_B) \\ &\leq E(UV) - E(UI_A)E(VI_B) - E(U)E(V) + E(U)E(V) \\ &= Cov(U,V) + E(U)E(V) - E(UI_A)E(VI_B) \\ &\leq Cov(U,V) + E(U)E(V) \leq |Cov(U,V)| + E(U)E(V). \end{aligned}$$

Also,

$$Cov\left(UI_A, VI_B\right) = E\left(UVI_AI_B\right) - E\left(UI_A\right)E\left(VI_B\right) \geq -E\left(UI_A\right)E\left(VI_B\right)$$
$$\geq -E\left(U\right)E\left(V\right) - \left|Cov\left(U,V\right)\right|,$$

proving (ii). □

Now, we define

$$\mu_{w,d,k}\left(h,s\right) = \sum_{d_1=1}^{K}\int \frac{1}{h}w\left(\frac{s-s_1}{h}\right)I\left(d_1=d\right)f\left(s_1,d_1\right)ds_1,$$
$$= \int \frac{1}{h}w\left(\frac{s-s_1}{h}\right)f\left(s_1,d\right)ds_1$$
$$\mu_w = \int w\left(s\right)ds$$

and

$$\sigma^2_{w,d,k}\left(h,s\right) = \sum_{d=1}^{K}\int\left(\frac{1}{h}w\left(\frac{s-s_1}{h}\right)I\left(d_1=d\right)\right)^2 f_k\left(s_1,d\right)ds_1 - \mu^2_{w,d,k}\left(h,s\right)$$
$$= \int\left(\frac{1}{h}w\left(\frac{s-s_1}{h}\right)\right)^2 f_k\left(s_1,d\right)ds_1 - \mu^2_{w,d,k}\left(h,s\right)$$
$$\sigma^2_w = \int w^2\left(s_1\right)ds_1.$$

It can be verified that

$$\lim_{h\searrow 0}\mu_{w,d_0,k}\left(h,s_0\right) = f_k\left(s_0,d_0\right)\mu_w \tag{5.8}$$

and

$$\lim_{h\searrow 0} h\sigma^2_{w,d_0,k}\left(h,s_0\right) = f_k\left(s_0,d_0\right)\sigma^2_w. \tag{5.9}$$

Straight forward calculations yield the following result.

Lemma 4. *For $J, J' \in \mathcal{J}$, let $A\left(J,J'\right) = \left\{1 \leq k \leq K : j_k = j'_k\right\}$, and $B\left(J,J'\right) = \left\{1,\ldots,K\right\} \setminus A\left(J,J'\right)$. Then, for $J, J' \in \mathcal{J}$, we have*

$$E\left(\tilde{w}_J\left(s\right)\tilde{\eta}_J\left(d\right)\right) = \prod_{k=1}^{K}\mu_{w,d,k}\left(h_k,s\right) \tag{5.10}$$

$$Cov\left(\tilde{w}_J(s)\,\tilde{\eta}_J(d), \tilde{w}_{J'}(s)\,\tilde{\eta}_{J'}(d)\right) = \prod_{k\in A(J,J')} \sigma^2_{w,d,k}(h_k,s) \prod_{k\in B(J,J')} \mu^2_{w,d,k}(h_k,s) \tag{5.11}$$

Now we prove Lemma 2.

Proof. First we analyze the numerator pf $\tilde{p}_i$. It can be shown that

$$E\left(I\left[Y_{ij_i} > \max_{k\neq i}\{Y_{kj_k}\}\right]\tilde{w}_J(s)\,\tilde{\eta}_J(d)\right) = p_i \prod_{k=1}^{K} f_k(s,d) + O_p(\max\{h_k\}). \tag{5.12}$$

The proof of this result is a standard procedure for kernel estimation of smooth functions and is omitted here. Next, consider the variance.

$$\begin{aligned}
&Var\left(\sum_{J\in\mathcal{J}} I\left[Y_{ij_i} > \max_{k\neq i}\{Y_{kj_k}\}\right]\tilde{w}_J(s)\,\tilde{\eta}_J(d)\right) \\
&\quad= \sum_{J\in\mathcal{J}}\sum_{J'\in\mathcal{J}} Cov\Big(I\left[Y_{ij_i} > \max_{k\neq i}\{Y_{kj_k}\}\right]\tilde{w}_J(s)\,\tilde{\eta}_J(d), \\
&\qquad\qquad I\left[Y_{ij'_i} > \max_{k\neq i}\{Y_{kj'_k}\}\right]\tilde{w}_{J'}(s)\,\tilde{\eta}_{J'}(d)\Big) \\
&\quad= \sum_{J,J'\in\mathcal{J}, A(J,J')\neq\emptyset} Cov\Big(I\left[Y_{ij_i} > \max_{k\neq i}\{Y_{kj_k}\}\right]\tilde{w}_J(s)\,\tilde{\eta}_J(d), \\
&\qquad\qquad I\left[Y_{ij'_i} > \max_{k\neq i}\{Y_{kj'_k}\}\right]\tilde{w}_{J'}(s)\,\tilde{\eta}_{J'}(d)\Big) \\
&\quad\leq \sum_{J,J'\in\mathcal{J}, A(J,J')\neq\emptyset} \Big|Cov\Big(I\Big[Y_{ij_i} > \max_{k\neq i}\{Y_{kj_k}\}\Big]\tilde{w}_J(s)\,\tilde{\eta}_J(d), \\
&\qquad\qquad I\left[Y_{ij'_i} > \max_{k\neq i}\{Y_{kj'_k}\}\right]\tilde{w}_{J'}(s)\,\tilde{\eta}_{J'}(d)\Big)\Big| \\
&\quad\leq \sum_{J,J'\in\mathcal{J}, A(J,J')\neq\emptyset} \left(|Cov(\tilde{w}_J(s), \tilde{w}_{J'}(s))| + E^2(\tilde{w}_J(s))\right) \text{ (by Lemma 3)} \\
&\leq \sum_{J,J'\in\mathcal{J}, A(J,J')\neq\emptyset} \Big(\prod_{k\in A(J,J')} \sigma^2_{w,d,k}(h_k,s) \prod_{k\in B(J,J')} \mu^2_{w,d,k}(h_k,s) + \prod_{k=1}^{K} \mu^2_{w,d,k}(h_k,s)\Big)
\end{aligned}$$

Thus, by 5.10 and 5.11, for large N,

$$Var\left(\sum_{J\in\mathcal{J}} I\left[Y_{ij_i} > \max_{k\neq i}\{Y_{kj_k}\}\right]\tilde{w}_J(s)\,\tilde{\eta}_J(d).\right) \leq \sum_{J,J'\in\mathcal{J}, A(J,J')\neq\emptyset} c_1 \prod_{k\in A(J,J')} \frac{1}{h_k}, \tag{5.13}$$

where c_1 is a value that is not dependent on N. The right-hand side of Equation 5.13 is the sum of $\prod n_k \left(\prod n_k - \prod (n_k - 1)\right)$ terms. The number of terms for which $\|A(J, J')\| = r$ is of order $O\left(N^r N^{2(K-r)}\right)$. These terms are of the form c_1 divided by the product of r of the h_k's, and thus, the sum of these terms is of order $o\left(N^{2K}\right)$. Therefore we conclude,

$$Var\left(\sum_{J\in\mathcal{J}} I\left[Y_{ij_i} > \max_{k\neq i}\{Y_{kj_k}\}\right]\tilde{w}_J(s)\,\tilde{\eta}_J(d).\right) = o\left(N^{2K}\right).$$

Combining **??** and **??** we have that

$$\frac{1}{\prod n_k}\sum_{J\in\mathcal{J}} I\left[Y_{ij_i} > \max_{k\neq i}\{Y_{kj_k}\}\right]\tilde{w}_J(s)\,\tilde{\eta}_J(d) \xrightarrow{P} p_i\prod_{k=1}^{K} f(s,d). \tag{5.14}$$

Following a similar procedure, we can show that

$$\frac{1}{\prod n_k}\sum_{J\in\mathcal{J}} \tilde{w}_J(s)\,\tilde{\eta}_J(d) \xrightarrow{P} \prod_{k=1}^{K} f(s,d). \tag{5.15}$$

Combining 5.14 and 5.15 we have the desired result.

□

Remark 2. *From the proof of Lemma 1 it can be seen that, to achieve optimal rate of convergence for variances of both the numerator and denominator of the right-hand side of 5.7, the bandwidth h_k need to be of order $N^{-1/5}$ for $k = 1, \ldots, K$.*

We will introduce some additional notation before we prove Theorem 1.

Define

$$F_i^{(n)}(s,d) = \frac{1}{n_i}\sum_{j=1}^{n_i} I\left(\hat{S}(X_{ij}) \leq s, \hat{\delta}(X_{ij}) = d\right)$$

and

$$\tilde{F}_i^{(n)}(s,d) = \frac{1}{n_i}\sum_{j=1}^{n_i} I\left(S(X_{ij}) \leq s, \delta(X_{ij}) = d\right).$$

Furthermore, let $\mathbf{y} = (y_1, \ldots, y_K)'$, $\mathbf{s} = (s_1, \ldots, s_K)'$, $\mathbf{d} = (d_1, \ldots, d_K)'$, $F^{(n)}(\mathbf{s}, \mathbf{d}) = F_1^{(n)} \times \ldots \times F_K^{(n)}$, $\tilde{F}^{(n)}(\mathbf{s}, \mathbf{d}) = \tilde{F}_1^{(n)} \times \ldots \times \tilde{F}_K^{(n)}$, and finally, let $F = F \times F_2 \times \ldots \times F_K$. We need the following assumption that is a very reasonable assumption from empirical distribution results.

Assumption 6. $\left\|\tilde{F}_i - F_i\right\|_\infty = o_p\left(N^{-1/2}\log(N)\right)$.

Proof. Proof of Theorem 1

Let $v(\mathbf{s}) = \prod_{k=1}^{K} \frac{1}{h_k} w\left(\frac{s_i - s_0}{h_k}\right) I\left(s_{d_0} > \max_{k \neq d_0} s_k\right)$. By assumption 2, we have that $v(\mathbf{s})$ is Riemann-integrable. Let $S_{i(j)}$ denote the jth largest of $S(X_{i1}), \ldots, S(X_{in_i})$. Similarly we can define $\hat{S}_{i(j)}$. Since $\sup_{\boldsymbol{x} \in S_{\boldsymbol{x}}} \left|\hat{g}_i\left(\hat{\boldsymbol{\beta}}_i'\boldsymbol{x}\right) - g\left(\boldsymbol{\beta}_i\boldsymbol{x}\right)\right|$ $= O_p\left(N^{-2/5}\log N\right)$, we have, $\sup_{\boldsymbol{x}} \left|S(\boldsymbol{x}) - \hat{S}(\boldsymbol{x})\right| = O_p\left(N^{-2/5}\log N\right)$. Additionally, $\max_{i,j} \left|\hat{S}(X_{ij}) - S(X_{ij})\right| = O_p\left(N^{-2/5}\log N\right)$. After some tedious calculations we can deduce that the above also implies $\max_{i,j} \left|\hat{S}_{i(j)} - S_{i(j)}\right| = O_p\left(N^{-2/5}\log N\right)$. Combine this with 6, and the fact that $\hat{d}_0 \to d_0$ (by the fact that s_0 is positive and the second part of Assumption 5), we can find sets $A_N \subset \Omega$ and positive numbers $a_N \propto N^{-2/5}\log N$ such that for $\omega \in A_N$, $\max_{i,j} \left|\hat{S}_{i(j)} - S_{i(j)}\right| + \left\|\tilde{F}^{(n)} - F^{(n)}\right\|_\infty \leq a_N$ and $\hat{d}_0 = d_0$. Define q_{ij}, $i = 1, \ldots, K; j = 1, \ldots, n_i$ to be values such that $F(q_{ij}, d) = \frac{j}{n_i F(\infty, d)+1}$. With condition 1, we can find $b_N \propto N^{-1/2}\log N$ such that $\max_{i,j} \left|q_{ij} - S_{i(j)}\right| \leq b_N$ if $q_{ij} \in (s_0 - d_1, s_0 + d_1)$, a neighborhood of s_0. Thus, by properly redefining $a_N \propto N^{-2/5}$, we can assume when $\omega \in A_N$, $\max_{i,j} \left|q_{ij} - \hat{S}_{i(j)}\right| \leq a_N$. By condition 2, without loss of generality we can assume that $w\left(\frac{s - s_0}{h_k}\right) = 0$ for s outside $(s_0 - d_1, s_0 + d_1)$. Let $c_1, \ldots, c_m$ be such that $c_i - c_{i-1} = 4b_N$, and the support of w is within $\left(\frac{c_1 - s_0 + 2b_N}{h_k^{(2)}}, \frac{c_m - s_0 - 2b_N}{h_k^{(2)}}\right)$. Define $c_i' = \frac{c_i + c_{i+1}}{2}$ for $i = 1, \ldots, m-1$. For $k = 0, \ldots, 2^K - 1$, let $\mathcal{I}_k$ be collection of K-dimensional intervals of the form $\left[c_{i_1}^* - c_{i_1 - 1}^*\right] \times \ldots \times \left[c_{i_K}^* - c_{i_K - 1}^*\right]$, where $c_{i_j}^*$ is either c_{i_j} or c_{i_j}' depending the whether the jth position of the k when written as a binary number is 0 or 1. For $I \in \mathcal{I}_k$, let $\bar{v}_{I,\mathbf{h}}\left(\underline{v}_{I,\mathbf{h}}\right)$ be the supremum (infimum) of $v(\mathbf{s})$ over $\mathbf{s} \in I$. For $J = (j_1, \ldots, j_K) \in \mathcal{J}$, let $I_J^{(k)}$ denote the I in $\mathcal{I}_k$ such that $\mathbf{s} = (q_{1j_1}, \ldots, q_{Kj_K}) \in I$. By condition 2, we

have

$$\begin{aligned}
&\frac{1}{n_1 \ldots n_K} \sum_{J \in \mathcal{J}} \left(\bar{v}_{I_J^{(k)}, \mathbf{h}} - \underline{v}_{I_J^{(k)}, \mathbf{h}} \right) \\
&\leq \frac{1}{n_1 \ldots n_K} \prod_{k=1}^{K} \lceil 4 b_N n_k \left\| f \right\|_\infty \rceil \sum_{I \in \mathcal{I}_k} \left(\bar{v}_{I, \mathbf{h}} - \underline{v}_{I, \mathbf{h}} \right) \\
&= O\left((4 b_N)^K \sum_{I \in \mathcal{I}_k} \left(\bar{v}_{I, \mathbf{h}} - \underline{v}_{I, \mathbf{h}} \right) \right) \to 0
\end{aligned}$$

for $k = 0, \ldots, 2^K - 1$. Note that the first inequality in the above expression is due to the fact that the number of q_{ij}'s that falls into $[c_k, c_{k+1}]$ ($\left[c'_k, c'_{k+1}\right]$) is less than $\lceil 4 b_N n_i \left\| f \right\|_\infty \rceil$. When $\omega \in A_N$, for $J = (j_1, \ldots, j_K) \in \mathcal{J}$, we have that $\mathbf{S} = (S(X_{1 j_1}), \ldots, S(X_{K j_K}))$ must be in one of the $I_J^{(k)}$'s. Thus,

$$\begin{aligned}
&\frac{1}{n_1 \ldots n_K} \Big(\sum_{J \in \mathcal{J}} I\left[Y_{i j_i} > \max_{k \neq i} \{ Y_{k j_k} \} \right] \tilde{w}_J(s) \, \tilde{\eta}_J(d) \\
&\qquad - \sum_{J \in \mathcal{J}} I\left[Y_{i j_i} > \max_{k \neq i} \{ Y_{k j_k} \} \right] \hat{w}_J(s) \, \hat{\eta}_J(d) \Big) \\
&\qquad\qquad \leq \frac{1}{n_1 \ldots n_K} \sum_{k=0}^{2^K - 1} \sum_{J \in \mathcal{J}_k} \left[\left(\bar{v}_{I_J, \mathbf{h}} - \underline{v}_{I_J, \mathbf{h}} \right) \right] \to 0
\end{aligned}$$

Similarly, we can show that when $\omega \in A_N$,

$$\frac{1}{n_1 \ldots n_K} \left(\sum_{J \in \mathcal{J}} \tilde{w}_J(s) \, \tilde{\eta}_J(d) - \sum_{J \in \mathcal{J}} \hat{w}_J(\hat{s}) \, \hat{\eta}_J(d) \right) \to 0$$

These combined with 5.14 and 5.15 give us the desired result. □

CURRICULUM VITAE

NAME: Chathura Siriwardhana

ADDRESS: Department of Biostatistics and Bioinformatics
University of Louisville
Louisville, KY 40292

EDUCATION: Bachelor of Sceince, University of Peradeniya, 2008
Masters in Applied Statistics, University of Peradeniya, 2012

PROFESSIONAL SOCIETIES: American Statistical Association

PUBLICATIONS: Siriwardhana, C.,Kulasekera, K.B. and Datta, S. Non-parametric regression approach to estimate state occupation probabilities in right-censored multistate models.
In preparation

Siriwardhana, C., Datta, S., and Kulasekera, K.B. Personalized treatment selection for survival outcome.
In preparation

Siriwardhana, C., Zhao, M., Datta, S., and Kulasekera, K.B. Personalized Plans with Multiple Treatments.

Submitted for publication

Siriwardhana, C., Datta, S., and Kulasekera, K.B. Personalized Plans with Multiple outcome measures.
Submitted for publication

Siriwardhana, C., Datta,S., and Datta,S. Inter-platform concordance of gene expression data for the prediction of chemical mode of action.
Submitted for publication

PRESENTATIONS

ORAL

Personalized Plans with Multiple Treatments,
Spring 2016 Seminar Series, Department of Bioinformatics and Biostatistics, University of Louisville, January 29, 2016.

University of Louisville Chronic Disease Management Program: Assessing Risks for Developing Chronic Diseases,
School of Public Health and Information Sciences,
University of Louisville, June 23, 2015.

POSTER

Joint Statistical Meetings-2015, Seattle, WA. August 10, 2016

Research Louisville-2014, Louisville, KY, September 17, 2014