8-2023

# Statistical methods for assessing treatment effects on ordinal outcomes and selecting optimal treatment on survival outcomes using observational data.

Huirong Hu
*University of Louisville*

# STATISTICAL METHODS FOR ASSESSING TREATMENT EFFECTS ON ORDINAL OUTCOMES AND SELECTING OPTIMAL TREATMENT ON SURVIVAL OUTCOMES USING OBSERVATIONAL DATA

By

Huirong Hu
B.S. in Applied Psychology, Anhui University of Chinese Medicine, 2014
M.S. in Biostatistics, University of Louisville, 2019

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

Department of Bioinformatics and Biostatistics
University of Louisville
Louisville, Kentucky

August 2023

# STATISTICAL METHODS FOR ASSESSING TREATMENT EFFECTS ON ORDINAL OUTCOMES AND SELECTING OPTIMAL TREATMENT FOR SURVIVAL OUTCOMES USING OBSERVATIONAL DATA

By

Huirong Hu
B.S. in Applied Psychology, Anhui University of Chinese Medicine, 2014
M.S. in Biostatistics, University of Louisville, 2019

A Dissertation Approved on

August 2, 2023

by the following Dissertation Committee:

---

Maiying Kong, Ph.D., Committee Chair

---

Mitra Riten, Ph.D., Co-Chair

---

Karunarathna B. Kulasekera, Ph.D.

---

Craig James McClain, M.D., Ph.D.

---

Qi Zheng, Ph.D.

# DEDICATION

This dissertation is dedicated to my parents

Mr. Jin Hu

and

Mrs. Cuifen He

who have given me invaluable love, encouragement and support.

# ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisor Dr. Maiying Kong and co-advisor Dr. Riten Mitra for their insightful guidance and unwavering support for my research over the last four years. It has been a great privilege and joy to study under their supervision. In particular, I have been greatly benefited from Dr. Kong's personality and diligence, which I will treasure in my whole life.

I would also like to extend my deep gratitude to my dissertation committee. My sincere and hearty gratitude goes to Dr. Qi Zheng for his ingenious suggestions and invaluable contributions to my research projects, to Dr. Craig James McClain for his valuable guidance on alcohol related diseases and treatments, and to Dr. Karunarathna B. Kulasekera for his consistent support and time to serve on my dissertation committee.

My deep gratitude also goes to the Department of Bioinformatics and Biostatistics for providing me with a great environment to do research, thank all the faculty members for enriching my knowledge in biostatistics and related fields.

My acknowledgements also go to my classmates and friends. Their support and generosity inspire me a lot. Last but not least, I am very thankful to my family members who have been supporting me on every stage of my life.

# ABSTRACT

## STATISTICAL METHODS FOR ASSESSING TREATMENT EFFECTS ON ORDINAL OUTCOMES AND SELECTING OPTIMAL TREATMENT FOR SURVIVAL OUTCOMES USING OBSERVATIONAL DATA

Huirong Hu

August 2, 2023

This dissertation consists of two projects investigating statistical methods in causal inference and personalized medication using observational data.

In the first project, we propose a parametric marginal structural ordinal logistic regression model (MS-OLRM) to assess treatment effects on ordinal outcomes. Average treatment effect (ATE) is used to measure the difference of the mean outcomes if all patients would have been treated compared with the outcomes if they would not have been treated. Many statistical methods have been developed to estimate ATE when the outcome is continuous or binary. The methodology on assessing treatment effect for an ordinal outcome is less studied. For an ordinal outcome, the concept of mean may not be appropriate. For example, the difference in breast cancer between stage II versus stage I is quite different from that between stage IV versus stage III. For an ordinal outcome, we propose use superiority score to measure the treatment effect. Superiority score measures whether the outcome under treatment is stochastically larger than the outcome under control. We propose using the MS-OLRM along with the inverse probability of treatment weighting (IPTW) to estimate

the superiority score under treatment compared with control. This methodology adjusts confounding factors between treatment and outcome by using IPTW. In the weighted sample, all covariates become balanced among different treatment groups. Extensive simulation studies are carried out to examine the performance of the proposed method. We apply the proposed method to assess the treatment effects of medications and behaviour therapies on patients' recovery from alcohol use disorders using the Kentucky Medicaid 2012-2019 database.

In the second project, we propose a doubly robust method for selecting optimal treatment regimen for survival outcome using observational data. In the proposed method, we apply the generalized partial linear single-index models (GPLSIMs) directly to model the contrast functions (i.e., the outcome difference between treatment and control). We consider the outcome under control as nuisance function, and we target to estimate the contrast functions using A-learning method and structural mean model. The optimal treatment regimen is defined as the treatment which results in the optimal outcome. The contrast functions can be consistently estimated if either the outcome model under control or the generalized propensity scores are correctly specified. When the outcome model under control is estimated using GPLSIM, the outcome model is less prone to mis-specification, which results in a more robust estimation for contrast functions and optimal treatment selection. Extensive simulation studies are carried out to examine the performance of the proposed method. The simulation results show the good performance of the proposed method. We apply the proposed method to select the optimal exercise level based on patients' comorbidity and other characteristics using the National Health and Nutrition Examination Survey (NHANES) III data sets.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# STATISTICAL METHODS FOR ASSESSING TREATMENT EFFECTS ON ORDINAL OUTCOMES USING OBSERVATIONAL DATA

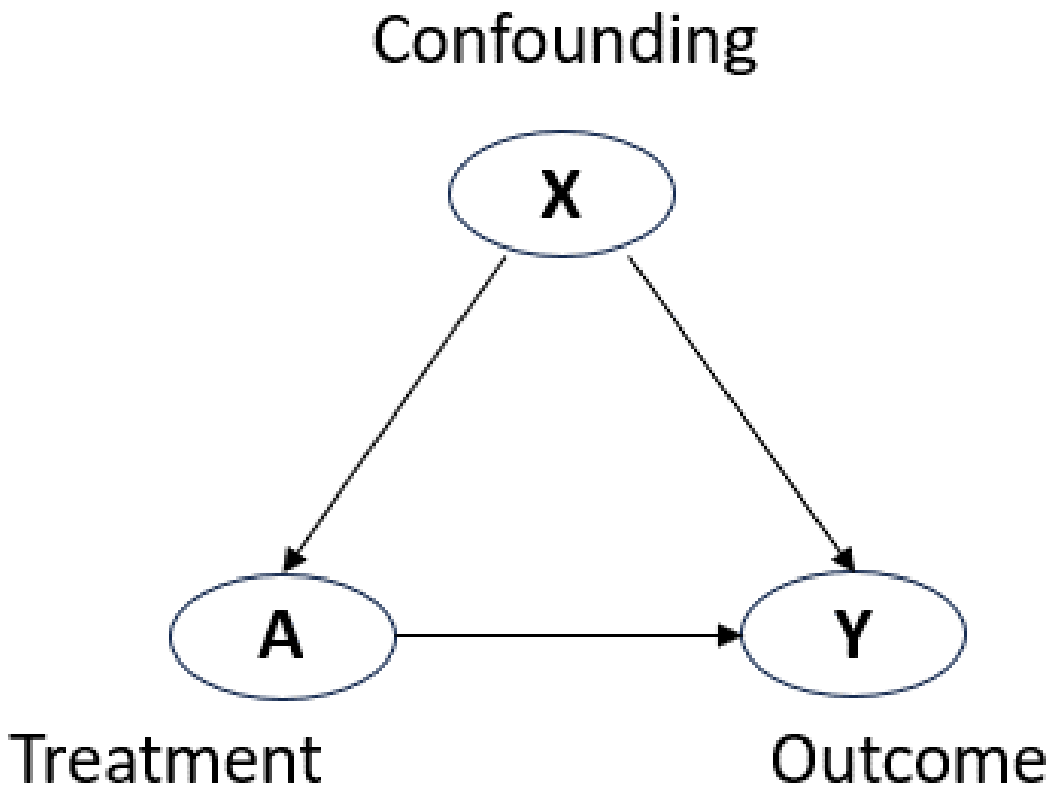
## 1.1 Introduction

Randomized controlled trials (RCTs) are the gold standard to assess treatment effect. However, observational data, such as electronic clinical records data and medical claims data become abundant, and researchers are attempted to evaluate treatment effect from observational data. In an observational study, the relationship between treatment and outcome is often confounded by the third variables (see Figure 1), called confounding variables. A vivid example is illustrated by Simpson's paradox phenomena, where a statistical association which holds in each subgroup is reversed in the entire population (Pearl, 2009). Simpson's paradox happens when both exposure and outcome are strongly associated with the confounding variables (Julious and Mullee, 1994). Sharma et al., (2022) provides an example in comparing the success rate in removing kidney stone: patients who received open surgery had 78% success rate while patients who received percutaneous nephrolithotomy had 83% success rate. However, in patients whose stones were smaller then 2 mm, open surgery had 93% success rate and percutaneous nephrolithotomy had 83% success rate. In patients whose stone were larger than 2 mm stone, open surgery had 73% success rate compared with 69% success rate using percutaneous nephrolithotomy. The reason of

this paradox is that patients who had larger stones were more likely to choose open surgery. Thus, the difference of success rates between two treatment groups are explained by the diameters of stone as well as the treatments. The size of stone, which is a confounding variable, is associated with success rate and treatment selection. As a result, to evaluate the treatment effect accurately and correctly, confounding variables should be controlled or adjusted (Charig et al., 1986). In a RCT, eligible patients are randomly assigned to different treatment groups, thus the treatment assignment is independent to all other variables, and there is no confounding between treatment and outcome. Thus, the treatment effect can be directly estimated by the difference of sample means between treatment and control. In an observed study, it is likely that treatment selection is impacted by patients' characteristics and health conditions, such as age, comorbidities, and severity of diseases. The outcome is not only impacted by treatment but also the patients' characteristics and health conditions. To evaluate treatment effect, generalized propensity score (GPS) and inverse probability of treatment weighting (IPTW) have been introduced to control confounding variables (Guo and Fraser, 2014). GPS is defined as the conditional probability of receiving a particular category of treatment given the confounding variables, which is an extension of propensity score (Rosenbaum and Rubin, 1983; Imbens, 2000). The IPTW method is a widely used GPS based method to estimate treatment effect, where the weight for each patient is obtained as the inverse of the probability of treatment the patient received. Upon weighting, the confounding variables among different treatment groups are expected to be balanced, and the treatment effects can be compared based on the weighted sample (Li et al., 2016; Chattopadhyay et al., 2020; Li and Li, 2021). In other words, the distributions of each confounding variable under different treatment groups are similar in the weighted sample. That is, the confounding is removed between treatment and outcome in the weighted sample.

When outcome is continuous or binary variable, average treatment effect (ATE)

Figure 1.1: Illustration of the relationship between treatment and outcome which is confounded by X.




is often used to measure treatment effect (Rubin, 1974). ATE is defined as the difference of mean outcome when the target population would have been treated versus the target population would have been under control. However, when the outcome is an ordinal variable with more than two categories, the concept of mean outcome is difficult to define because the ordinal outcome themselves are not well-quantified and may not be compared in a numeric scale (Lu, 2018). For example, the stages of a cancer are often graded as stage I through stage IV. However, we can not say that the difference between stage IV and stage III is the same as the difference between stage II and stage I (National Cancer Institute, 2022). Thus, it is not appropriate to take an average for an ordinal outcome. Volfovsky et al. (2015) propose a multidimen-

sional estimand that describes differences in the distributions of potential outcomes under treatment and control in each outcome level. Lu et al. (2018) propose to use two causal parameters, which are defined as the probabilities that the treatment is beneficial and strictly beneficial for the experimental units, under fixed marginal distributions of the potential outcomes. Agresti and Kateri (2017) propose superiority measurement as the probability that an outcome variable under treatment is stochastically superior to the outcome under control, defined as $Pr(Y_1 > Y_0)$, where $Y_1$ is the outcome under treatment and $Y_0$ is the outcome under control. Ryu and Agresti (2018) propose a measure of stochastic superiority score of an outcome under treatment over control as $Pr(Y_1 > Y_0) + 0.5Pr(Y_1 = Y_0)$. Following the literature in causal inference, we are interested in estimating the stochastic superiority score which measures the beneficial effect of treatment over control for a target population using the potential outcome framework.

The statistical methods to evaluate causal effect for the ordinal outcome based on observational study are scarce in the literature. In this project, we surveyed the causal parameters which are suitable for ordinal outcomes. Based on potential outcome framework and the common assumptions on causal inference, we propose using the marginal structure ordinal logistic regression model (MS-OLRM) with IPTW to evaluate the superiority score for potential outcomes under treatment versus control. The remaining part of this paper is organized as follows. Section 2 introduces the superiority score for ordinal outcomes and proposes using the MS-OLRM along with IPTW to estimate the treatment effect for ordinal outcomes. Section 3 presents the simulation studies to assess the performance of the proposed method. Section 4 provides a case study to evaluate the treatment effects of medication and psychotherapy on patients diagnosed with alcohol use disorder. The last section is devoted as conclusion and discussion.

## 1.2 Statistical method to assess treatment effects for ordinal outcomes

### 1.2.1 Basic settings and underlying assumptions

Let $(X, A, Y)$ denote the random variables of observable triplet, where $X$ is confounding variables, $A$ is a treatment variable with $K$ ($K \geq 2$) choices (i.e., $A \in \{0, 1, \cdots, K-1\}$). $Y$ denotes an ordinal outcome variable of $c$ ($c > 2$) levels with possible values from 1 to $c$, a large value indicating a better outcome. To illustrate the estimand we are interested, we use the concept of potential outcomes. Let $Y^{(a)}$ denote the potential outcome when a patient receives treatment level $a$, $a \in \{0, 1, \cdots, K-1\}$. Thus, there are $K$ potential outcomes for a patient, say $Y^{(0)}, Y^{(1)}, \cdots, Y^{(K-1)}$. However, only one potential outcome is observed for each patient, which corresponds to the treatment the patient receives. To compare the treatment $a$ versus control, the target estimand is $\theta_a = Pr(Y^{(0)} < Y^{(a)}) + 0.5Pr(Y^{(0)} = Y^{(a)})$, where $a \in \{1, 2, \cdots, K-1\}$. $\theta_a$ measures the stochastic superiority of treatment $a$ over control in a target population of interest. In this article, we focus on estimating the superiority score $\theta_a$ in the target population from which the sample comes from. We propose using MS-OLRM along with IPTW method to estimate the parameters $\theta_a s$. The IPTW method is based on the generalized propensity scores, which are $Pr(A = a|X)$ for $a = 0, 1, \cdots, K-1$. To estimate $\theta_a$ ($a = 1, 2, \cdots, K-1$) appropriately based on observational data, we need some underlying assumptions, which are presented later in this section. The MS-OLRM is presented in Section 1.2.2, and the inference for $\theta^{(a)}$ is presented in Section 1.2.3.

To make a valid estimation and inference for $\theta_a$, we make the following assumptions: ($i$) Consistency: for a patient, the observed outcome $Y$ is the potential outcome corresponding to the treatment the patient receives (Cole and Frangakis, 2009). That is, $Y = \sum_{a=0}^{K-1} I_{\{A=a\}} Y^{(a)}$ . ($ii$) Exchageability (no unmeasured confounding): given confounding variables $X$, a potential outcome is independent of

treatment assignment, that is, $Y^{(a)} \perp\!\!\!\perp A|X$ for $a = 1, 2, 3$. $(iii)$ Positivity: the probability of being assigned to each treatment group is positive (Hernán, 2012; Hernan and Robins, 2018). That is, $Pr(A = a|X) > 0$ for $a \in \{0, 1, \cdots, K-1\}$. $(iv)$ Correct specification of GPS models, that is, the GPS model needs to be specified.

### 1.2.2 Marginal structural ordinal logistic regression model

Marginal structural ordinal logistic regression model (MS-OLRM) with IPTW is proposed to estimate the superiority scores in observational studies. MS-OLRM describes the causal relationships between treatment and ordinal potential outcomes with the following form:

$$logit\ Pr(Y^{(a)} \leq j) = \alpha_j - \sum_{k=1}^{K-1} I_{\{a=k\}}\tau_k, \tag{1.1}$$

for $j = 1, 2, ..., c-1$, and $a = 0, 1, \cdots, K-1$. Here $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_{c-1})$ captures the distribution of the potential outcome under control in the target population. To illustrate, we set $a = 0$ in equation (1.1), we have $logit\ Pr(Y^{(0)} \leq j) = \alpha_j$ which implies that

$$Pr(Y^{(0)} \leq j) = \frac{exp(\alpha_j)}{1 + exp(\alpha_j)},\ for\ j = 1, 2, \cdots, c-1. \tag{1.2}$$

Similarly, we can see that the parameter component $\tau_a$ in $\boldsymbol{\tau} = (\tau_1, \tau_2, \cdots, \tau_{K-1})$ and $\boldsymbol{\alpha}$ jointly capture the distribution of the potential outcome under treatment $a$. From equation (1.1), we have $logit\ Pr(Y^{(a)} \leq j) = \alpha_j - \tau_a$ $(j = 1, 2, \cdots, c-1)$, which implies that

$$Pr(Y^{(a)} \leq j) = \frac{exp(\alpha_j - \tau_a)}{1 + exp(\alpha_j - \tau_a)}. \tag{1.3}$$

Note that the MS-OLRM describes the relationship between treatment and potential outcomes. In the presence of confounding variables, one cannot directly estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ based on the observed treatment assignment $A$ and outcome $Y$. Instead, a weighted sample is formed, where each observation is weighted by the inverse of the probability of treatment assigned. The distributions of each confounding variable across different treatment groups are similar in the weighted sample. That is, the relationship between treatment and outcome is not confounded any more in the weighted sample. The parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ in the MS-OLRM can be estimated based on the weighted sample. For an observation with a realization $(x, a, y)$, the weight is defined as $w(a, x) = \frac{Pr(A=a)}{Pr(A=a|X=x)}$, where $Pr(A = a)$ is the marginal probability for treatment assignment for treatment level $a$, $Pr(A = a|X = x)$ is the conditional probability, which is often referred as the generalized propensity score (GPS) and estimated using multinomial logistic regression model based on the maximum likelihood method or covariates balance criteria (Deb et al., 2016). The latter is often referred as covariates balance propensity score (CBPS) model (Imai and Ratkovic, 2014). Once the weight for each observation is obtained, one can estimate the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ in the association model $log(Y \leq j|A = a) = \alpha_j - \sum_{k=1}^{K} I_{\{a=k\}}\tau_a$ using the weighted sample. This estimation procedure is implemented using the weighted likelihood approach, where the weights are calculated using the R-package WeightIt. Under the underlying assumptions presented in section 1.2.1 (i.e., consistency, exchangeability, positivity, and correct specification of GPS model), the estimation $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\tau}}$ from the association model have the causal interpretation.

### 1.2.3 Estimation and inference for superiority score $\theta_a$

Once the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ in the MS-OLRM are obtained, one can calculate the superiority score $\theta_a$ $(a = 1, \cdots, K-1)$ and make inference for treatment effect. To calculate the superiority score $\theta_a$, let us denote $\pi_j^{(0)} = Pr(Y^{(0)} = j)$

and $\pi_j^{(a)} = Pr(Y^{(a)} = j),\ j = 1, 2, \cdots, c.$ Set $\boldsymbol{\pi}^{(0)} = (\pi_1^{(0)}, \pi_2^{(0)}, \cdots, \pi_c^{(0)})^T$, and $\boldsymbol{\pi}^{(a)} = (\pi_1^{(a)}, \pi_2^{(a)}, \cdots, \pi_c^{(a)})^T$, then we can get

$$\begin{aligned}
&\theta_a = Pr(Y^{(0)} < Y^{(a)}) + 0.5Pr(Y^{(0)} = Y^{(a)}) \\
=&\sum_{j=1}^{c-1} Pr(Y^{(a)} > j)Pr(Y^{(0)} = j) + 0.5\sum_{j=1}^{c} Pr(Y^{(a)} = j)Pr(Y^{(0)} = j) \\
=&\sum_{j=1}^{c-1}\sum_{k>j}^{c} \pi_j^{(0)}\pi_k^{(a)} + 0.5\sum_{j=1}^{c} \pi_j^{(0)}\pi_j^{(a)} \\
=&\boldsymbol{\pi}^{(a)T} D_\pi \boldsymbol{\pi}^{(0)}, \qquad (1.4)
\end{aligned}$$

where

$$D_\pi = \begin{bmatrix} 0.5 & 0 & \cdots & 0 & 0 \\ 1 & 0.5 & \cdots & 0 & 0 \\ & & \vdots & & \\ 1 & 1 & \cdots & 0.5 & 0 \\ 1 & 1 & \cdots & 1 & 0.5 \end{bmatrix} \in \mathbb{R}^{c\times c}$$

Let denote $\gamma_j^{(0)} = Pr(Y^{(0)} \le j)$ and $\gamma_j^{(a)} = Pr(Y^{(a)} \le j)$, and set $\boldsymbol{\gamma}^{(0)} = (\gamma_1^{(0)}, \gamma_2^{(0)}, \cdots, \gamma_{c-1}^{(0)})^T$ and $\boldsymbol{\gamma}^{(a)} = (\gamma_1^{(a)}, \gamma_2^{(a)}, \cdots, \gamma_{c-1}^{(a)})^T$. $\theta_a$ can also be expressed as

$$\theta_a = \boldsymbol{\gamma}^{(0)T} D_\gamma \boldsymbol{\gamma}^{(a)} + 0.5(1 + \gamma_{c-1}^{(0)} - \gamma_{c-1}^{(a)}) \qquad (1.5)$$

where

$$D_\gamma = \begin{bmatrix} 0 & 0.5 & \cdots & 0 & 0 \\ -0.5 & 0 & \cdots & 0 & 0 \\ & & \vdots & & \\ 0 & 0 & \cdots & 0 & 0.5 \\ 0 & 0 & \cdots & -0.5 & 0 \end{bmatrix} \in \mathbb{R}^{(c-1)\times(c-1)}$$

A proof of this relationship is provided in appendix.

The hypothesis test on whether treatment $a$ is significantly different from control can be written as:

$$H_0 : \theta_a = 0.5 \quad vs \quad H_1 : \theta_a \neq 0.5$$

Upon obtaining the estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$ in the MS-OLRM, one can obtain the estimates for $\gamma^{(0)}$ and $\gamma^{(a)}$, here $\hat{\gamma}_j^{(0)} = \hat{Pr}(Y^{(0)} \leq j) = \frac{exp(\hat{\alpha}_j)}{1+exp(\hat{\alpha}_j)}$ and $\hat{\gamma}_j^{(a)} = \hat{Pr}(Y^{(a)} \leq j) = \frac{exp(\hat{\alpha}_j+\hat{\tau}_a)}{1+exp(\hat{\alpha}_j+\hat{\tau}_a)}$ $(j = 1, 2, \cdots, c-1)$. Further, one can obtain the estimates for $\theta_a$: $\hat{\theta}_a = \hat{\gamma}^{(0)T} D_\gamma \hat{\gamma}^{(a)} + 0.5(1 + \hat{\gamma}_{c-1}^{(0)} - \hat{\gamma}_{c-1}^{(a)})$. To develop test statistics for $H_0 : \theta_a = 0.5$ versus $H_a : \theta_a \neq 0.5$, we need to obtain the variance estimate for $\theta_a$. Let $\Psi(x) = \frac{e^x}{1+e^x}$ denote the cumulative distribution function of a standard logistic distribution, and $\psi(x) = \frac{e^x}{(1+e^x)^2}$ denote its probability density function. From equation (1.5), we

can obtain the first derivatives of $\theta_a$ related to $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$:

$$\begin{aligned}
\frac{\partial \theta_a}{\partial \alpha_1} &= 0.5[-\Psi(\alpha_2)\psi(\alpha_1-\tau_a)+\psi(\alpha_1)\Psi(\alpha_2-\tau_a)],\\
\frac{\partial \theta_a}{\partial \alpha_j} &= 0.5\psi(\alpha_j)[\Psi(\alpha_{j+1}-\tau_a)-\Psi(\alpha_{j-1}-\tau_a)]+0.5(\alpha_j-\tau_a)[\Psi(\alpha_{j-1})-\Psi(\alpha_{j+1})],\\
&\quad \text{for } j=2,\cdots,c-2,\\
\frac{\partial \theta_a}{\partial \alpha_{c-1}} &= 0.5[-\psi(\alpha_{c-1})\Psi(\alpha_{c-2}-\tau_a)+\psi(\alpha_{c-1}-\tau_a)\Psi(\alpha_{c-2})+\psi(\alpha_{c-1})-\psi(\alpha_{c-1}-\tau_a)],\\
\frac{\partial \theta_a}{\partial \tau_a} &= -\Psi(\boldsymbol{\alpha})^T D_\gamma \psi(\boldsymbol{\alpha}-\tau_a)+0.5(\alpha_{c-1}-\tau_a),\\
\frac{\partial \theta_a}{\partial \tau_k} &= 0 \text{ for } k\in\{1,\cdots,a-1,a+1,\cdots,K-1\},
\end{aligned} \tag{1.6}$$

Here $\Psi(\boldsymbol{\alpha})=(\Psi(\alpha_1),\cdots,\Psi(\alpha_{c-1}))^T$ and $\Psi(\boldsymbol{\alpha}-\tau_a)=(\Psi(\alpha_1-\tau_a),\cdots,\Psi(\alpha_{c-1}-\tau_a))^T$. Let $D=(\frac{\partial\theta_a}{\partial\alpha_1},\cdots,\frac{\partial\theta_a}{\partial\alpha_{c-1}},\frac{\partial\theta_a}{\partial\tau_1},\cdots,\frac{\partial\theta_a}{\partial\tau_a})^T$. Let denote $\hat{D}$ as the quantities obtained by replacing $\hat{\boldsymbol{\alpha}}$ and $\hat{\tau}$ by their estimates. Denote $V$ as the variance covariance matrix for $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\tau}})$. Based on the delta method, the asymptotic variance for $\hat{\theta}_a$ can be estimated as

$$\hat{var}(\hat{\theta}_a)=\hat{D}^T\hat{V}\hat{D}. \tag{1.7}$$

The estimated standard error for $\hat{\theta}_a$ can be obtained as $\hat{se}(\hat{\theta}_a)=\sqrt{\hat{var}(\hat{\theta}_a)}$, and $100(1-\alpha)\%$ confidence interval (CI) for $\theta_a$ is obtained as $\hat{\theta}_a \pm z_{\alpha/2}\hat{se}(\hat{\theta}_a)$.

The variance estimates for $\theta_a$ in Equation (1.7) did not consider the variation in estimating the generalized propensity scores. Thus, the variance estimates for $\hat{\theta}_a$ based on the delta method may not be appropriate. An alternative variance estimate is based on the bootstrap resampling method, where B bootstrap samples are drawn, and $\hat{\theta}_a^{(b)}$ $(b=1,2,...,B)$ is obtained from $b^{th}$ bootstrap sample. The variance of $\hat{\theta}_a$ can be estimated as (Horowitz, 2001):

$$\hat{var}_{bs}(\hat{\theta}_a)=\frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_a^{(b)}-\overline{\theta}_a)^2, \tag{1.8}$$

where $\overline{\theta}_a = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_a^{(b)}$. The 95% CI for $\hat{\theta}_a$ can be constructed as $[\ \hat{\theta}_a - 1.96\sqrt{\hat{var}_{bs}(\hat{\theta}_a)}, \hat{\theta}_a + 1.96\sqrt{\hat{var}_{bs}(\hat{\theta}_a)}\ ]$.

## 1.3 Simulation studies

### 1.3.1 Simulation design

In this simulation study, we examine the performance of our proposed method on estimating and testing treatment effect for ordinal outcomes. Assume that there are four groups (K = 4): one control group and three treatment groups. The outcome variable $Y$ has three levels (c = 3), a higher level indicates a better outcome. The confounding variables, say $\mathbf{X}$ with three components, are generated from standard normal distribution with mean 0 and variance 1. The treatment assignment $A$ is generated from a multinomial distribution with probabilities determined by a regression model $log\frac{Pr(A=a|\mathbf{X})}{1-Pr(A=a|\mathbf{X})} = \mathbf{X}^T\delta^{(a)}$, where $a = 0, 1, 2, 3$, $\delta^{(0)} = (0, 0, 0)^T$, $\delta^{(1)} = (0.1, 0.2, 0.3)^T$, $\delta^{(2)} = (-0.1, 0.1, 0.4)^T$, and $\delta^{(3)} = (0.2, -0.1, -0.5)^T$. Given $\mathbf{X}$ and $A$, the outcome is generated from the following cumulative ordinal regression model:

$$ln(\frac{Pr(Y \le j|\mathbf{X}, A)}{1 - Pr(Y \le j|\mathbf{X}, A)}) = \alpha_1 I_{\{j=1\}} + \alpha_2 I_{\{j=2\}} + \mathbf{X}^T\beta - \tau_1 I_{\{A=1\}} - \tau_2 I_{\{A=2\}} - \tau_3 I_{\{A=3\}} \quad (1.9)$$

Here we set $\beta = (-0.1, -0.2, -0.3)^T$, $\alpha_1 = -2$, $\alpha_2 = 1$. $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3) = \kappa\tau^*$ with $\tau^* = (1, 1, 1.5)$ and $\kappa$ varying from 0 to 2.

The simulation study is carried out under three different sample sizes, say $n = 500, 1000$, and 5000. For each sample size, we generate 1000 simulated data sets. For each data set, we estimate $\theta_a$ and test the hypotheses on treatment effect $H_0: \theta_a = 0.5$ vs $H_a: \theta_a \neq 0.5$ for $a = 1, 2, 3$. The simulation studies are carried out in the following Steps:

Step 1: Generate a covariate vector $\mathbf{X}_i$ $(i = 1, 2, \cdots, n)$, where $\mathbf{X}_i \sim MVN(0, I^2_{3\times 3})$.

Step 2: Generate treatment assignment $A_i \in (0, 1, 2, 3)$ for $i = 1, 2, \cdots, n$. Given $\mathbf{X}_i$, the treatment $A_i$ follows a multinomial distribution with the following probabilities:

$$P(A_i = 0|\mathbf{X}_i) = \frac{1}{1 + \sum_{k=1}^{3} exp(\mathbf{X}_i^T \delta^{(k)})}, \text{ and}$$
$$P(A_i = a|\mathbf{X}_i) = \frac{exp(\mathbf{X}_i^T \delta^{(a)})}{1 + \sum_{k=1}^{3} exp(\mathbf{X}_i^T \delta^{(k)})}, \ a = 1, 2, 3.$$

where $\delta^{(1)} = (0.1, 0.2, 0.3)^T$, $\delta^{(2)} = (-0.1, 0.1, 0.4)^T$, and $\delta^{(3)} = (0.2, -0.1, -0.5)^T$.

Step 3: Generate observed outcome $Y$ based on the equation (1.9), given $A$ and $\mathbf{X}$.

Step 4: Estimate the parameters $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\tau}}$ in the MS-OLRM using IPTW method, then $\theta_a$ was estimated as:

$$\hat{\theta}_a = \hat{\gamma}^{(0)T} D_\gamma \hat{\gamma}^{(a)} + 0.5(1 + \hat{\gamma}_2^{(0)} - \hat{\gamma}_2^{(a)}), \text{ where } \quad D_\gamma = \begin{bmatrix} 0 & 0.5 \\ -0.5 & 0 \end{bmatrix}.$$

The variance of $\hat{\theta}_a$ is estimated by the delta method as well as bootstrap method based on $B = 100$ bootstrap samples, and the resulting variances are denoted as $MSM.SE$ and $BS.SE$ respectively. The hypothesis test for $H_0 : \ \theta_a = 0.5$ versus $H_a : \ \theta_a \neq 0.5$ is carried out by examining whether the 95% CI for $\theta_a$ includes 0.5.

Step5: Estimate the parameters $\hat{\alpha}$ and $\hat{\tau}$ in the MS-OLRM without IPTW method, and obtain the variance estimates using the delta method and the bootstrap method, respectively.

Step 6: Repeat Steps 1-5 for 1000 times. The simulation results for the 1000 samples are summarized by (i) the empirical standard error for the 1000 estimates of $\theta_a$ (denoted as $Emp.SE$ in Table 1.1 - 1.3), which is the standard deviation of

the 1000 estimated $\hat{\theta}_a$ $(a = 1, 2, 3)$; (ii) the mean of estimated standard errors based on the delta method and bootstrap method (denoted as $MSM.SE$ and $BS.SE$ respectively in Table 1.1 - 1.3); (iii) the proportion of rejection for the hypothesis $H_0 : \theta_a = 0.5$ vs $H_a : \theta_a \neq 0.5$.

Step 7: Repeat Steps 1-6 for each $\tau$, where $\tau = \kappa \times \tau^*$ with $\kappa$ taking values from 0 to 2 by 0.2.

In addition, the quantile-quantile plots (Q-Q plots) based on 1000 simulated data sets for testing the hypothesis $H_0 : \theta_a = 0.5$ versus $H_a : \theta_a \neq 0.5$ $(a = 1, 2, 3)$ are obtained when the null hypothesis is correct (i.e., $\kappa = 0$). The Q-Q plots are presented in Figure 1.2 for different estimation methods under different sample sizes.

### 1.3.2 Simulation results

The empirical standard error ($Emp.SE$), the mean of standard errors based on bootstrap method ($BS.SE$) and sandwich method ($MSM.SE$) are presented in Table 1.1 for different $\kappa$ for sample size $n = 500$. The empirical standard error, the mean of standard error based on bootstrap method and sandwich estimates for sample sizes 1000 and 5000 are presented in Table 1.2 and 1.3, respectively. Based on Tables 1.1 - 1.3, it is clear that the mean of standard error estimates based on bootstrap method is close to the empirical standard error, while the empirical standard error is much larger than the mean of standard error from delta method ($MSM.SE$), indicating that the bootstrap method provides a more accurate standard error estimator than delta method.

The Q-Q plots (Figure 1.2), where the null hypotheses are correct, show that the empirical p-values and the theoretical p-values are almost identical when the IPTW method and the bootstrap standard error estimates are applied. The Q-Q plots are far from the diagonal straight lines when unweighted method is applied or the delta method based standard error estimates is applied. These results indicate that the

IPTW method and the bootstrap standard error estimates provide correct rejection rates for the hypothesis tests. In the following, we only present the simulation results with standard error estimated by bootstrap method. The additional simulation results are presented in Figure 1.3 to Figure 1.5 under different sample sizes (n = 500, 1000, and 5000). In each figure, the first row (panels $A_1$, $A_2$, and $A_3$) summarizes the empirical standard error and the mean of estimated standard error based on bootstrap method, which shows that the mean of estimated standard errors is close to the empirical standard error regardless IPTW is used or not. The second row (panels $B_1$, $B_2$, and $B_3$) shows the power of the test for $H_0: \theta_a = 0.5$ VS $H_a: \theta_a \neq 0.5$ for $a = 1, 2, 3$ based on weighted and unweighted method with the standard errors estimated by bootstrap method. Based on the simulation results in Figures 1.3 - 1.5, we conclude that (1) the empirical standard error ($Emp.SE$) is close to the bootstrap standard error ($BS.SE$) for both weighted and unweighted method (see panels $A_1$ - $A_3$), indicating the bootstrap SE estimates are more appropriate to capture the variation of the estimates of $\theta_1$, $\theta_2$, and $\theta_3$; (2) the size of the test under the null hypothesis (i.e., $\kappa = 0$ in panels $B_1$ - $B_3$) based on the IPTW method and the bootstrap error estimate is about 5%, while the test based on unweighted method has a larger type I error rate, indicating that the tests based on unweighted method do not control type I error rate; (3) when the sample size increases from 500 (Figure 1.3) to 1000 (Figure 1.4) and 5000 (Figure 1.5), the standard errors decreased, the power of the test increased for each fixed $\kappa$ for the IPTW estimates, while the type I errors based on the unweighted estimates increased as sample size increased for $\theta_2$ and $\theta_3$. Therefore, the proposed MS-OLR model using IPTW method and bootstrap standard error estimate performs well in our simulation studies.

Figure 1.2: Q-Q plots for testing $H_0 : \theta_a = 0.5$ vs $H_a : \theta_a \neq 0.5$ $(a = 1, 2, 3)$ when the null hypothesis is correct under different sample sizes.

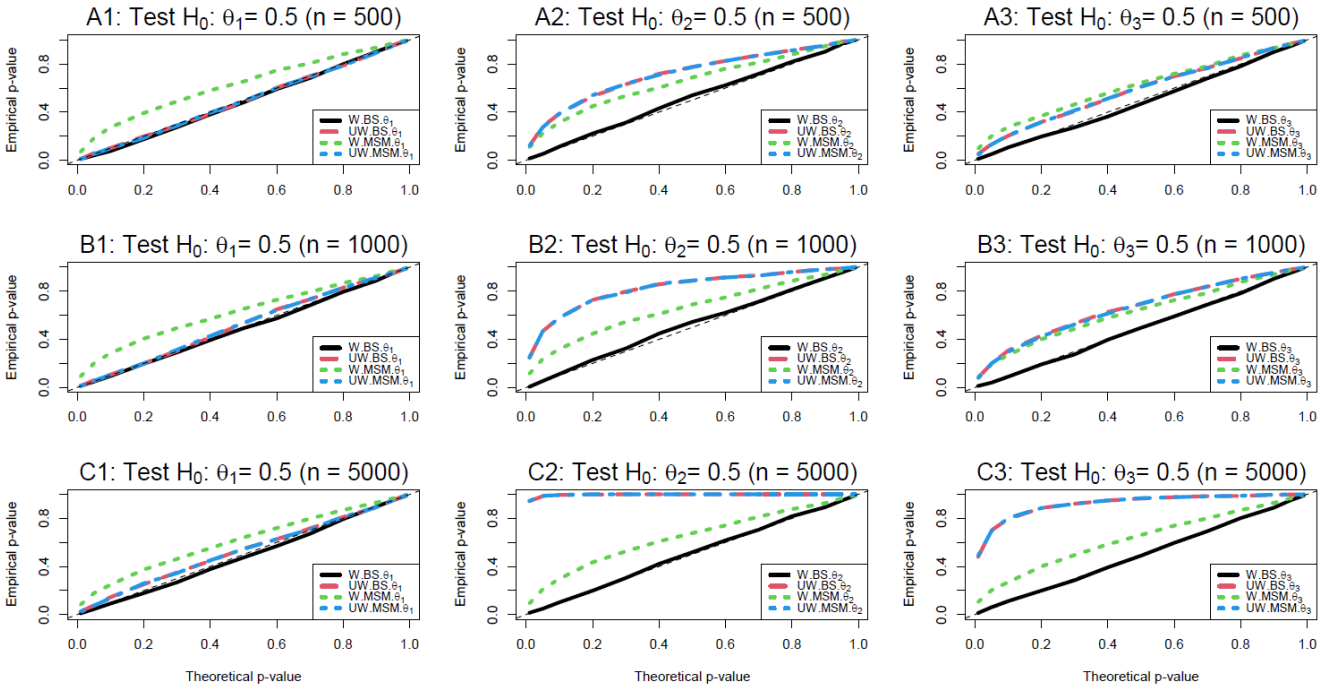
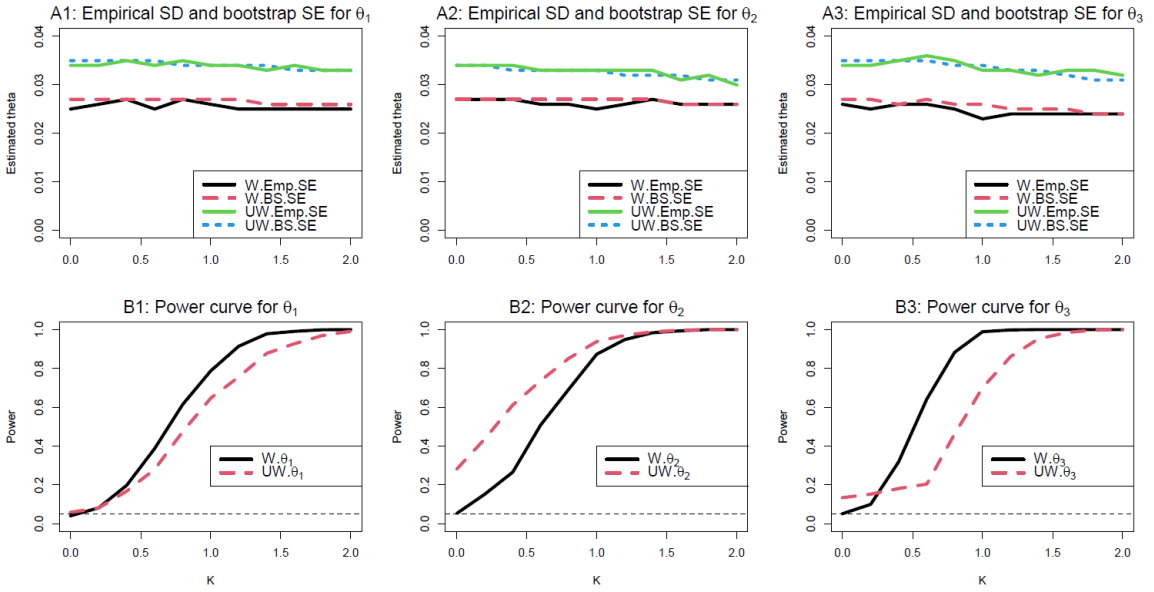



Figure 1.3: Simulation results for testing $\theta_a = 0.5$ vs $\theta_a \neq 0.5$ for sample size n = 500, where $a = 1, 2, 3$.




where the first row shows the estimated SE and empirical SD under different methods, and the third row shows the power of the test for $H_0 : \theta_a = 0.5$ $(a = 1, 2, 3)$.

Figure 1.4: Simulation results for standard error estimation and testing $\theta_a = 0.5$ vs $\theta_a \neq 0.5$ for sample size n = 1000, where $a = 1, 2, 3$.

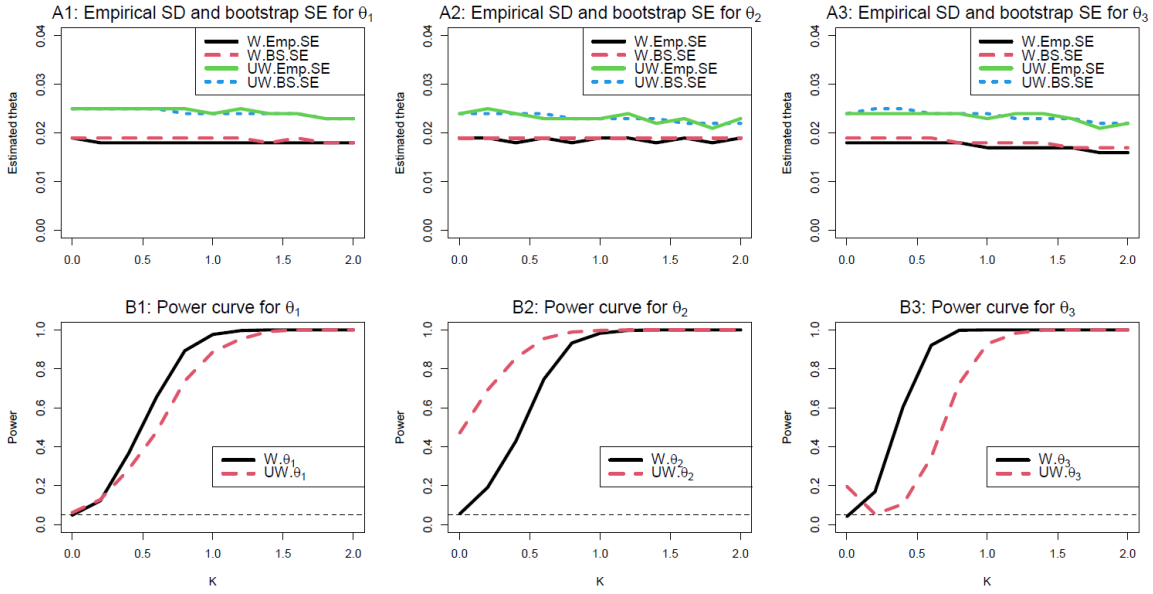



Figure 1.5: Simulation results for estimating the causal parameters ($\theta_1$, $\theta_2$, and $\theta_3$) for sample size n = 5000.

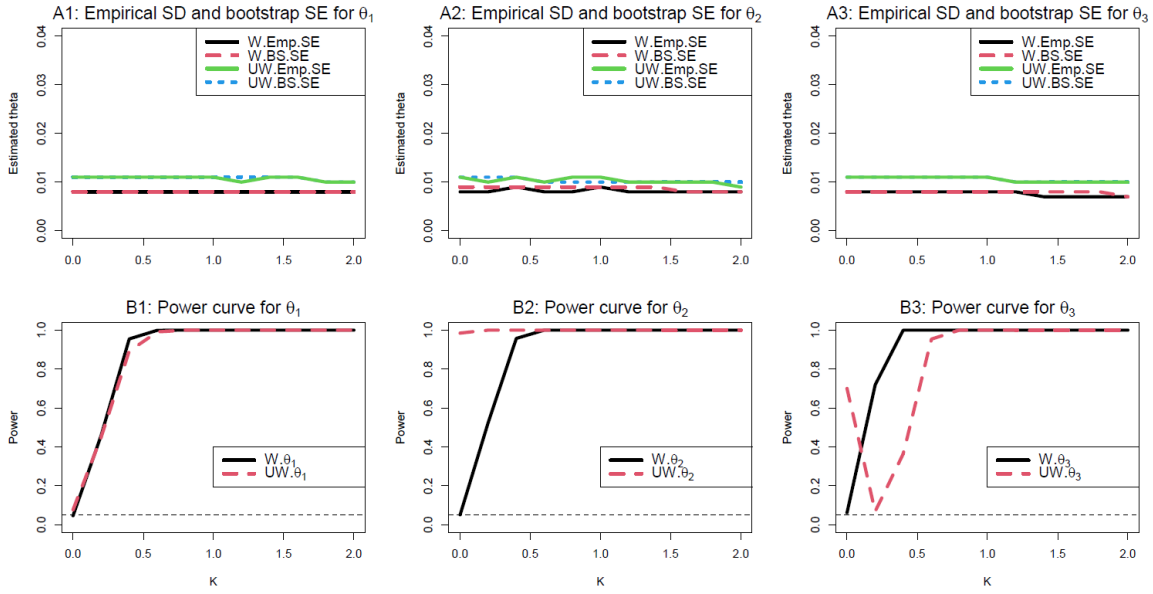


Table 1.1: Evaluation of standard error for sample size n = 500.

| | kappa | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1 | 1.2 | 1.4 | 1.6 | 1.8 | 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_1$ | Emp.SE | 0.025 | 0.026 | 0.027 | 0.025 | 0.027 | 0.026 | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 |
| | BS.SE | 0.027 | 0.027 | 0.027 | 0.027 | 0.027 | 0.027 | 0.027 | 0.026 | 0.026 | 0.026 | 0.026 |
| | MSM.SE | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.016 | 0.016 | 0.016 | 0.016 |
| $\theta_2$ | Emp.SE | 0.027 | 0.027 | 0.027 | 0.026 | 0.026 | 0.025 | 0.026 | 0.027 | 0.026 | 0.026 | 0.026 |
| | BS.SE | 0.027 | 0.027 | 0.027 | 0.027 | 0.027 | 0.027 | 0.027 | 0.027 | 0.026 | 0.026 | 0.026 |
| | MSM.SE | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.016 | 0.016 | 0.016 | 0.016 |
| $\theta_3$ | Emp.SE | 0.026 | 0.025 | 0.026 | 0.026 | 0.025 | 0.023 | 0.024 | 0.024 | 0.024 | 0.024 | 0.024 |
| | BS.SE | 0.027 | 0.027 | 0.026 | 0.027 | 0.026 | 0.026 | 0.025 | 0.025 | 0.025 | 0.024 | 0.024 |
| | MSM.SE | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 | 0.016 | 0.016 | 0.016 | 0.016 | 0.015 | 0.015 |

Table 1.2: Evaluation of standard error for sample size n = 1000.

| | kappa | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1 | 1.2 | 1.4 | 1.6 | 1.8 | 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_1$ | Emp.SE | 0.019 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 |
| | BS.SE | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.018 | 0.019 | 0.018 | 0.018 |
| | MSM.SE | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 |
| $\theta_2$ | Emp.SE | 0.019 | 0.019 | 0.018 | 0.019 | 0.018 | 0.019 | 0.019 | 0.018 | 0.019 | 0.018 | 0.019 |
| | BS.SE | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
| | MSM.SE | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 | 0.011 |
| $\theta_3$ | Emp.SE | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 | 0.017 | 0.017 | 0.017 | 0.017 | 0.016 | 0.016 |
| | BS.SE | 0.019 | 0.019 | 0.019 | 0.019 | 0.018 | 0.018 | 0.018 | 0.018 | 0.017 | 0.017 | 0.017 |
| | MSM.SE | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 | 0.011 | 0.011 | 0.01 |

Table 1.3: Evaluation of standard error for sample size n = 5000.

| | kappa | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1 | 1.2 | 1.4 | 1.6 | 1.8 | 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_1$ | Emp.SE | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
| | BS.SE | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
| | MSM.SE | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| $\theta_2$ | Emp.SE | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.009 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
| | BS.SE | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.008 | 0.008 | 0.008 |
| | MSM.SE | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| $\theta_3$ | Emp.SE | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.009 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
| | BS.SE | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.009 | 0.008 | 0.007 | 0.007 | 0.007 | 0.007 |
| | MSM.SE | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |

## 1.4 Case study

Alcohol use becomes the seventh leading risk factor for both disabilities and deaths and contributes to three million deaths each year globally (Esser et al., 2020). Excessive alcohol consumption can cause various organ diseases, including but not

limited to alcoholic cardiomyopathy, alcoholic liver disease, alcoholic gastritis, and alcoholic polyneuropathy (Vittadini et al., 2001; Piano, 2002; O'shea et al., 2010). Medication for alcohol abuse and alcohol dependence includes three Food and Drug Administration (FDA) approved drugs (naltrexone, acamprosate, and disulfiram) and one off-label drug (topiramate) which is approved by the U.S. Department of Veterans Affairs (VA) (Litten et al., 2016; Kranzler and Soyka, 2018; Witkiewitz et al., 2019). Other than medications, behaviour treatments, such as consultation or psychotherapy, can help recovery from alcohol use disorder (National Institutes of Health, 2014). In this study, we evaluate the treatments effect on patients with alcohol dependence or alcohol abuse using Kentucky Medicaid data from 2012 to 2019.

Figure 1.6: Covariate balance scores for the case study of patients with alcohol abuse/dependence.

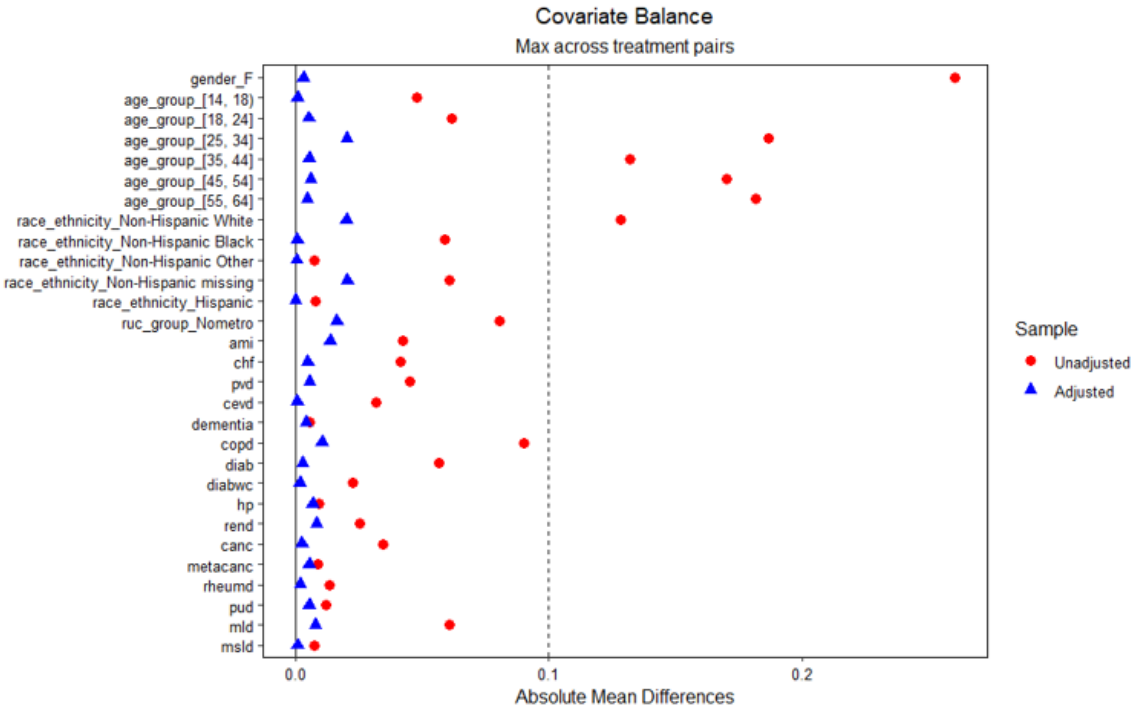


Table 1.4: The distributions of confounding variables among four different treatment groups.

| Variables | Levels | Total N (%) | No treatment N (%) | Behaviour therapy Only N (%) | Medication Only N (%) | Combination N (%) | P-Value |
|---|---|---|---|---|---|---|---|
| Overall | | 11819 | 5848 (49.5%) | 5053 (42.7%) | 352 (3.0%) | 566 (4.8%) | |
| Gender | Male | 6886 (58.3%) | 3786 (64.7%) | 2702 (53.5%) | 179 (50.9%) | 219 (38.7%) | < 0.001 |
| | Female | 4933 (41.7%) | 2062 (35.3%) | 2351 (46.5%) | 173 (49.1%) | 347 (61.3%) | |
| Age | [14, 18) | 579 (4.8%) | 183 (3.0%) | 371 (7.3%) | 9 (2.6%) | 16 (2.8%) | < 0.001 |
| | [18, 24] | 976 (8.0%) | 318 (5.2%) | 586 (11.5%) | 21 (6.0%) | 51 (9.0%) | |
| | [25, 34] | 2513 (20.6%) | 840 (13.6%) | 1400 (27.5%) | 86 (24.4%) | 187 (33.0%) | |
| | [35, 44] | 2713 (22.3%) | 1169 (18.9%) | 1254 (24.6%) | 102 (29.0%) | 188 (33.2%) | |
| | [45, 54] | 3255 (26.7%) | 1997 (32.4%) | 1058 (20.8%) | 103 (29.3%) | 97 (17.1%) | |
| | [55, 64] | 1783 (14.6%) | 1341 (21.7%) | 384 (7.5%) | 31 (8.8%) | 27 (4.8%) | |
| Race and Ethnicity | Non-Hispanic White | 8396 (71.0%) | 3941 (67.4%) | 3733 (73.9%) | 268 (76.1%) | 454 (80.2%) | < 0.001 |
| | Non-Hispanic Black | 1356 (11.5%) | 738 (12.6%) | 552 (10.9%) | 28 (8.0%) | 38 (6.7%) | |
| | Non-Hispanic Other | 102 (0.9%) | 51 (0.9%) | 47 (0.9%) | 3 (0.9%) | 1 (0.2%) | |
| | Non-Hispanic Missing | 1905 (16.1%) | 1088 (18.6%) | 697 (13.8%) | 49 (13.9%) | 71 (12.5%) | |
| | Hispanic | 60 (0.5%) | 30 (0.5%) | 24 (0.5%) | 4 (1.1%) | 2 (0.4%) | |
| RUC | Metro | 6464 (54.7%) | 3073 (52.5%) | 2850 (56.4%) | 198 (56.2%) | 343 (60.6%) | < 0.001 |
| | No Metro | 5354 (45.3%) | 2775 (47.5%) | 2202 (43.6%) | 154 (43.8%) | 223 (39.4%) | |
| Acute Myocardial Infarction | Yes | 465 (3.9%) | 330 (5.6%) | 116 (2.3%) | 11 (3.1%) | 8 (1.4%) | < 0.001 |
| Congestive Heart Failure | Yes | 524 (4.4%) | 376 (6.4%) | 122 (2.4%) | 13 (3.7%) | 13 (2.3%) | < 0.001 |
| Peripheral Vascular Disease | Yes | 619 (5.2%) | 431 (7.4%) | 154 (3.0%) | 10 (2.8%) | 24 (4.2%) | < 0.001 |
| Cerebrovascular Disease | Yes | 464 (3.9%) | 317 (5.4%) | 114 (2.3%) | 16 (4.5%) | 17 (3.0%) | < 0.001 |
| Dementia | Yes | 48 (0.4%) | 31 (0.5%) | 16 (0.3%) | 1 (0.3%) | 0 (0%) | 0.143 |
| Chronic Obstructive Pulmonary Disease | Yes | 3729 (31.6%) | 2089 (35.7%) | 1368 (27.1%) | 94 (26.7%) | 178 (31.4%) | < 0.001 |
| Rheumatoid Disease | Yes | 218 (1.8%) | 135 (2.3%) | 62 (1.2%) | 9 (2.6%) | 12 (2.1%) | <0.001 |
| Peptic Ulcer Disease | Yes | 189 (1.6%) | 119 (2.0%) | 61 (1.2%) | 3 (0.9%) | 6 (1.1%) | 0.004 |
| Mild Liver Disease | Yes | 1275 (10.8%) | 664 (11.4%) | 501 (9.9%) | 29 (8.2%) | 81 (14.3%) | 0.001 |
| Moderate or Severe Liver Disease | Yes | 88 (0.7%) | 59 (1.0%) | 26 (0.5%) | 1 (0.3%) | 2 (0.4%) | 0.014 |
| Diabetes without Complication | Yes | 1549 (13.1%) | 899 (15.4%) | 550 (10.9%) | 45 (12.8%) | 55 (9.7%) | < 0.001 |
| Diabetes with Complication | Yes | 452 (3.8%) | 287 (4.9%) | 133 (2.6%) | 17 (4.8%) | 15 (2.7%) | <0.001 |
| Hemiplegia or Paraplegia | Yes | 119 (1.0%) | 84 (1.4%) | 30 (0.6%) | 2 (0.6%) | 3 (0.5%) | <0.001 |
| Renal Disease | Yes | 345 (2.9%) | 220 (3.8%) | 105 (2.1%) | 13 (3.7%) | 7 (1.2%) | < 0.001 |
| Cancer (Any Malignancy) | Yes | 322 (2.7%) | 242 (4.1%) | 70 (1.4%) | 6 (1.7%) | 4 (0.7%) | < 0.001 |
| Metastatic Solid Tumor | Yes | 62 (0.5%) | 51 (0.9%) | 9 (0.2%) | 2 (0.6%) | 0 (0%) | <0.001 |

Table 1.5: The distributions of confounding variables stratified by outcomes.

| Variables | Levels | Total N (%) | Alcohol related organ disease N (%) | Alcohol abuse/dependence N (%) | Remission N (%) | P-Value |
|---|---|---|---|---|---|---|
| Overall | | 11819 | 669 (5.7%) | 10730 (90.8%) | 420 (3.5%) | |
| Gender | Male | 6886 (58.3%) | 477 (66.8%) | 6211 (57.9%) | 228 (54.3%) | < 0.001 |
| | Female | 4933 (41.7%) | 222 (33.2%) | 4519 (42.1%) | 192 (45.7%) | |
| Age | [14, 18) | 579 (4.8%) | 3 (0.4%) | 570 (5.2%) | 6 (1.3%) | < 0.001 |
| | [18, 24] | 976 (8.0%) | 12 (1.7%) | 937 (8.5%) | 27 (6.0%) | |
| | [25, 34] | 2513 (20.6%) | 73 (10.4%) | 2363 (21.4%) | 77 (17.0%) | |
| | [35, 44] | 2713 (22.3%) | 150 (21.4%) | 2457 (22.3%) | 106 (23.5%) | |
| | [45, 54] | 3255 (26.7%) | 280 (39.9%) | 2851 (25.9%) | 124 (27.4%) | |
| | [55, 64] | 1783 (14.6%) | 151 (21.5%) | 1552 (14.1%) | 80 (17.7%) | |
| Race and Ethnicity | Non-Hispanic White | 8396 (71.0%) | 473 (70.7%) | 7602 (70.8%) | 321(76.4%) | 0.019 |
| | Non-Hispanic Black | 1356 (11.5%) | 66 (9.9%) | 1259 (11.7%) | 31 (7.4%) | |
| | Non-Hispanic Other | 102 (0.9%) | 3 (0.4%) | 98 (0.9%) | 1 (0.2%) | |
| | Non-Hispanic Missing | 1905 (16.1%) | 126 (18.8%) | 1714 (16.0%) | 65 (15.5%) | |
| | Hispanic | 60 (0.5%) | 1 (0.1%) | 57 (0.5%) | 2 (0.4%) | |
| RUC | Metro | 6464 (54.7%) | 346 (51.7%) | 5872 (54.7%) | 246 (58.6%) | 0.085 |
| | No Metro | 5354 (45.3%) | 323 (48.3%) | 4857 (45.3%) | 174 (41.1%) | |
| Acute Myocardial Infarction | Yes | 465 (3.9%) | 47 (7.0%) | 396 (3.7%) | 22 (5.2%) | < 0.001 |
| Congestive Heart Failure | Yes | 524 (4.4%) | 57 (8.5%) | 444 (4.1%) | 23 (5.5%) | < 0.001 |
| Peripheral Vascular Disease | Yes | 619 (5.2%) | 52 (7.8%) | 537 (5.0%) | 30 (7.1%) | 0.001 |
| Cerebrovascular Disease | Yes | 464 (3.9%) | 34 (5.1%) | 409 (3.8%) | 21 (5.0%) | 0.121 |
| Dementia | Yes | 48 (0.4%) | 0 (0.0%) | 45 (0.4%) | 3 (0.7%) | 0.106 |
| Chronic Obstructive Pulmonary Disease | Yes | 3729 (31.6%) | 258 (38.6%) | 3315 (30.9%) | 156 (37.1%) | < 0.001 |
| Rheumatoid Disease | Yes | 218 (1.8%) | 11 (1.6%) | 192 (1.8%) | 15 (3.6%) | 0.046 |
| Peptic Ulcer Disease | Yes | 189 (1.6%) | 21 (3.1%) | 159 (1.5%) | 9 (2.1%) | 0.005 |
| Mild Liver Disease | Yes | 1275 (10.8%) | 205 (30.6%) | 1017 (9.5%) | 53 (12.6%) | < 0.001 |
| Moderate or Severe Liver Disease | Yes | 88 (0.7%) | 39 (5.8%) | 45 (0.4%) | 4 (1.0%) | < 0.001 |
| Diabetes without Complication | Yes | 1549 (13.1%) | 128 (19.1%) | 1356 (12.6%) | 65 (15.5%) | < 0.001 |
| Diabetes with Complication | Yes | 452 (3.8%) | 53 (7.9%) | 372 (3.5%) | 27 (6.4%) | < 0.001 |
| Hemiplegia or Paraplegia | Yes | 119 (1.0%) | 8 (1.2%) | 100 (0.9%) | 11 (2.6%) | 0.005 |
| Renal Disease | Yes | 345 (2.9%) | 33 (4.9%) | 288 (2.7%) | 24 (5.7%) | < 0.001 |
| Cancer (Any Malignancy) | Yes | 322 (2.7%) | 27 (4.0%) | 276 (2.6%) | 19 (4.5%) | 0.007 |
| Metastatic Solid Tumor | Yes | 62 (0.5%) | 7 (1.0%) | 51 (0.5%) | 4 (1.0%) | 0.048 |

To examine the treatment effects on patients with alcohol abuse/dependence,

Table 1.6: Causal association between treatment and outcome for patients diagnosed with alcohol abuse/dependence

| | | Observed outcomes | | | Estimated outcomes with IPTW | | |
|---|---|---|---|---|---|---|---|
| Treatment group | sample size | Alcohol related organ disease | Alcohol abuse/dependence | Remission | Alcohol related organ disease | Alcohol abuse/dependence | Remission |
| Overall | 11819 | 5.7% | 90.8% | 3.6% | | | |
| No treatment | 5848 | 8.2% | 88.1% | 3.7% | 7.3% | 89.0% | 3.7% |
| Behaviour therapy only | 5053 | 3.1% | 93.6% | 3.3% | 3.5% | 92.8% | 3.7% |
| Medication only | 352 | 5.1% | 90.3% | 4.5% | 5.7% | 89.6% | 4.7% |
| Combination | 566 | 2.7% | 94.0% | 3.4% | 2.6% | 94.4% | 3.0% |

Table 1.7: Estimated treatment effect with and without IPTW.

| | MS-OLRM without IPTW | | | MS-OLRM with IPTW | | |
|---|---|---|---|---|---|---|
| Comparison groups | OR | 95% CI | P-Value | OR | 95% CI | P-Value |
| Behaviour therapy only | 1.738 | (1.521, 1.988) | $< 0.001$ | 1.588 | (1.450, 1.740) | $<0.001$ |
| Medication only | 1.581 | (1.083, 2.318) | 0.019 | 1.380 | (1.260, 1.512) | $<0.001$ |
| Combination | 1.841 | (1.359, 2.496) | $<0.001$ | 1.623 | (1.481, 1.779) | $<0.001$ |
| $\theta_1$ (Behaviour therapy only vs Control) | 0.523 | (0.517, 0.528) | $<0.001$ | 0.518 | (0.515, 0.522) | $<0.001$ |
| $\theta_2$ (Medication only vs Control) | 0.519 | (0.504, 0.534) | 0.015 | 0.513 | (0.509, 0.516) | $<0.001$ |
| $\theta_3$ (Combination vs Control) | 0.525 | (0.513, 0.537) | $<0.001$ | 0.519 | (0.515, 0.523) | $<0.001$ |
| $\theta_{21}$ (Medication only vs Behaviour therapy only) | 0.496 | (0.481, 0.512) | 0.629 | 0.495 | (0.492, 0.499) | 0.009 |
| $\theta_{31}$ (Combination vs Behaviour therapy only) | 0.502 | (0.490, 0.514) | 0.711 | 0.501 | (0.497, 0.504) | 0.606 |
| $\theta_{32}$ (Combination vs Medication only) | 0.506 | (0.487, 0.525) | 0.526 | 0.506 | (0.502, 0.509) | 0.001 |

we obtain patients diagnosed with alcohol abuse/dependence but without alcohol related organ diseases. We consider four treatment groups: medication only group, behaviour therapy only group, both medication and behaviour therapy, and none of these (control group). The outcome variable is an ordinal variable with outcome levels as progressing to alcohol related organ diseases, remaining as alcohol abuse/dependence, and remission from alcohol abuse/dependence. We obtain a study cohort of patients diagnosed with alcohol abuse/dependence who did not have alcoholic related organ diseases or remission within one-year prior of alcohol abuse/dependence diagnosis or within 6 months post alcohol abuse/dependence diagnosis. We use the diagnosis codes and procedure codes to capture treatment utilization with 6-month post alcohol abuse/dependence diagnosis, and we capture the outcomes from 7 to 18 months post diagnosis of alcohol abuse/dependence. The study cohort is formed by the following inclusion/exclusion criteria: (1) exclude patients who did not have any diagnosis of alcohol abuse or dependence; (2) exclude patients who had diagnosis of alcohol related organ diseases or remission prior the first diagnosis of alcohol abuse/dependence within one year to exclude patients with history of organ dis-

eases or remission (Smothers et al., 2004); (3) exclude patients with age $< 14$ or age $\geq 65$ years old; (4) exclude patients who did not have medical claims within 6 months of diagnosis of alcohol abuse/dependence (Rogal et al., 2020); (5) exclude patients who did not have medical claims from 7-18 months since diagnosis of alcohol abuse/dependence (Weisner et al., 2003). By following the inclusion/exclusion criteria, we form a cohort of 11819 patients to evaluate treatment effect. The alcohol dependence/abuse is captured from the International Classification of Disease nine and tenth revision (ICD-9 and ICD-10) diagnosis codes from Medicaid claims data (see codes in Table S1.1). Behaviour therapy is defined by the Healthcare Common Procedure System (HCPCS) and the Current Procedural Terminology (CPT) codes (see Table S1.2), and medication treatment includes naltrexone, acamprosate, topiramate, and disulfiram, which are obtained from Medicaid pharmacy claim data using national drug codes (NDC) (see Table S1.3). The outcome variable on whether a patient developed organ diseases, remained in alcohol abuse/dependence or remitted are obtained from ICD-9/ICD-10 diagnosis codes (see Table S1.1).

Let assume that the order of outcome is alcohol related organ diseases, alcohol abuse/dependence, and remission. The confounding variables are demographic variables (i.e., age, gender, race and ethnicity, metro versus non-metro) and Charlson comorbidities (Hall et al., 2004; Quan et al., 2005; Hu et al., 2022). The associations between confounding variables and treatment groups are presented in Table 1.4, and the association between confounding variables and outcomes are presented in Table 1.5. Based on the two tables, it is clear that these variables are both associated with treatment and outcome, thus need to be adjusted to evaluate treatment effect. We evaluated the effects of different treatments with the proposed method with and without IPTW.

We summarize the relationship between the outcomes and the treatment groups using MS-OLRM with and without IPTW (see Table 1.6). The distribution of covari-

ates in treatment groups are shown in Figure 1.5, which indicates that all covariates are balanced with absolute mean difference below the cut point 0.1. Table 1.5 shows the results from MS-OLRM with/without IPTW. Based on the simulation studies as well as the theoretical argument, we draw conclusions based on the MS-OLRM with IPTW method: (1) comparing with control group, the odds of getting better outcome (alcohol abuse/dependence or remission versus organ diseases) are 1.59 in behaviour therapy only group, 1.38 in medication only group, and 1.62 in combination group; (2) $\theta_1$, $\theta_2$, and $\theta_3$ are greater than 0.5, and their p-values are less than 0.05 with IPTW, indicating that treatment groups have stochastic superiority than control group, and patients in treatment groups are more likely to get a better outcome; (3) $\theta_{21} = 0.495$ (p-value = 0.009) indicates that medication only is inferior than behaviour therapy; $\theta_{31} = 0.501$ (p-value = 0.606) indicates no significant difference between combination therapy and behaviour therapy; $\theta_{32} = 0.506$ (p-value = 0.001) indicates that combination therapy is superior than medication therapy.

## 1.5 Conclusion and discussion

In this study, we used stochastic superiority score to assess the treatment effect on an ordinal outcome, which is estimated using MS-OLRM with IPTW. Further we develop the test procedure to test whether the superiority score effect between treatment and outcome is statistically significantly different from 0.5. The simulation results show that the stochastic superiority score estimated using MS-OLRM with IPTW controls the type I error rate and has an increased power as the effect size increases. The case study shows that the beneficial effect of behaviour therapy or combination therapy on treating patients with alcohol abuse or dependence, when comparing with control or medication only.

The valid inference for the stochastic superiority score relies on the four underlying assumptions: exchangeability, consistency, positivity, and correct specification

of generalized propensity scores. The positivity requires that the probability for each treatment (given all confounding variables) is positive, which can be illustrated by the predicted density plots with zero probability at 0 and 1. The correct specification of generalized propensity scores could be relaxed by developing a doubly robust approach which uses both generalized propensity scores model and outcome model. The doubly robust approach will be investigated in our future research.

# CHAPTER 2

# DOUBLY ROBUST SINGLE INDEX MODEL FOR SELECTING OPTIMAL TREATMENT ON SURVIVAL OUTCOMES IN OBSERVATIONAL STUDIES

## 2.1 Introduction

Personalized medicine has become an area of interest in modern biomedical research (Guo et al., 2021). Personalized medicine involves selecting appropriate and optimal therapies based on the context of a patient's demographics, comorbidity, genetic content or other molecular or cellular analysis (Jain, 2002). A treatment that works for a majority patients may fail to work for a subgroup of patients with specific characteristics (Pan and Zhao, 2021). For example, olanzapine was considered effective, safe and well tolerated treatment for schizophrenia. However, Ishigooka et al. (2000) reported that 79.5% of patients diagnosed with schizophrenia had improvement while the remaining patients did not have any improvement, or even became worse when treated with olanzpine (Qian and Murphy, 2011). Personalized treatment that selects an optimal treatment for a patient with a particular set of characteristics has received much attention (VanderWeele et al., 2019).

The important component of personalized medicine is to estimate individualized treatment rules (IRTs) and to select an optimal treatment regime. Randomized experiments could stratify patients into different subgroups by their specific characteristics, and randomly assign patients in each subgroup to different treatments.

Kaplan-Meier curves and/or Cox proportional model are applied to evaluate the optimal treatment (Murphy, 2002; Jiang et al., 2017). Removing heterogeneity and keeping homogeneity are the common strategies to find which sub-population can get the most benefit from a particular treatment. However, the way of controlling confounding variables in a randomized study is unlikely to work in an observational study (Yao and Tarpey, 2022). To select the optimal treatment in an observational study, the primary methods are Q-learning and A-learning methods (Zhao et al., 2009, 2011). Q-learning is based on outcome regression models (Bather, 2000), while A-learning requires only posited models for contrast function between treatment and control and the propensity score model for treatment assignment (Schulte et al., 2014). Q-learning is more efficient in estimating the parameters that defines the optimal regime when the outcome regression models are correctly specified, while A-learning trends to offer robust estimation of parameters when the outcome model is mis-specified (Schulte et al., 2014). In addition to Q- and A- learning methods, machine-learning methods have been used to directly estimate optimal treatment regime (Chen and Tsiatis, 2001). Non-parametric methods often provide more flexible and less biased estimation for the contrast function, however, non-parametric models are less interpretable in practice (Song et al., 2017). On the contrary, the parameters from parametric models are usually interpretable. However, the parametric models are often prone to model mis-specification. In this project, we propose to use generalized partial linear single-index model (GPLSIM) as outcome model to estimate the contrast functions. The GPLSIM model is a bridge between parametric model and non-parametric model, its parameters could be interpretable and the model self is quite flexible.

Time-to-event data is very common in medical research. In this project, we propose a semi-parametric accelerated failure time (AFT) mean model which composes an outcome component and contrast component (Wallace et al., 2019; Chatton et al., 2020; Simoneau et al., 2020). The contrast component is modeled using

GPLSIM model. We develop a doubly robust method to estimate the optimal treatment regimes. The proposed method is doubly robust in estimating the contrast component if either the propensity score model or the outcome model under control is correctly specified. The structure of this paper is organized as follows. Section 2 introduces the proposed structural accelerated failure time (AFT) mean model and estimation procedure. Section 3 presents the simulation studies to examine the performance of proposed method. Section 4 applies the proposed method to the National Health and Nutrition Examination Survey (NHANES) data to identify the proper level of physical activity for a patient's survival. Section 5 draws conclusions and provides discussions.

## 2.2 Proposed statistical method

Let us consider the outcome variable as a time-to-event outcome, which could be right-censored. Let denote $(\mathbf{X}, \mathbf{Z}, A, Tr, \triangle)$ as the observed variables for each subject. $\mathbf{X}$ denotes a vector of $p$ continuous covariates (i.e., $\mathbf{X} \in R^p$), $\mathbf{Z}$ denotes a vector of $q$ binary variables (i.e., $\mathbf{Z} \in R^q$). $A$ denotes the treatment assigned among $K+1$ treatment choices (i.e., $A \in \{0, 1, \cdots, K\}$). $T_r$ denotes the time to event or time of censoring, which is identified by $\triangle$ taking a value of 1 if $T_r$ is an event time and 0 if $Tr$ is a censoring time.

Following the literature, we assume that $(Tr, \triangle)$ are from two random variables: the survival time $T$ and the censoring time $C$. The observed time $Tr = min\{T, C\}$, and $\triangle = I_{\{T<C\}}$. To obtain the contrast functions, we propose using the following AFT mean model (Wang et al., 2010):

$$log(T) = h_0(X, Z) + \sum_{a=1}^{K} I_{\{A=a\}} \left[ Z^T \theta^{(a)} + g_a(X^T \beta^{(a)}) \right] + \epsilon$$

Here $h_0(X, Z)$ captures the outcome under control model, and the second term

$Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})$ does capture the difference between treatment $a$ and control. To illustrate these point, we introduce the potential outcomes. Let denote the survival time as $T^{(a)}$ if the subject would have received treatment $a$, $a \in \{0, 1, \cdots, K\}$. Thus, we have $K+1$ potential survival times which correspond to the $K+1$ treatment choices. Let denote $Y^{(a)} = log\ T^{(a)}$ for $a \in \{0, 1, \cdots, K\}$. Following the literature in causal inference, we assume that exchangeability and consistency hold. Exchangeability (no unmeasured confounding) assumes that given confound variables $(X, Z)$, a potential outcome is independent of treatment assignment given confound variables. That is, $Y^{(a)} \perp A|(X, Z)$ for $a \in (0, 1, \cdots, K)$. Consistency indicates that for each subject, the observed outcome is the potential outcome corresponding to the treatment the subject receives, that is, $Y = \sum_{a=0}^{K} I_{\{A=a\}} Y^{(a)}$. With the notation that $Y = log\ T$, and $Y^{(a)} = log\ T^{(a)}$, we can rewrite the proposed model as:

$$Y = h_0(X, Z) + \sum_{a=1}^{K} I_{\{A=a\}} \left[ Z^T\theta^{(a)} + g_a(X^T\beta^{(a)}) \right] + \epsilon$$

Based on the assumptions of exchangeability and consistency, we have:

$$E(Y|A = 0, X, Z) = E(Y^{(0)}|A = 0, X, Z) = E(Y^{(0)}|X, Z) = h_0(X, Z)$$

Thus, $h_0(X, Z)$ captures the response profile when a patient with covariates $(\mathbf{X}, \mathbf{Z})$ is in control group. Based on the assumptions of exchangeability and consistency, we also have:

$$E(Y|A = a, X, Z) = E(Y^{(a)}|A = a, X, Z) = E(Y^{(a)}|X, Z)$$

Note that $E(Y|A = a, X, Z) = h_0(X, Z) + Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})$. Thus, the average

treatment effects in subjects with covariates $(\mathbf{X}, \mathbf{Z})$ can be expressed as:

$$E(Y^{(a)}|X,Z) - E(Y^{(0)}|X,Z) = Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})$$

Hence, $Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})$ is a contrast function between treatment $a$ and control. $g_a(.)$ is an unknown function which is estimated using B-splines. The contrast function $Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})$ is described as a generalized partially linear single-index model (GPLSIM) (Carroll et al., 1997). This GPLSIM captures the benefits from treatment. The optimal treatment (say $a^*$) can be selected as the treatment which maximizes the contrast functions (Geng et al., 2015). That is

$$a^* = argmax_{a\in\{0,1,\cdots,K\}} \left\{0, Z^T\theta^{(1)} + g_1(X^T\beta^{(1)}), \cdots, Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})\right\}$$

Note that:

$$\begin{aligned}
E(Y|X,Z) &= E\left\{\left(I_{\{A=0\}}Y^{(0)} + \sum_{a=1}^{K} I_{\{A=a\}}Y^{(a)}\right)|X,Z\right\} \\
&= E\left(I_{\{A=0\}}Y^{(0)}|X,Z\right) + E\left\{\sum_{a=1}^{K} I_{\{A=a\}}Y^{(a)}|X,Z\right\} \\
&= E(Y^{(0)}|X,Z)\pi_0(X,Z) + \sum_{a=1}^{K} E(Y^{(a)}|X,Z)\pi_a(X,Z) \\
&= h_0(X,Z)\pi_0(X,Z) + \sum_{a=1}^{K}\left[h_0(X,Z) + Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})\right]\pi_a(X,Z) \\
&= h_0(X,Z) + \sum_{a=1}^{K}\left[Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})\right]\pi_a(X,Z) \\
&= h(X,Z)
\end{aligned}$$

Here, $\pi_0(X,Z) = Pr(A=0|X,Z)$ and $\pi_a(X,Z) = Pr(A=a|X,Z)$ $(a = 1,\cdots,K)$ are the generalized propensity scores, which satisfy $\pi_0(X,Z) + \sum_{a=1}^{K}\pi_a(X,Z) = 1$. With the equation $h_0(X,Z) = h(X,Z) - \sum_{a=1}^{K}\left[Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})\right]\pi_a(X,Z)$, we

can rewrite the semiparametric AFT mean model as:

$$Y = h(X, Z) + \sum_{a=1}^{K} \left[ I_{\{A=a\}} - \pi_a(X, Z) \right] \left[ Z^T \theta^{(a)} + g_a(X^T \beta^{(a)}) \right] + \epsilon, \tag{2.1}$$

Note that $Y = log\ T$, where $T$ is the time to event. $T$ is not observed when the time to event is censored, which is indicated by $\triangle = 0$. Let assume that the censoring time and event time are independent. Let denote $G(t) = Pr(C > t)$. That is, $G(t)$ is the probability of censoring time being greater than $t$. Let $(X_i, Z_i, A_i, Tr_i, \triangle_i)$ $(i = 1, \cdots, n)$ denote the observed data, $G(t)$ could be estimated by using Kaplan-Meier curve using $(T_{ri}, \triangle_i)$ $(i = 1, \cdots, n)$ (Sugihara, 2010). The estimation of the parameters and link functions are based on the observed uncensored data but with each uncensored data point being weighted by the inverse of $G(Tr_i)$. That is,

$$L(\beta, \theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\triangle_i}{G(Tr_i)} \left\{ Y_i - h(X_i, Z_i) - \sum_{a=1}^{K} \left[ I_{\{A_i=a\}} - \pi_a(X_i, Z_i) \right] \left[ Z_i^T \theta^{(a)} + g_a(X_i^T \beta^{(a)}) \right] \right\}^2 \tag{2.2}$$

For the unknown function $g_a(.)$, we suppose $g_a(.)$ can be represented as a linear combination of B-spline basis functions. That is, $g_a(.)$ can be described as $g_a(X^T \beta^{(a)}) = B_a(X^T \beta^{(a)})^T \gamma^{(a)}$, where $B_a(.)$ is the spline basis with $k$ knots, and $\gamma^{(a)}$ is the spline coefficient vector with $k + 4$ dimensions (Boehm, 1980). To estimate $\theta^{(a)}$, $\beta^{(a)}$, and the link function $g_a$, we take the first derivative of the loss function in equation (2.2) with respect to $\theta$, $\beta$, and $\gamma$ in equation (2.2). Here, $\theta = (\theta^{(1)T}, \cdots, \theta^{(K)T})^T$, $\beta = (\beta^{(1)T}, \cdots, \beta^{(K)T})^T$, and $\gamma = (\gamma^{(1)T}, \cdots, \gamma^{(K)T})^T$. By setting $\frac{\partial l}{\partial \theta} = 0$, we have

$$\begin{bmatrix} \frac{\partial l}{\partial \theta^{(1)}} \\ \frac{\partial l}{\partial \theta^{(2)}} \\ \vdots \\ \frac{\partial l}{\partial \theta^{(K)}} \end{bmatrix} = \frac{2}{n} \nu_\theta - \frac{2}{n} D_\theta \begin{bmatrix} \theta^{(1)} \\ \theta^{(2)} \\ \vdots \\ \theta^{(K)} \end{bmatrix}, \tag{2.3}$$

where

$$\nu_\theta = \begin{bmatrix} \sum_{i=1}^n \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=1\}} - \pi_1(X_i, Z_i)\right] Z_i \left[Y_i - h_0(X_i, Z_i) - \sum_{a=1}^K I_{\{A_i=a\}} g_a(X_i^T \beta^{(a)})\right] \\ \sum_{i=1}^n \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=2\}} - \pi_2(X_i, Z_i)\right] Z_i \left[Y_i - h_0(X_i, Z_i) - \sum_{a=1}^K I_{\{A_i=a\}} g_a(X_i^T \beta^{(a)})\right] \\ \vdots \\ \sum_{i=1}^n \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=K\}} - \pi_K(X_i, Z_i)\right] Z_i \left[Y_i - h_0(X_i, Z_i) - \sum_{a=1}^K I_{\{A_i=a\}} g_a(X_i^T \beta^{(a)})\right] \end{bmatrix}$$

and

$$D_\theta = \sum_{i=1}^n \begin{bmatrix} \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=1\}} - \pi_1(X_i, Z_i)\right] \left[I_{\{A_i=1\}}, I_{\{A_i=2\}}, \cdots, I_{\{A_i=K\}}\right] \bigotimes Z_i Z_i^T \\ \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=2\}} - \pi_2(X_i, Z_i)\right] \left[I_{\{A_i=1\}}, I_{\{A_i=2\}}, \cdots, I_{\{A_i=K\}}\right] \bigotimes Z_i Z_i^T \\ \vdots \\ \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=K\}} - \pi_K(X_i, Z_i)\right] \left[I_{\{A_i=1\}}, I_{\{A_i=2\}}, \cdots, I_{\{A_i=K\}}\right] \bigotimes Z_i Z_i^T \end{bmatrix}.$$

By setting $\frac{\partial l}{\partial \gamma} = 0$, we have

$$\begin{bmatrix} \frac{\partial l}{\partial \gamma^{(1)}} \\ \frac{\partial l}{\partial \gamma^{(2)}} \\ \vdots \\ \frac{\partial l}{\partial \gamma^{(K)}} \end{bmatrix} = \frac{2}{n}\nu_\gamma - \frac{2}{n} D_\gamma \begin{bmatrix} \gamma^{(1)} \\ \gamma^{(2)} \\ \vdots \\ \gamma^{(K)} \end{bmatrix}, \tag{2.4}$$

where

$$\nu_\gamma = \begin{bmatrix} \sum_{i=1}^n \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=1\}} - \pi_1(X_i, Z_i)\right] B_1(X_i^T \beta^{(1)}) \left[Y_i - h_0(X_i, Z_i) - \sum_{a=1}^K I_{\{A_i=a\}} Z_i^T \theta^{(a)}\right] \\ \sum_{i=1}^n \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=2\}} - \pi_2(X_i, Z_i)\right] B_2(X_i^T \beta^{(2)}) \left[Y_i - h_0(X_i, Z_i) - \sum_{a=1}^K I_{\{A_i=a\}} Z_i^T \theta^{(a)}\right] \\ \vdots \\ \sum_{i=1}^n \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=K\}} - \pi_K(X_i, Z_i)\right] B_K(X_i^T \beta^{(K)}) \left[Y_i - h_0(X_i, Z_i) - \sum_{a=1}^K I_{\{A_i=a\}} Z_i^T \theta^{(a)}\right] \end{bmatrix}$$

and

$$D_\gamma = \sum_{i=1}^{n} \begin{bmatrix} \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=1\}} - \pi_1(X_i, Z_i)\right] \left[I_{\{A_i=1\}}, I_{\{A_i=2\}}, \cdots, I_{\{A_i=K\}}\right] \bigotimes B_1(X_i^T\beta^{(1)}) B_1(X_i^T\beta^{(1)})^T \\ \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=2\}} - \pi_2(X_i, Z_i)\right] \left[I_{\{A_i=1\}}, I_{\{A_i=2\}}, \cdots, I_{\{A_i=K\}}\right] \bigotimes B_2(X_i^T\beta^{(2)}) B_2(X_i^T\beta^{(2)})^T \\ \vdots \\ \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=K\}} - \pi_K(X_i, Z_i)\right] \left[I_{\{A_i=1\}}, I_{\{A_i=2\}}, \cdots, I_{\{A_i=K\}}\right] \bigotimes B_K(X_i^T\beta^{(1)}) B_K(X_i^T\beta^{(1)})^T \end{bmatrix}.$$

By setting $\frac{\partial l}{\partial \beta} = 0$, we have

$$\begin{bmatrix} \frac{\partial l}{\partial \beta^{(1)}} \\ \frac{\partial l}{\partial \beta^{(2)}} \\ \vdots \\ \frac{\partial l}{\partial \beta^{(K)}} \end{bmatrix} = \frac{2}{n}\nu_\beta - \frac{2}{n} D_\beta \begin{bmatrix} \beta^{(1)} \\ \beta^{(2)} \\ \vdots \\ \beta^{(K)} \end{bmatrix}, \tag{2.5}$$

where

$$D_\beta = \sum_{i=1}^{n} \begin{bmatrix} \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=1\}} - \pi_1(X_i, Z_i)\right] g_1'(X_i^T\beta^{(1)}) \left[I_{\{A_i=1\}} g_1'(X_i^T\beta_{old}^{(1)}), \cdots, I_{\{A_i=K\}} g_K'(X_i^T\beta_{old}^{(K)})\right] \bigotimes X_i X_i^T \\ \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=2\}} - \pi_2(X_i, Z_i)\right] g_2'(X_i^T\beta^{(2)}) \left[I_{\{A_i=1\}} g_1'(X_i^T\beta_{old}^{(1)}), \cdots, I_{\{A_i=K\}} g_K'(X_i^T\beta_{old}^{(K)})\right] \bigotimes X_i X_i^T \\ \vdots \\ \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=K\}} - \pi_K(X_i, Z_i)\right] g_K'(X_i^T\beta^{(K)}) \left[I_{\{A_i=1\}} g_1'(X_i^T\beta_{old}^{(1)}), \cdots, I_{\{A_i=K\}} g_K'(X_i^T\beta_{old}^{(K)})\right] \bigotimes X_i X_i^T \end{bmatrix},$$

and

$$\nu_\beta = \begin{bmatrix} \sum_{i=1}^n \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=1\}} - \pi_1(X_i, Z_i)\right] g_1'(X_i^T\beta_{old}^{(1)}) RES^* X_i \\ \sum_{i=1}^n \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=2\}} - \pi_2(X_i, Z_i)\right] g_2'(X_i^T\beta_{old}^{(2)}) RES^* X_i \\ \vdots \\ \sum_{i=1}^n \frac{\triangle_i}{G(T_i)} \left[I_{\{A_i=k\}} - \pi_k(X_i, Z_i)\right] g_k'(X_i^T\beta_{old}^{(k)}) RES^* X_i \end{bmatrix}$$

with

$$RES^* = Y_i - h_0(X_i, Z_i) - \sum_{a=1}^{K} I_{\{A_i=a\}} \left[Z_i^T\theta^{(a)} + g_a(X_i^T\beta_{old}^{(a)}) - g_a'(X_i^T\beta_{old}^{(a)}) X_i^T\beta_{old}^{(a)})\right].$$

**Remark 2.2.1.** *Note that $\beta^{(a)}$ in the contrast function $Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})$ is not in the linear form of $\beta^{(a)}$. To solve for $\beta^{(a)}$, one needs to use the first order Taylor's expansion and the Newton-Raphson iteration method. Based on the first order Taylor's expansion,*

$$g_a(X^T\beta^{(a)}) \approx g_a(X^T\beta^{(a)}_{old}) + g_a(X^T\beta^{(a)}_{old})X^T(\beta^{(a)} - \beta^{(a)}_{old}) \tag{2.6}$$

*The derivation for $\beta^{(a)}$ is obtained from replacing $g_a(X^T\beta^{(a)})$ by the right-hand side in equation (2.6).*

**Remark 2.2.2.** *For the identifiability, in the single-index model $g_a(X^T\beta^{(a)})$, we restrict $\|\beta^{(a)}\|^2 = 1$ and $\beta^{(a)}_1 > 0$ (Zhang et al., 2010). This is implemented by standardizing $\beta^{(a)}$ first and then times $sign(\beta^{(a)}_1)$.*

To estimate the contrast functions and to select the optimal treatment, we propose the following algorithms:

Step 1: Estimate the probabilities of censoring function $G(t)$ and generalized propensity scores.

Step 2: Obtain the response profile when all subjects would have been in control group: First building an outcome model based on observations in control group, and then predicting the outcomes for all subjects regardless the treatment assignment. Thus, we obtain $\hat{h_0}(X_i, Z_i)$ for $i = 1, \cdots, n$.

Step 3: Estimate the parameters $\theta^{(a)}$, $\beta^{(a)}$ and link function $g_a$ in the contrast functions $Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})$ $(a = 1, \cdots, K)$ by the following iterations:

S0: Replace censoring probabilities, the GPS, and the response profile under control by their estimates;

S1: Obtain the initial values for parameter $\beta^{(a)}$, say $\beta^{(a)}_{old}$, by minimizing equation (2.2) with taking $g_a(.)$ as identity function;

S2: Obtain the initial values for parameters $\theta^{(a)}$ and $\gamma^{(a)}$ given the estimated resulting value $\beta_{old}^{(a)}$ from S1, denoted as $\theta_{old}^{(a)}$ and $\gamma_{old}^{(a)}$;

S3: Given $\beta_{old}^{(a)}$, $\theta_{old}^{(a)}$ and $\gamma_{old}^{(a)}$, update $\beta^{(a)}$ as $\beta_{new}^{(a)}$;

S4: Update $\theta_{new}^{(a)}$ and $\gamma_{new}^{(a)}$ with $\beta^{(a)}$ obtained from S3;

S5: Set $\theta_{new}^{(a)}$, $\beta_{new}^{(a)}$, and $\gamma_{new}^{(a)}$ as $\beta_{old}^{(a)}$, $\theta_{old}^{(a)}$ and $\gamma_{old}^{(a)}$, repeat S2 - S4 until converge.

Step 4: Given $(X, Z)$, the optimal treatment assignment is obtained by

$a^* = argmax_{a\in\{0,1,\cdots,K\}} \left\{0, Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})\right\}$.

## 2.3 Simulation studies

### 2.3.1 Simulation design

In this section, we conduct a set of simulation studies to investigate the performance of the proposed method under settings. We first set up three continuous variables, say $\boldsymbol{X} = (X_1, X_2, X_3)$, and two categorical variables, say $\boldsymbol{Z} = (Z_1, Z_2)$, as confounding variables. $\boldsymbol{X}$ follows a multivariate normal distribution with mean 0 and variance $I_{3\times 3}$. Two variables in $\boldsymbol{Z}$ follows Bernoulli distributions with probabilities 0.4 and 0.5 respectively. We consider a control group and three treatment groups ($K = 3$). The treatment assignments $A \in (0, 1, 2, 3)$ are generated from a multinomial distribution which is causally associated with confounding variables $(\boldsymbol{X}, \boldsymbol{Z})$ with the probabilities specified by the following two GPS models:

$$P(A = a|\mathbf{X}, \mathbf{Z}) = \frac{exp((\mathbf{X}, \mathbf{Z})^T\delta^{(a)})}{1 + \sum_{k=1}^{3} exp((\mathbf{X}, \mathbf{Z})^T\delta^{(k)})} \tag{2.7}$$

and

$$P(A = a|\mathbf{X}, \mathbf{Z}) = \frac{exp((\mathbf{X}, \mathbf{Z})^T\delta^{(a)})}{1 + \sum_{k=1}^{3} exp((\mathbf{X}, \mathbf{Z})^T\delta^{(k)} + (X_1^2 + X_2^2 + X_3^2)\delta_2 + (Z_1^2 + Z_2^2 + Z_3^2)\delta_3 + (X_2Z_1 + X_3Z_2)\delta_4} \tag{2.8}$$

for $a = 1, 2, 3$. $P(A = 0|\mathbf{X}, \mathbf{Z}) = 1 - \sum_{a=1}^{3} P(A = a|\mathbf{X}, \mathbf{Z})$, $\delta^{(1)} = (-0.1, 1, -1, 1)^T$, $\delta^{(2)} = (-0.1, 1, 1, -1, 1)^T$, $\delta^{(3)} = (-0.1, -1, -1, 1, -1)^T$, $\delta_2 = -0.5$, $\delta_3 = 0.8$, and $\delta_4 = -0.3$. However, to estimate the GPS, we only use the multinomial model in the form of equation (2.7). Thus, when treatment assignment is generated from equation (2.7), the fitted GPS model is correctly specified model. When treatment assignment is generated from model (2.8), the fitted model using equation (2.7) is a mis-specified GPS model.

Given $(\boldsymbol{X}, \boldsymbol{Z})$ and treatment assignment $A$, the outcome $Y$ is generated by each of the following models (Guo et al., 2021):

**Model 1**: $Y = 1 + \mathbf{X}^T\alpha_1 + \mathbf{Z}^T\alpha_2 + \sum_{a=1}^{K} I_{\{A=a\}}[\mathbf{Z}^T\theta^{(a)} + \mathbf{X}^T\beta^{(a)}] + \epsilon$,

**Model 2**: $Y = 1+\mathbf{X}^T\alpha_1(1-\mathbf{X}^T\alpha_1)+\mathbf{Z}^T\alpha_2+\sum_{a=1}^{K} I_{\{A=a\}}[\mathbf{Z}^T\theta^{(a)}+\mathbf{X}^T\beta^{(a)}(1-\mathbf{X}^T\beta^{(a)})]+\epsilon$,

**Model 3**: $Y = 1 + \mathbf{X}^T\alpha_1(1 - \mathbf{X}^T\alpha_1) + \mathbf{Z}^T\alpha_2 +$
$\sum_{a=1}^{K} I_{\{A=a\}} \left\{\mathbf{Z}^T\theta^{(a)} + \mathbf{X}^T\beta^{(a)}(1 - \mathbf{X}^T\beta^{(a)}) + sin[\pi(\mathbf{X}^T\beta^{(a)})/2]\right\} + \epsilon$,

**Model 4**: $Y = exp(\mathbf{X}^T\alpha_1) + sin[\pi(\mathbf{X}^T\alpha_3)/2] + \mathbf{Z}^T\alpha_2 +$
$\sum_{a=1}^{K} I_{\{A=a\}}[\mathbf{Z}^T\theta^{(a)} + \mathbf{X}^T\beta^{(a)}(1 - \mathbf{X}^T\beta^{(a)})] + \epsilon$.

We set $\beta^{(1)} = (1, 0.4, 0.3)^T$, $\beta^{(2)} = (1, -0.5, 0.2)^T$, $\beta^{(3)} = (1, 0.1, 0.7)^T$, $\theta^{(1)} = (-0.8, 0.6)^T$, $\theta^{(2)} = (0.2, -0.3)^T$, $\theta^{(3)} = (0.8, 0.6)^T$, $\alpha_1 = (0.1, 0.2, 0.3)^T$, $\alpha_2 = (0.5, 0.1)^T$, and $\alpha_3 = (-0.1, 0.3, 0.5)^T$. The error term $\epsilon$ follows a normal distribution with mean 0 and variance 0.1. Note that we propose to use either linear model or GPLSIM to capture outcome under control, and we use GPLSIM to capture contrast functions. When linear model is used to model the outcome under control, only Model 1 is correctly specified outcome under control model. When GPLSIM is used to model the outcome under control, Model 1 to Model 3 all have correctly specified outcome model under control, while Model 4 is a mis-specified outcome model for control group. Note that we generate treatment assignment using two different GPS models. However, we only

apply multinomial regression model to estimate GPS. When treatment assignment is generated from the multinomial regression model (equation (2.7)), the GPS model is correctly specified. When the treatment assignment is generated from the complex multinomial model (equation (2.8)), the GPS model is mis-specified. In addition, we also generate censoring observations with censoring time from a gumbel distribution with parameters $(3, c)$, where $c$ is chosen to achieve 20% censoring rates. For each outcome model, we generate data under the following three scenarios:

Scenario I: Treatment assignment $A$ is generated from the multinomial model equation (2.7), and there is no censored observations,

Scenario II: Treatment assignment $A$ is generated from the multinomial model (2.7), and there are censored observations,

Scenario III: Treatment assignment $A$ is generated from the complex multinomial model (2.8), and there are censored observations. Since we apply multinomial regression to obtain GPS, we have a mis-specified GPS model in this scenario.

We carry out simulation study with two sample sizes ($n$ = 2000 and 5000) for each outcome model under each of three scenarios related treatment assignment and censoring. For each setting, we generate 1000 data sets. For each data set, we estimate contrast function with the outcome model under control as linear model **O1** : $h_0(X, Z) = X^T\beta^* + Z^T\theta^*$ and as GPLSIM **O2** : $h_0(X, Z) = g(X^T\beta^*) + Z^T\theta^*$ respectively. The estimates for contrast functions are captured by the estimates of $\beta^{(a)}$ and $\theta^{(a)}$ in the contrast function $Z^T\theta^{(a)} + g_a(X^t\beta^{(a)})$, and the selected optimal treatment $a^* = argmax_{a\in\{0,\cdots,K\}}\{Z^T\hat{\theta}^{(a)} + g_a(X^T\hat{\beta}^{(a)})\}$.

### 2.3.2 Simulation results

We evaluate the performance of proposed method by (i) the box plots of the estimated parameters in the contrast function for each outcome model under each

scenario (see Figures 2.1 - 2.12); (ii) mean squared error (MSE) for each parameter in the contrast function (see Table 2.1 and 2.2); and (iii) the percentage of correct decision (PCD) (see Table 2.3) for the optimal treatment selection.

Figures 2.1 - 2.4 show the box plots of the estimates under four different outcome models with data generated under Scenario I without censoring and with a correctly specified GPS model. Figures 2.5 - 2.8 show the box plots of the estimates under four different outcome models in Scenario II with data generated with censoring and with a correctly specified GPS model. Figures 2.9 - 2.12 show the box plots of the estimates in scenario III with data generated with censoring and with a mis-specified GPS model. In each figure, the three rows show the box plots of the estimates in three contrast functions (i.e., treatment $a$ vs control with $a = 1, 2, 3$), and the left panels show the estimates with sample size $n = 2000$, and the right panels show the estimates with sample size $n = 5000$. The x-axis shows the parameters to be estimated, and the y-axis shows the estimated values. From these figures, we can see that (i) when the underlying outcome model is linear (Model 1) (see Figures 2.1, 2.5, and 2.9), the estimates with linear outcome model are similar to the ones with GPLSIM. When GPS is a mis-specified model, the proposed GPLSIM outcome model with CBPS performs better than multinomial GPS model (see Figure 2.9); (ii) when the outcome model is non-linear but with partial linear single-index form (Models 2 and 3), the proposed GPLSIM outcome model performs better than those with linear outcome model. Among those with GPLSIM outcome models, their estimates are similar to those with true GPSs (see Figures 2.2 - 2.3, 2.6 - 2.7, and 2.10 - 2.11); (iii) when the outcome model is not in a partial linear single-index form (Model 4), the proposed model with GPLSIM performs much better than those with linear outcome model (see Figure 2.4, 2.8, and 2.12), and the proposed model with CBPS has smaller variation (similar to those from true GPS), than those from multinomial GPS model; which is similar to use true GPS; (iv) when the GPS model is mis-specified,

the proposed method with correctly specified outcome models (Models 1 - 3) provide unbiased estimates, the results with CBPS have smaller variation (Figures 2.9 - 2.11); (v) when both GPS model and the outcome model (Model 4) are mis-specified (see Figure 2.12), the results with GPLSIM outcome model and CBPS are compariable with those with GPLSIM outcome and true GPS model. The results are less biased and have smaller variation, indicating the advantage using the proposed methods with GPLSIM as outcome model and CBPS as GPS. The MSE for $\beta^{(a)}$ presented in Table 2.1 and $\theta^{(a)}$ in Table 2.2 further support these findings.

Table 2.1: Summarized mean square errors for the estimates of $\beta^{(a)}$ in the contrast function $Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})$ for different sample sizes and different scenarios.

| | | Linear.true.ps | | | Linear.multinomial.ps | | | Linear.CBPS | | | Spline.true.ps | | | Spline.multinomial.ps | | | Spline.CBPS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sample size (n) | $\beta^{(1)}$ | $\beta^{(2)}$ | $\beta^{(3)}$ | $\beta^{(1)}$ | $\beta^{(2)}$ | $\beta^{(3)}$ | $\beta^{(1)}$ | $\beta^{(2)}$ | $\beta^{(3)}$ | $\beta^{(1)}$ | $\beta^{(2)}$ | $\beta^{(3)}$ | $\beta^{(1)}$ | $\beta^{(2)}$ | $\beta^{(3)}$ | $\beta^{(1)}$ | $\beta^{(2)}$ | $\beta^{(3)}$ |
| Scenario I: without censoring and with correct GPS model | | | | | | | | | | | | | | | | | | | |
| | 2000 | 0.008 | 0.014 | 0.009 | 0.010 | 0.013 | 0.011 | 0.009 | 0.012 | 0.009 | 0.008 | 0.014 | 0.010 | 0.010 | 0.013 | 0.011 | 0.008 | 0.012 | 0.010 |
| Model 1 | 5000 | 0.002 | 0.004 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.003 | 0.001 | 0.002 | 0.004 | 0.002 | 0.001 | 0.003 | 0.001 | 0.001 | 0.003 | 0.001 |
| | 2000 | 0.040 | 0.059 | 0.029 | 0.040 | 0.060 | 0.027 | 0.035 | 0.054 | 0.024 | 0.009 | 0.006 | 0.006 | 0.006 | 0.004 | 0.003 | 0.005 | 0.004 | 0.002 |
| Model 2 | 5000 | 0.043 | 0.062 | 0.036 | 0.043 | 0.058 | 0.034 | 0.037 | 0.057 | 0.031 | 0.001 | 0.002 | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | 0.001 | <0.001 |
| | 2000 | 0.014 | 0.046 | 0.017 | 0.017 | 0.045 | 0.017 | 0.013 | 0.043 | 0.014 | 0.001 | 0.001 | 0.001 | <0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 |
| Model 3 | 5000 | 0.017 | 0.055 | 0.021 | 0.019 | 0.057 | 0.020 | 0.017 | 0.055 | 0.019 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | 2000 | 0.064 | 0.072 | 0.058 | 0.062 | 0.062 | 0.063 | 0.052 | 0.059 | 0.044 | 0.049 | 0.072 | 0.041 | 0.050 | 0.071 | 0.037 | 0.043 | 0.065 | 0.032 |
| Model 4 | 5000 | 0.057 | 0.047 | 0.059 | 0.057 | 0.047 | 0.059 | 0.051 | 0.044 | 0.053 | 0.038 | 0.060 | 0.034 | 0.039 | 0.060 | 0.035 | 0.036 | 0.059 | 0.033 |
| Scenario II: with censoring and with correct GPS model | | | | | | | | | | | | | | | | | | | |
| | 2000 | 0.018 | 0.016 | 0.018 | 0.017 | 0.014 | 0.016 | 0.016 | 0.013 | 0.013 | 0.017 | 0.014 | 0.015 | 0.017 | 0.018 | 0.016 | 0.016 | 0.014 | 0.014 |
| Model 1 | 5000 | 0.002 | 0.005 | 0.002 | 0.002 | 0.005 | 0.002 | 0.002 | 0.004 | 0.002 | 0.002 | 0.005 | 0.003 | 0.002 | 0.005 | 0.003 | 0.002 | 0.004 | 0.002 |
| | 2000 | 0.037 | 0.058 | 0.034 | 0.039 | 0.057 | 0.031 | 0.032 | 0.053 | 0.026 | 0.006 | 0.008 | 0.004 | 0.007 | 0.010 | 0.005 | 0.009 | 0.009 | 0.004 |
| Model 2 | 5000 | 0.025 | 0.042 | 0.023 | 0.026 | 0.042 | 0.026 | 0.023 | 0.041 | 0.025 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | 2000 | 0.020 | 0.045 | 0.016 | 0.016 | 0.041 | 0.015 | 0.014 | 0.041 | 0.014 | 0.001 | 0.001 | 0.001 | 0.002 | 0.004 | 0.002 | 0.002 | 0.004 | 0.001 |
| Model 3 | 5000 | 0.017 | 0.039 | 0.015 | 0.016 | 0.038 | 0.015 | 0.014 | 0.038 | 0.014 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | 2000 | 0.033 | 0.036 | 0.027 | 0.036 | 0.035 | 0.030 | 0.024 | 0.031 | 0.023 | 0.009 | 0.015 | 0.009 | 0.016 | 0.014 | 0.010 | 0.013 | 0.014 | 0.006 |
| Model 4 | 5000 | 0.016 | 0.016 | 0.013 | 0.016 | 0.018 | 0.014 | 0.014 | 0.020 | 0.012 | 0.004 | 0.006 | 0.002 | 0.005 | 0.006 | 0.003 | 0.004 | 0.006 | 0.002 |
| Scenario III: with censoring and with mis-specified GPS model | | | | | | | | | | | | | | | | | | | |
| | 2000 | 0.009 | 0.021 | 0.009 | 0.055 | 0.052 | 0.065 | 0.014 | 0.022 | 0.014 | 0.008 | 0.017 | 0.009 | 0.060 | 0.049 | 0.064 | 0.016 | 0.025 | 0.016 |
| Model 1 | 5000 | 0.002 | 0.007 | 0.002 | 0.046 | 0.034 | 0.041 | 0.009 | 0.010 | 0.007 | 0.002 | 0.006 | 0.002 | 0.049 | 0.034 | 0.047 | 0.010 | 0.009 | 0.008 |
| | 2000 | 0.096 | 0.089 | 0.081 | 0.399 | 0.132 | 0.301 | 0.141 | 0.091 | 0.113 | 0.004 | 0.020 | 0.003 | 0.146 | 0.056 | 0.104 | 0.024 | 0.016 | 0.013 |
| Model 2 | 5000 | 0.069 | 0.060 | 0.046 | 0.515 | 0.190 | 0.396 | 0.212 | 0.146 | 0.195 | <0.001 | 0.003 | <0.001 | 0.142 | 0.046 | 0.086 | 0.012 | 0.010 | 0.10 |
| | 2000 | 0.057 | 0.057 | 0.054 | 0.175 | 0.094 | 0.185 | 0.093 | 0.068 | 0.085 | 0.002 | 0.012 | 0.002 | 0.049 | 0.023 | 0.044 | 0.008 | 0.006 | 0.006 |
| Model 3 | 5000 | 0.039 | 0.036 | 0.043 | 0.202 | 0.122 | 0.227 | 0.114 | 0.089 | 0.116 | <0.001 | 0.001 | <0.001 | 0.038 | 0.015 | 0.041 | 0.006 | 0.003 | 0.005 |
| | 2000 | 0.085 | 0.086 | 0.064 | 0.560 | 0.267 | 0.722 | 0.274 | 0.160 | 0.267 | 0.031 | 0.049 | 0.024 | 0.484 | 0.180 | 0.464 | 0.169 | 0.097 | 0.139 |
| Model 4 | 5000 | 0.042 | 0.075 | 0.031 | 0.510 | 0.196 | 0.703 | 0.164 | 0.086 | 0.154 | 0.016 | 0.035 | 0.010 | 0.470 | 0.140 | 0.455 | 0.108 | 0.065 | 0.082 |

Table 2.2: Summarized mean square errors for the estimates of $\theta^{(a)}$ in the contrast function $Z^T\theta^{(a)} + g_a(X^T\beta^{(a)})$ for different sample sizes and different scenarios.

| | | Linear.true.ps | | | Linear.multinomial.ps | | | Linear.CBPS | | | Spline.true.ps | | | Spline.multinomial.ps | | | Spline.CBPS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sample size (n) | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ |
| Scenario I: without censoring and with correct GPS model | | | | | | | | | | | | | | | | | | | |
| Model 1 | 2000 | 0.002 | 0.015 | 0.002 | 0.002 | 0.013 | 0.003 | 0.002 | 0.010 | 0.002 | 0.002 | 0.019 | 0.003 | 0.002 | 0.015 | 0.003 | 0.002 | 0.011 | 0.002 |
| | 5000 | <0.001 | 0.001 | 0.001 | <0.001 | 0.002 | 0.001 | <0.001 | 0.002 | <0.001 | <0.001 | 0.002 | 0.001 | <0.001 | 0.002 | <0.001 | <0.001 | 0.002 | <0.001 |
| Model 2 | 2000 | 0.033 | 0.565 | 0.045 | 0.017 | 0.299 | 0.013 | 0.014 | 0.281 | 0.011 | 0.013 | 0.103 | 0.060 | 0.001 | 0.005 | 0.002 | 0.001 | 0.004 | 0.001 |
| | 5000 | 0.011 | 0.182 | 0.009 | 0.010 | 0.170 | 0.009 | 0.009 | 0.164 | 0.007 | 0.001 | 0.005 | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | 0.001 | <0.001 |
| Model 3 | 2000 | 0.015 | 0.531 | 0.015 | 0.018 | 0.537 | 0.019 | 0.013 | 0.500 | 0.016 | 0.001 | 0.014 | 0.002 | 0.001 | 0.018 | 0.002 | 0.001 | 0.017 | 0.001 |
| | 5000 | 0.009 | 0.240 | 0.012 | 0.009 | 0.243 | 0.012 | 0.009 | 0.235 | 0.011 | <0.001 | 0.001 | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | 0.001 | <0.001 |
| Model 4 | 2000 | 0.057 | 0.715 | 0.072 | 0.057 | 0.549 | 0.075 | 0.048 | 0.460 | 0.058 | 0.017 | 0.268 | 0.020 | 0.031 | 0.549 | 0.033 | 0.019 | 0.403 | 0.024 |
| | 5000 | 0.022 | 0.012 | 0.029 | 0.024 | 0.118 | 0.038 | 0.020 | 0.112 | 0.025 | 0.006 | 0.063 | 0.008 | 0.006 | 0.066 | 0.009 | 0.006 | 0.068 | 0.008 |
| Scenario II: with censoring and with correct GPS model | | | | | | | | | | | | | | | | | | | |
| Model 1 | 2000 | 0.007 | 0.049 | 0.035 | 0.009 | 0.058 | 0.060 | 0.009 | 0.057 | 0.060 | 0.006 | 0.033 | 0.016 | 0.009 | 0.067 | 0.029 | 0.008 | 0.069 | 0.030 |
| | 5000 | 0.001 | 0.006 | 0.001 | 0.001 | 0.009 | 0.001 | 0.001 | 0.008 | 0.001 | 0.001 | 0.003 | 0.001 | 0.001 | 0.004 | 0.001 | 0.001 | 0.004 | 0.001 |
| Model 2 | 2000 | 0.020 | 0.452 | 0.017 | 0.017 | 0.411 | 0.017 | 0.017 | 0.544 | 0.021 | 0.002 | 0.014 | 0.002 | 0.002 | 0.124 | 0.002 | 0.002 | 0.018 | 0.002 |
| | 5000 | 0.006 | 0.082 | 0.006 | 0.007 | 0.095 | 0.007 | 0.006 | 0.089 | 0.006 | <0.001 | 0.001 | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | 0.001 | <0.001 |
| Model 3 | 2000 | 0.026 | 0.725 | 0.020 | 0.024 | 0.665 | 0.019 | 0.021 | 0.630 | 0.016 | 0.002 | 0.005 | 0.002 | 0.002 | 0.007 | 0.002 | 0.002 | 0.006 | 0.002 |
| | 5000 | 0.009 | 0.201 | 0.010 | 0.009 | 0.195 | 0.010 | 0.008 | 0.193 | 0.010 | <0.001 | 0.001 | 0.001 | <0.001 | 0.001 | 0.001 | <0.001 | 0.001 | 0.001 |
| Model 4 | 2000 | 0.020 | 0.414 | 0.015 | 0.013 | 0.155 | 0.019 | 0.014 | 0.238 | 0.034 | 0.004 | 0.081 | 0.004 | 0.004 | 0.068 | 0.005 | 0.004 | 0.071 | 0.004 |
| | 5000 | 0.004 | 0.024 | 0.004 | 0.004 | 0.018 | 0.004 | 0.003 | 0.019 | 0.004 | 0.001 | 0.017 | 0.001 | 0.001 | 0.012 | 0.001 | 0.001 | 0.012 | 0.001 |
| Scenario III: with censoring and with mis-specified GPS model | | | | | | | | | | | | | | | | | | | |
| Model 1 | 2000 | 0.007 | 0.088 | 0.007 | 0.014 | 0.199 | 0.018 | 0.003 | 0.070 | 0.004 | 0.006 | 0.077 | 0.007 | 0.041 | 0.517 | 0.023 | 0.003 | 0.077 | 0.004 |
| | 5000 | 0.001 | 0.011 | 0.001 | 0.006 | 0.026 | 0.009 | 0.001 | 0.005 | 0.002 | 0.001 | 0.006 | 0.001 | 0.006 | 0.024 | 0.010 | 0.001 | 0.005 | 0.002 |
| Model 2 | 2000 | 0.283 | 2.636 | 0.383 | 3.436 | 65.636 | 3.637 | 0.375 | 11.485 | 0.860 | 0.007 | 0.146 | 0.009 | 0.059 | 0.520 | 0.127 | 0.009 | 0.033 | 0.011 |
| | 5000 | 0.072 | 0.371 | 0.080 | 1.158 | 21.526 | 2.618 | 0.189 | 6.666 | 0.251 | 0.001 | 0.005 | 0.001 | 0.017 | 0.226 | 0.046 | 0.001 | 0.007 | 0.002 |
| Model 3 | 2000 | 0.219 | 1.712 | 0.241 | 0.614 | 19.139 | 1.327 | 0.177 | 8.728 | 0.237 | 0.017 | 0.119 | 0.008 | 0.034 | 0.365 | 0.052 | 0.003 | 0.023 | 0.004 |
| | 5000 | 0.063 | 0.212 | 0.069 | 0.326 | 8.325 | 0.584 | 0.065 | 1.765 | 0.129 | 0.001 | 0.005 | 0.001 | 0.009 | 0.030 | 0.018 | 0.001 | 0.003 | 0.002 |
| Model 4 | 2000 | 0.523 | 10.728 | 1.162 | 17.424 | 522.079 | 29.306 | 1.371 | 17.647 | 2.180 | 0.059 | 1.056 | 0.044 | 3.092 | 97.579 | 8.688 | 0.338 | 7.853 | 0.537 |
| | 5000 | 0.052 | 0.772 | 0.046 | 2.532 | 12.111 | 6.868 | 0.048 | 0.666 | 0.071 | 0.014 | 0.145 | 0.012 | 0.692 | 4.084 | 1.340 | 0.024 | 0.502 | 0.039 |

Figure 2.1: Simulation results with outcome Model 1 without censoring.

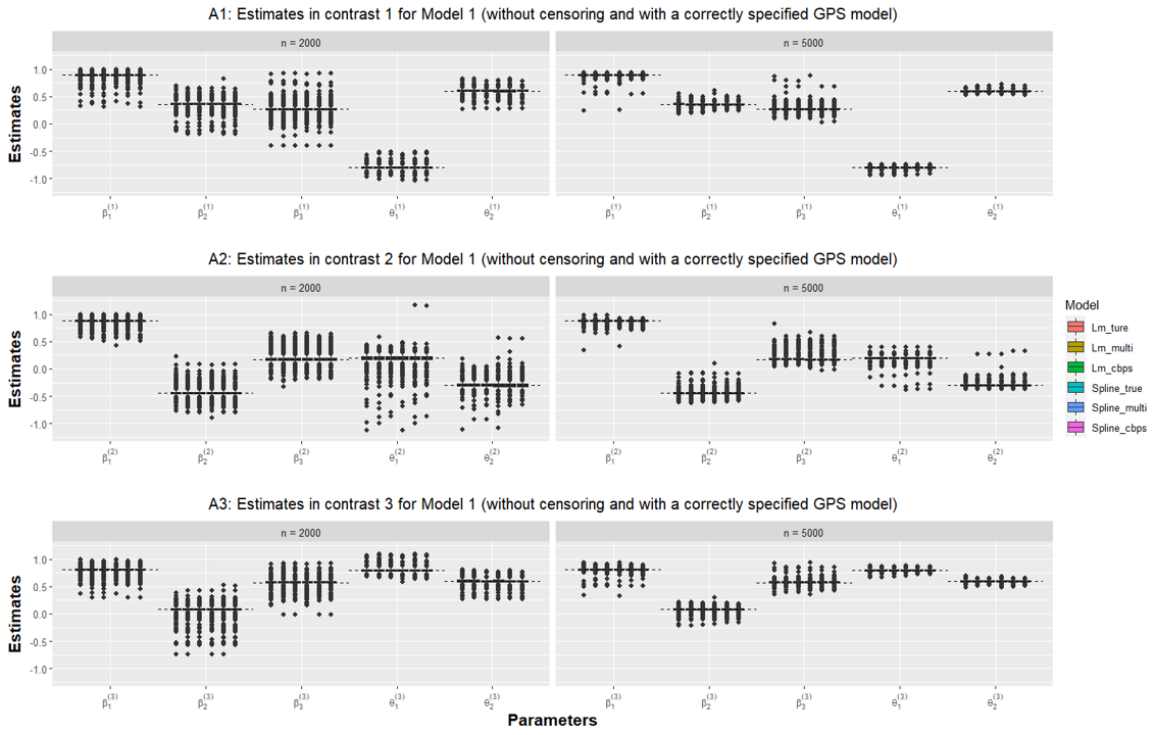


Table 2.3: Summarized percentage of correct decisions for different sample sizes and different scenarios.

| | Sample size (n) | Linear.true.ps | Linear.multinomial.ps | Linear.CBPS | Spline.true.ps | Spline.multinomial.ps | Spline.CBPS |
|---|---|---|---|---|---|---|---|
| Scenario I: without censoring and with correct GPS model | | | | | | | |
| | 2000 | 0.937 | 0.937 | 0.940 | 0.936 | 0.935 | 0.939 |
| Model 1 | 5000 | 0.973 | 0.972 | 0.973 | 0.973 | 0.972 | 0.973 |
| | 2000 | 0.782 | 0.780 | 0.791 | 0.960 | 0.961 | 0.964 |
| Model 2 | 5000 | 0.781 | 0.784 | 0.790 | 0.981 | 0.981 | 0.981 |
| | 2000 | 0.814 | 0.815 | 0.823 | 0.973 | 0.973 | 0.975 |
| Model 3 | 5000 | 0.815 | 0.812 | 0.818 | 0.988 | 0.987 | 0.988 |
| | 2000 | 0.722 | 0.734 | 0.744 | 0.773 | 0.775 | 0.785 |
| Model 4 | 5000 | 0.755 | 0.759 | 0.768 | 0.803 | 0.803 | 0.807 |
| Scenario II: with censoring and with correct GPS model | | | | | | | |
| | 2000 | 0.861 | 0.860 | 0.865 | 0.859 | 0.858 | 0.863 |
| Model 1 | 5000 | 0.889 | 0.889 | 0.891 | 0.890 | 0.890 | 0.892 |
| | 2000 | 0.769 | 0.773 | 0.783 | 0.947 | 0.944 | 0.948 |
| Model 2 | 5000 | 0.806 | 0.802 | 0.807 | 0.969 | 0.969 | 0.970 |
| | 2000 | 0.816 | 0.819 | 0.825 | 0.967 | 0.965 | 0.967 |
| Model 3 | 5000 | 0.823 | 0.825 | 0.829 | 0.972 | 0.972 | 0.972 |
| | 2000 | 0.858 | 0.860 | 0.869 | 0.908 | 0.904 | 0.907 |
| Model 4 | 5000 | 0.896 | 0.895 | 0.899 | 0.937 | 0.934 | 0.936 |
| Scenario III: with censoring and with mis-specified GPS model | | | | | | | |
| | 2000 | 0.849 | 0.798 | 0.873 | 0.853 | 0.796 | 0.870 |
| Model 1 | 5000 | 0.922 | 0.856 | 0.929 | 0.926 | 0.856 | 0.928 |
| | 2000 | 0.636 | 0.481 | 0.599 | 0.905 | 0.781 | 0.907 |
| Model 2 | 5000 | 0.687 | 0.424 | 0.526 | 0.954 | 0.806 | 0.939 |
| | 2000 | 0.682 | 0.555 | 0.659 | 0.931 | 0.851 | 0.938 |
| Model 3 | 5000 | 0.736 | 0.513 | 0.616 | 0.967 | 0.883 | 0.953 |
| | 2000 | 0.651 | 0.371 | 0.548 | 0.775 | 0.470 | 0.648 |
| Model 4 | 5000 | 0.690 | 0.373 | 0.663 | 0.807 | 0.472 | 0.721 |

Figure 2.2: Simulation results with outcome Model 2 without censoring.

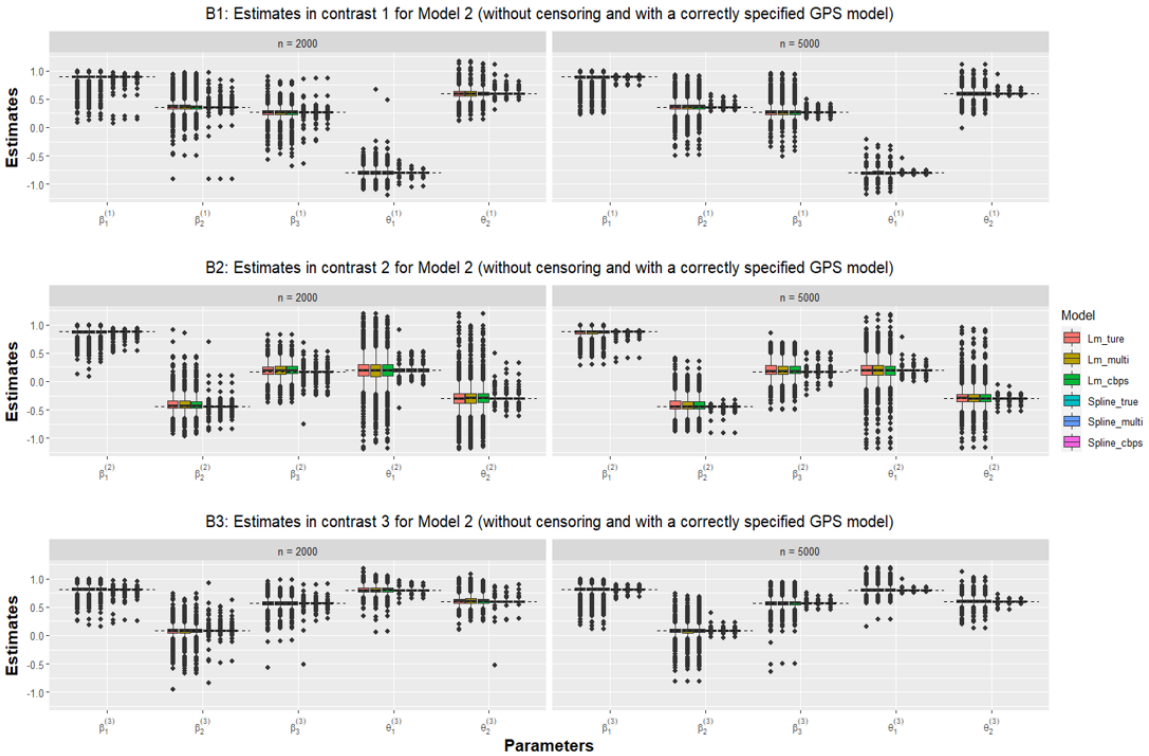


Figure 2.3: Simulation results with outcome Model 3 without censoring.

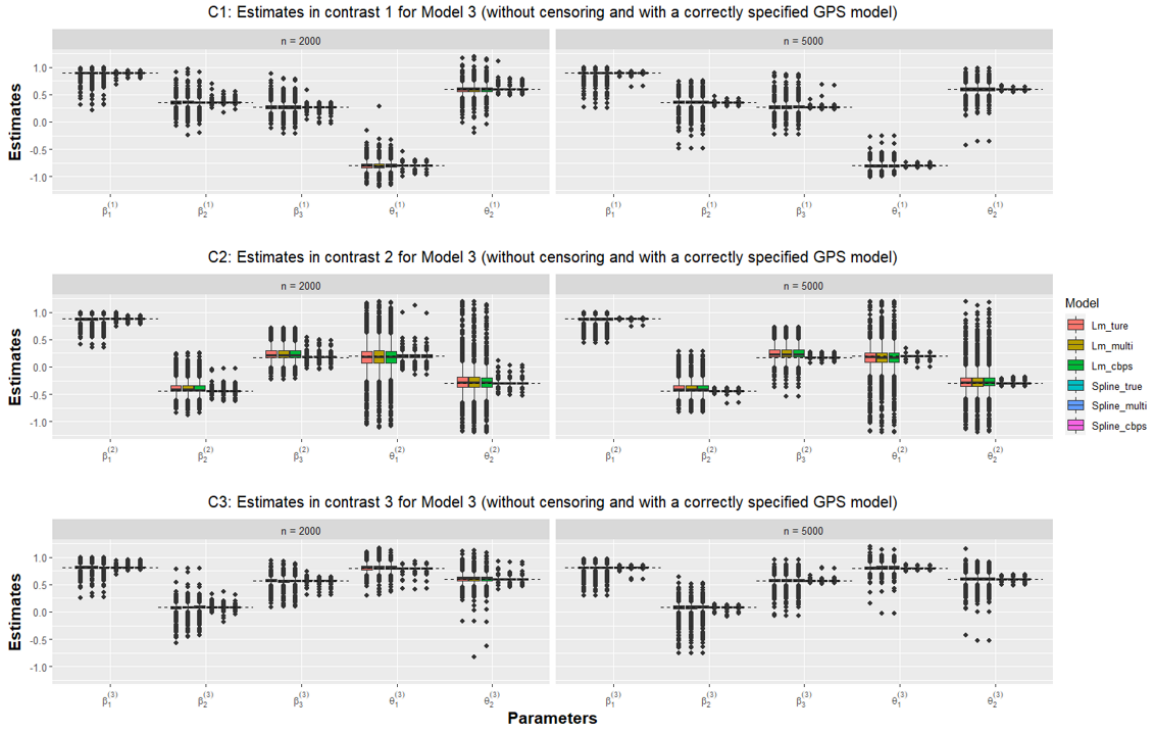



Figure 2.4: Simulation results with outcome Model 4 without censoring.

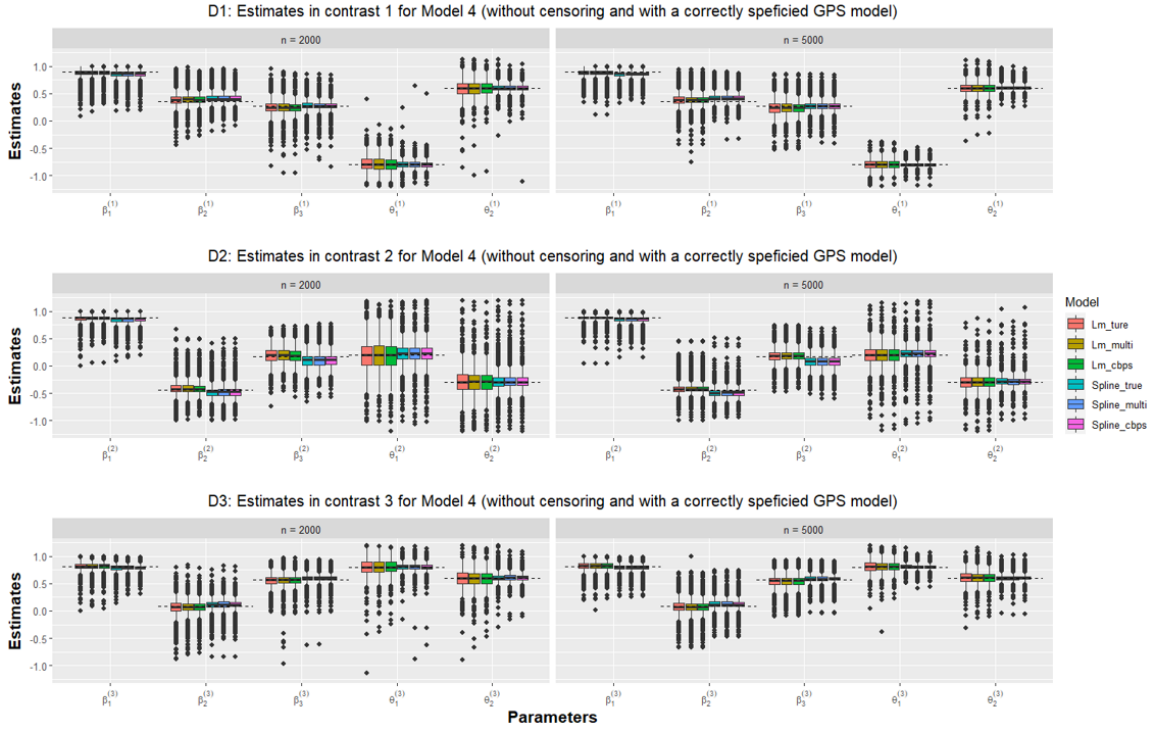


Figure 2.5: Simulation results for outcome Model 1 with censoring and with a correctly specified GPS model.

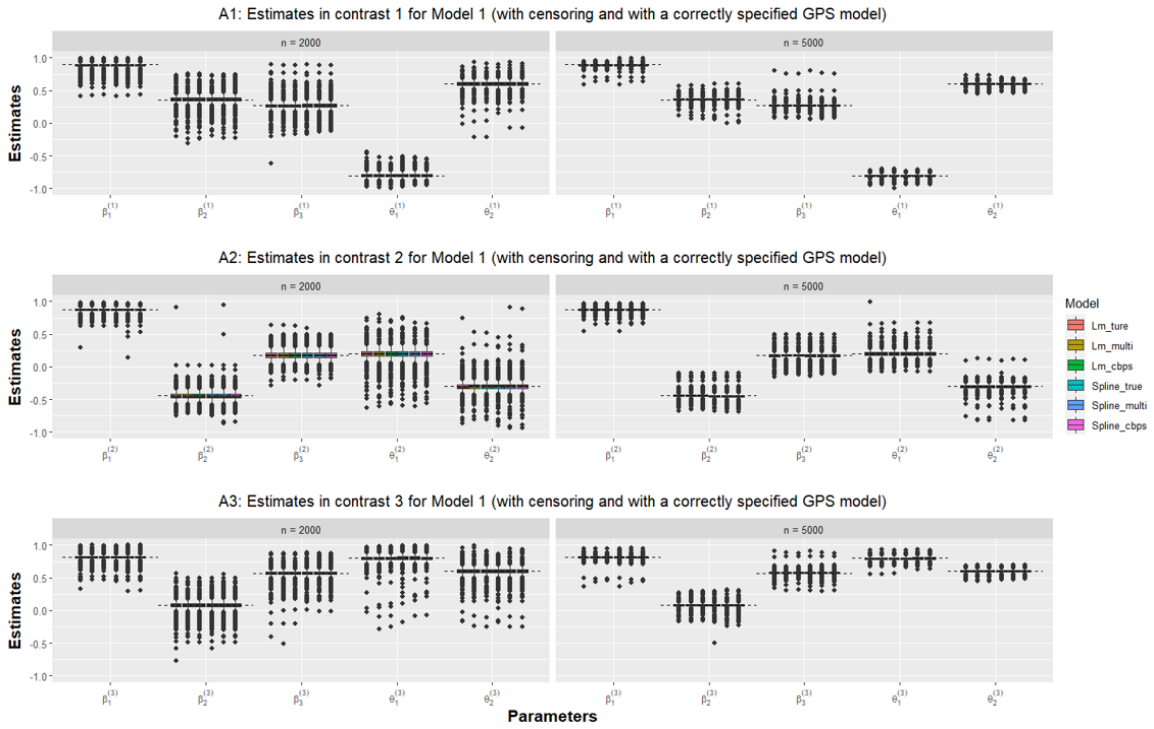



Figure 2.6: Simulation results for outcome Model 2 with censoring and with a correctly specified GPS model.

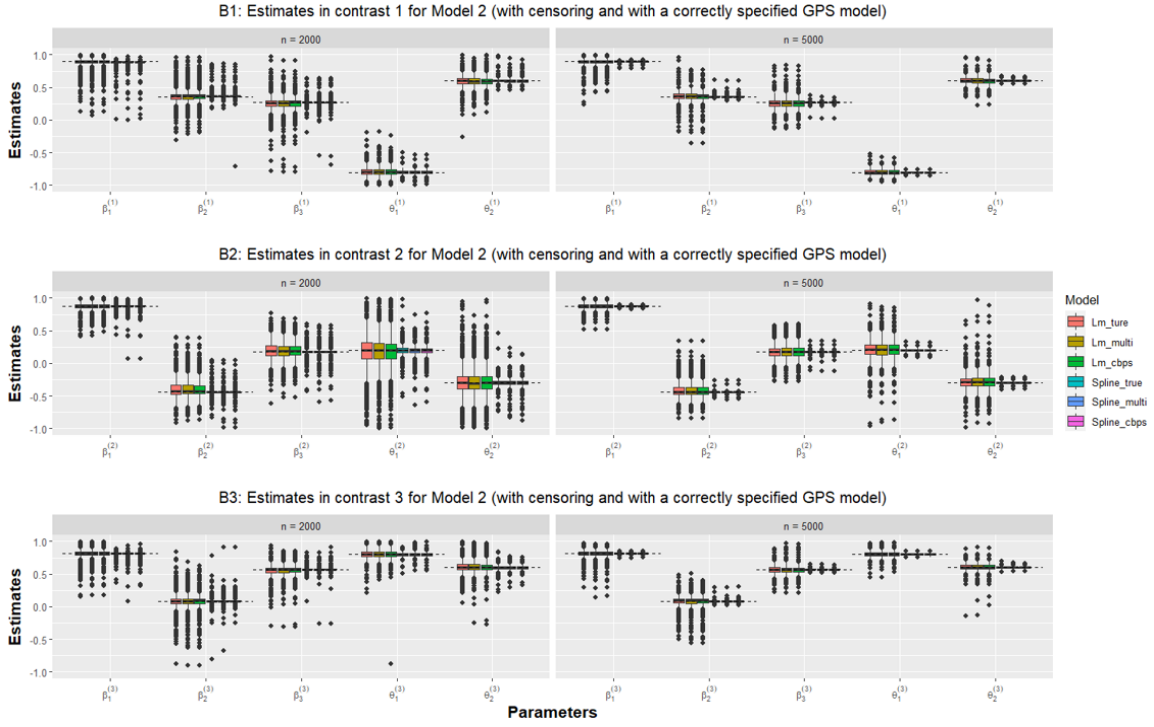


Figure 2.7: Simulation results for outcome Model 3 with censoring and with a correctly specified GPS model.

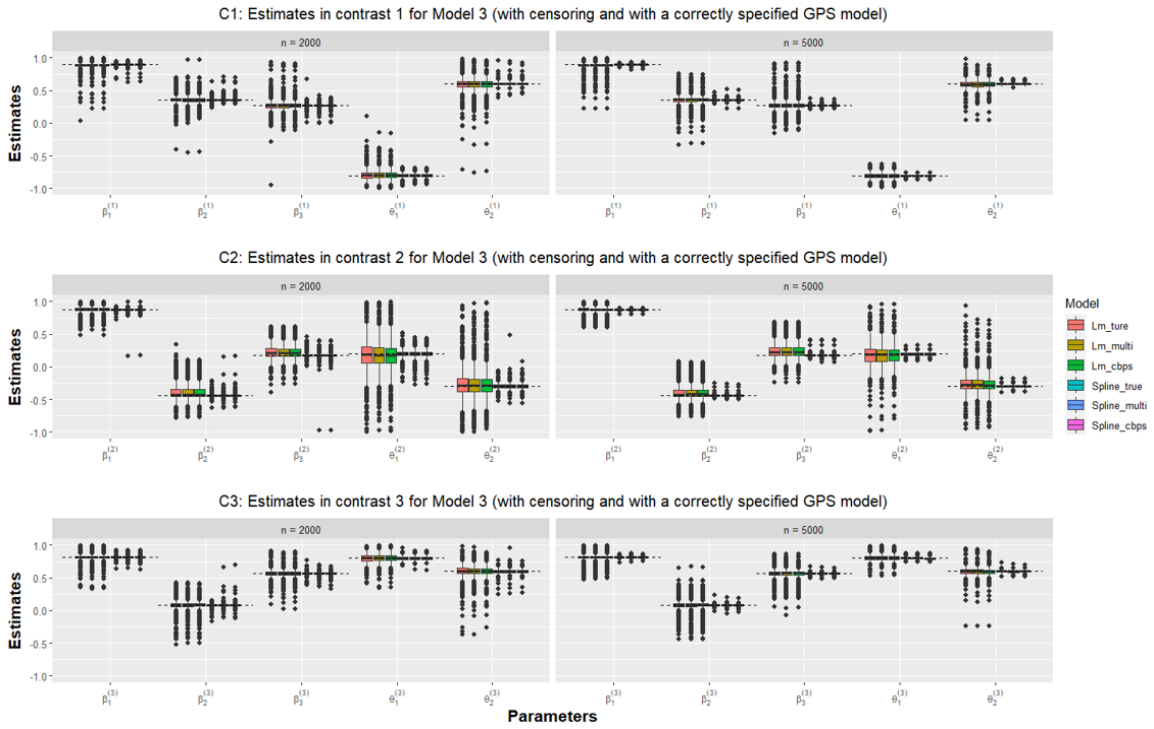



Figure 2.8: Simulation results for outcome Model 4 with censoring and with a correctly specified GPS model.

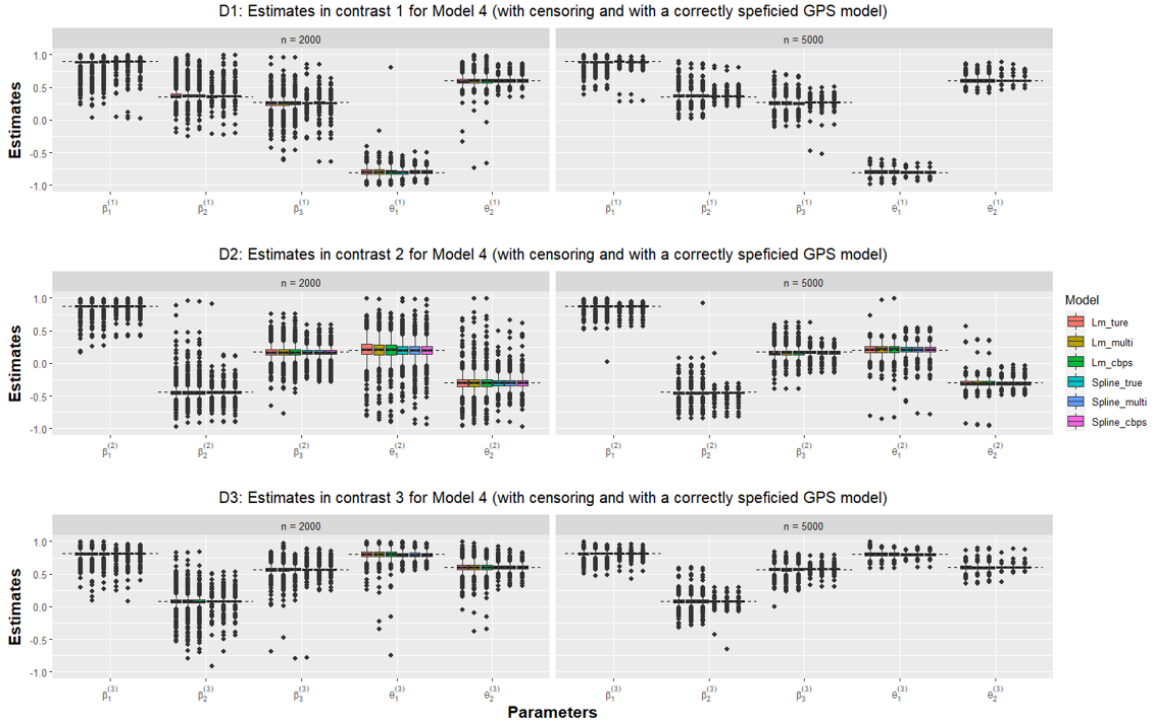


Figure 2.9: Simulation results for outcome Model 1 with censoring and with a mis-specified GPS model.

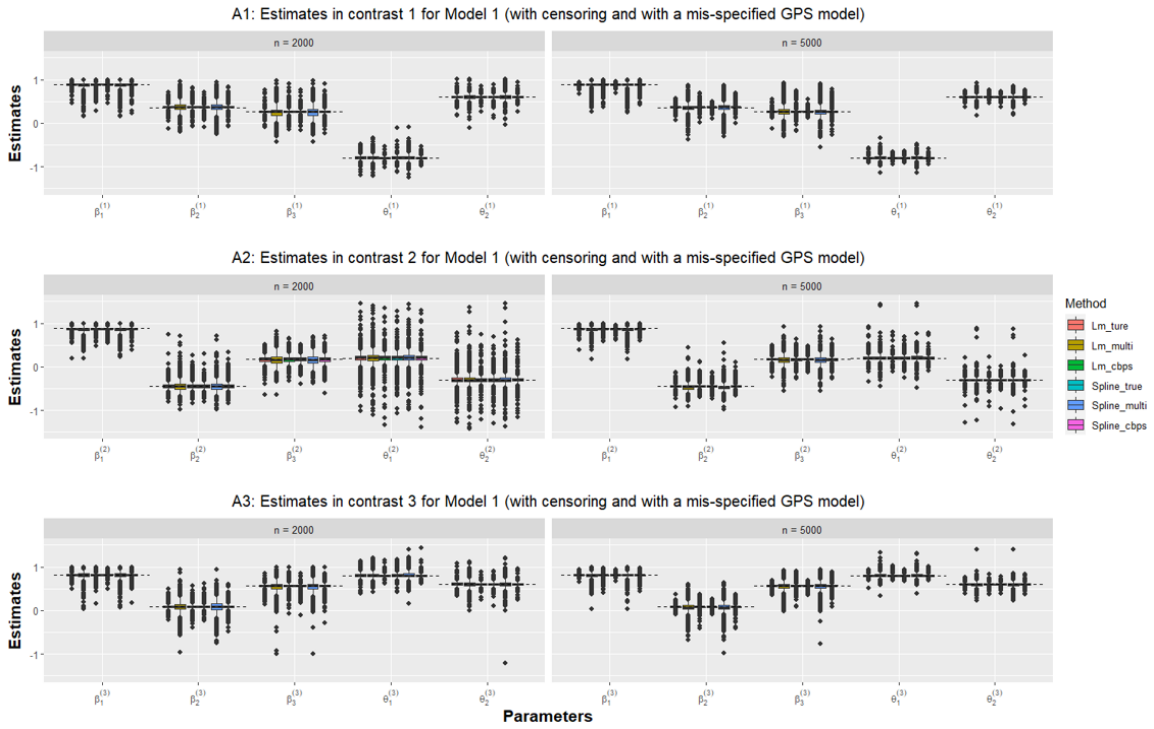



Figure 2.10: Simulation results for outcome Model 2 with censoring and with a mis-specified GPS model.

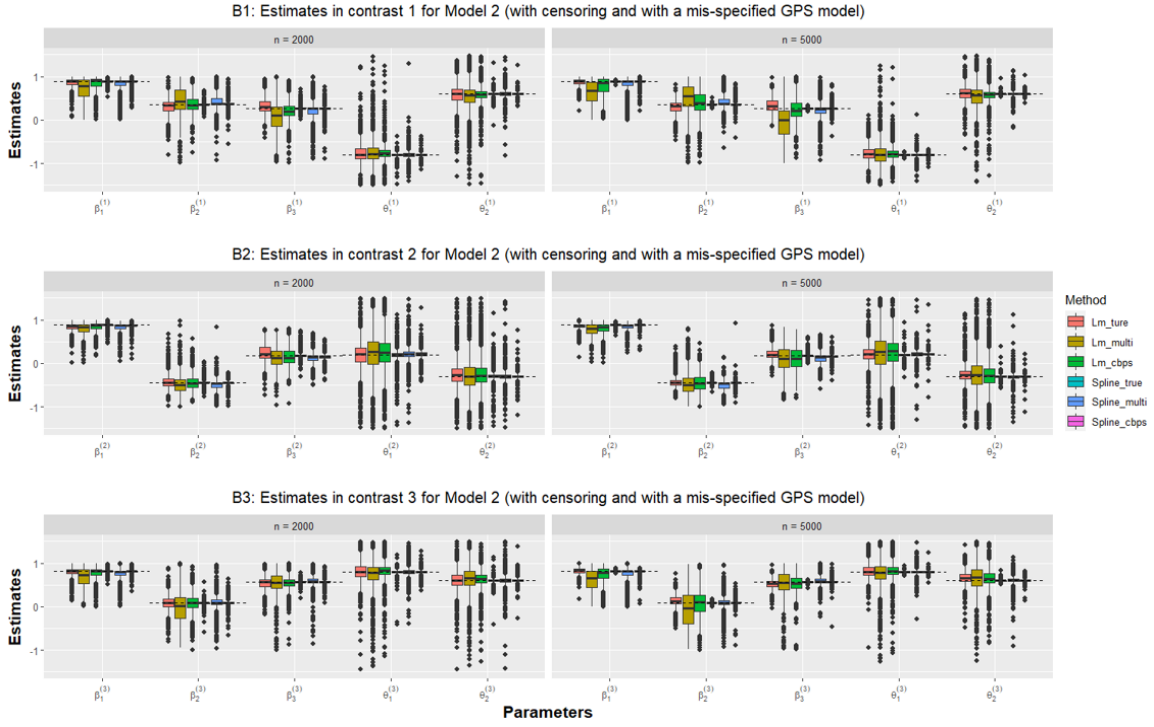


Figure 2.11: Simulation results for outcome Model 3 with censoring and with a mis-specified GPS model.

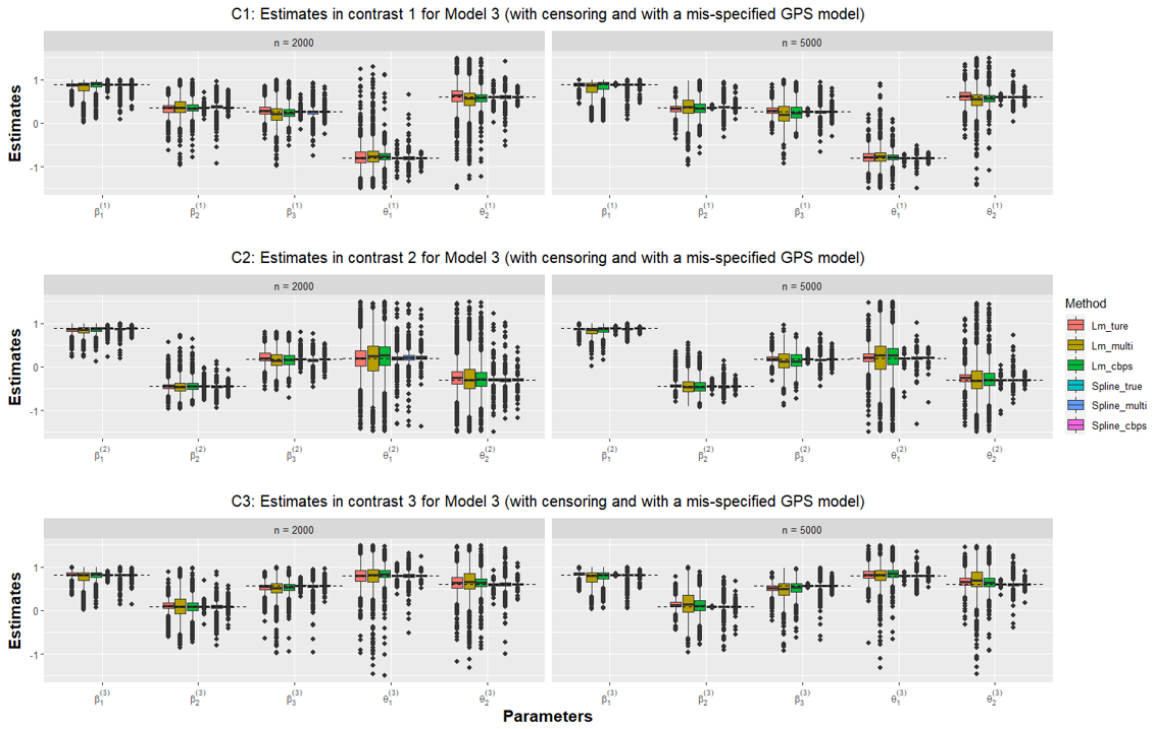



Figure 2.12: Simulation results for outcome Model 4 with censoring and with a mis-specified GPS model.

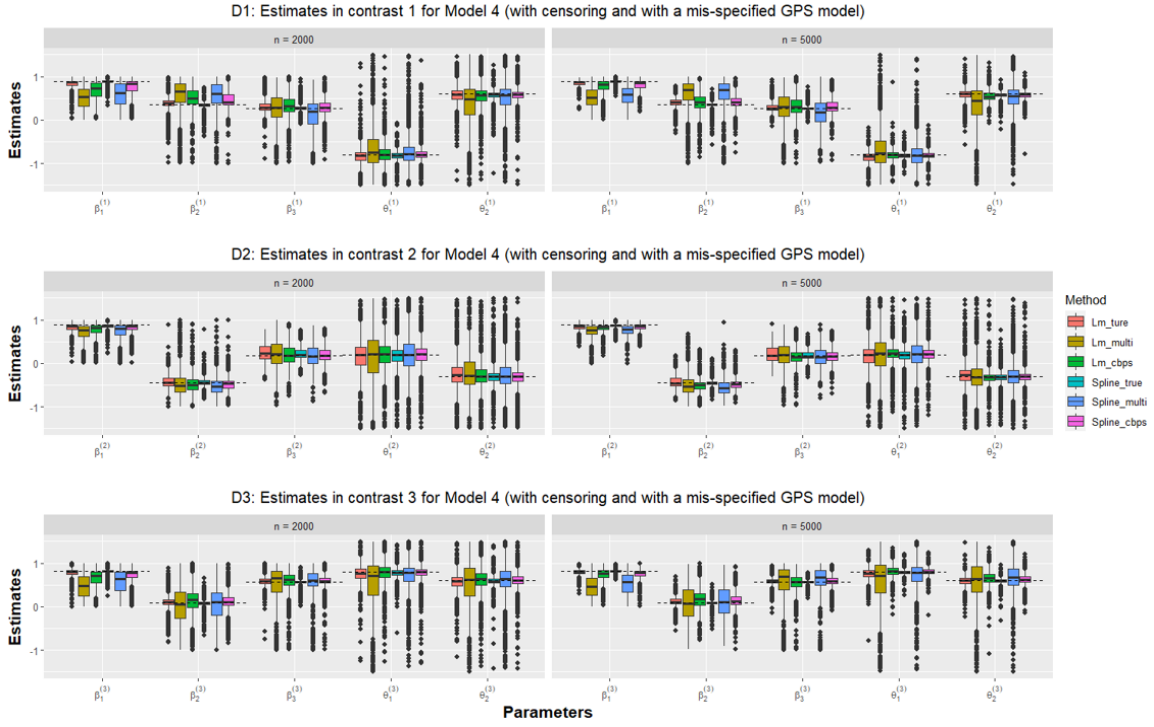


Table 2.3 summarizes the accuracy of each simulation setting for different sample sizes ($n$ = 2000 and 5000) in the three scenarios. Based on this table, we conclude that (1) the PCD increases with sample size increased, no matter how the data is generated; (2) PCDs are close between linear outcome model and GPLSIM outcome model when the outcome model is a simple linear model (Model 1), and PCDs are higher in GPLSIM outcome model when the outcome model is generated from a GPLSIM (Model 2 and 3) or more complex model (Model 4); (3) When a corrected propensity score model is applied, the PCDs are close regardless the GPS models used; (4) When a mis-specified GPS model is used, the proposed method with CBPS has a better performance than with multinomial GPS model, and the model with outcome GPLSIM performs better than linear outcome model.

## 2.4 Case study

In this section, we apply the proposed method to examine who would benefit more from physical exercise using the third national health and nutrition examination survey (NHANES III: 1988-1994) data set. The mortality and survival information by year 2020 on the subjects are obtained from the Centers for Disease Control and Prevention (`https://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/linked_mortality/`).

Literature shows that physical activity is inversely associated with mortality (Blair et al., 1989). A lower physical activity is associated with an increased mortality risk, and a greater moderate-to-vigorous physical activity is associated with lower mortality (Saint-Maurice et al., 2018). In this study, we take different levels of physical activity as treatment groups, and subjects are classified to inactive, insufficient active and active physical exercise groups according to the frequency and intensity of physical activity in a week. We only include subjects aged 50-70 years old when they were interviewed in NHANES III to have a more homogeneous group in age.

Figure 2.13: Estimated survival time for subjects witout physical activity.

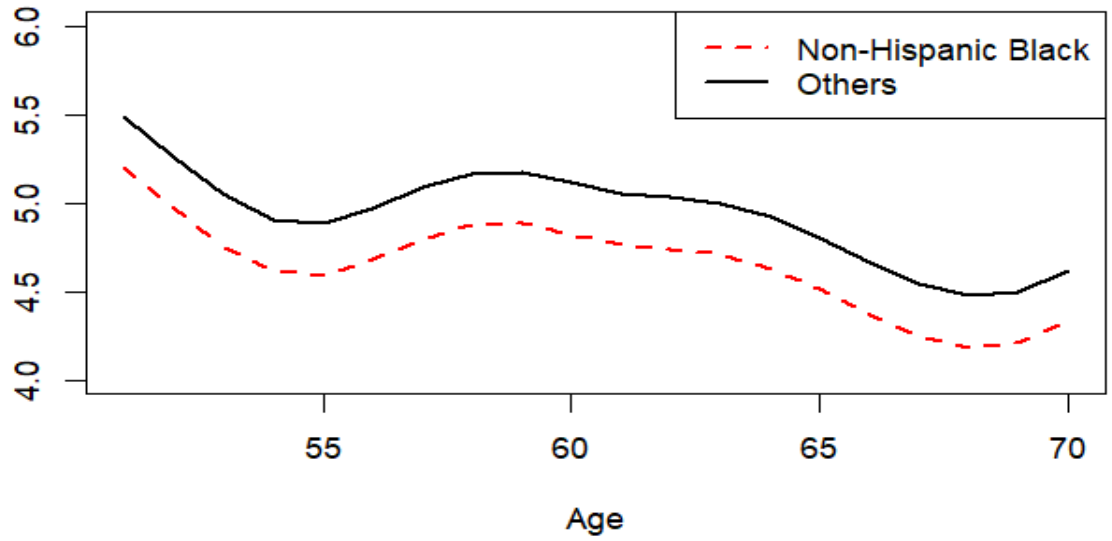



Figure 2.14: Estimated contrast function versus age for insufficient physical activity and active physical activity comparing with inactive group, stratified by race (Non-Hispanic black versus others).

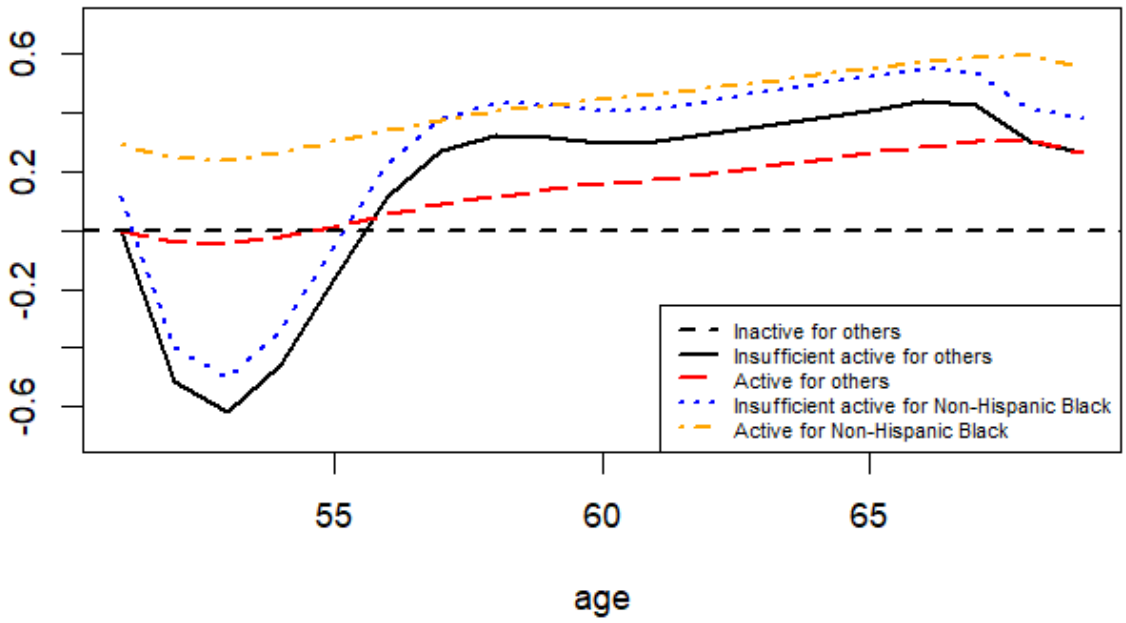


Table 2.4: The distributions of confounding variables among three different physical exercise groups.

| | | | Levels of physical activity | | | |
|---|---|---|---|---|---|---|
| | | | Inactive | Insufficient active | Active | |
| Continuous variables | | | Mean (SD) | Mean (SD) | Mean (SD) | P-Value |
| Age | | | 61.20 (5.60) | 60.45 (5.59) | 61.38 (5.69) | $< 0.001$ |
| BMI | | | 29.32 (6.50) | 28.61 (5.38) | 27.68 (5.05) | $< 0.001$ |
| Education | | | 8.60 (4.58) | 10.44 (4.09) | 11.11 (4.03) | $< 0.001$ |
| Poverty ratio | | | 1.98 (1.64) | 2.85 (1.92) | 3.12 (2.06) | $< 0.001$ |
| | | Total | Inactive | Insufficient active | Active | |
| Categorical variables | Level | N (%) | N (%) | N (%) | N (%) | P-Value |
| Overall | | 3084 | 732 (23.7%) | 1279 (41.5%) | 1073 (34.8%) | |
| Gender | Male | 1495 | 257 (17.2%) | 647 (43.3%) | 591 (39.5%) | $< 0.001$ |
| | Female | 1589 | 475 (29.9%) | 632 (39.8%) | 482 (30.3%) | |
| Non-Hispanic Black | Yes | 750 | 229 (30.5%) | 292 (38.9%) | 229 (30.5%) | $< 0.001$ |
| | No | 2334 | 503 (21.6%) | 987 (42.3%) | 844 (36.2%) | |
| RUC | Metro | 1390 | 332 (23.9%) | 543 (39.1%) | 515 (37.1%) | 0.026 |
| | No metro | 1694 | 400 (23.6%) | 736 (43.4%) | 558 (32.9%) | |
| Counts of comorbidities | $< 2$ | 1084 | 197 (18.2%) | 457 (42.2%) | 430 (39.7%) | $< 0.001$ |
| | $\geq 2$ | 2000 | 535 (26.8%) | 822 (41.1%) | 643 (32.1%) | |

Table 2.5: The distributions of confounding variables stratified by outcomes.

| | | | Survival | Death | |
|---|---|---|---|---|---|
| Continuous variables | | | Mean (SD) | Mean (SD) | P-Value |
| Age | | | 57.57 (4.89) | 62.31 (5.35) | $< 0.001$ |
| BMI | | | 28.33 (5.08) | 28.51 (5.78) | 0.672 |
| Education | | | 10.91 (4.39) | 9.96 (4.23) | $< 0.001$ |
| Poverty ratio | | | 3.16 (1.99) | 2.51 (1.92) | $< 0.001$ |
| | | Total | Survival | Death | |
| Categorical variables | Level | N | N (%) | N (%) | P-Value |
| Overall | | 3084 | 885 (28.7%) | 2199 (71.3%) | |
| Gender | Male | 1495 | 358 (23.9%) | 1137 (76.1%) | $< 0.001$ |
| | Female | 1589 | 527 (33.2%) | 1062 (66.8%) | |
| Non-Hispanic Black | Yes | 750 | 179 (23.9%) | 571 (76.1%) | 0.001 |
| | No | 2334 | 706 (30.2%) | 1628 (69.8%) | |
| RUC | Metro | 1390 | 425 (30.6%) | 965 (69.4%) | 0.040 |
| | No metro | 1694 | 460 (27.2%) | 1234 (72.8%) | |
| Counts of comorbidities | $\leq 1$ | 1084 | 449 (41.4%) | 635 (58.6%) | $< 0.001$ |
| | $\geq 2$ | 2000 | 436 (21.8%) | 1564 (78.2%) | |

The treatment group, in terms of level of physical activity is created based on the survey questions on Household Adult Questionnaire (HAQ). The frequency and metabolic equivalent of task (MET) score for different types of activities are

Table 2.6: Estimated parameters in the outcome under control model and in the two contrast functions.

| | Outcome under control | | | Contrasts | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Inactive | | | Insufficient active vs Inactive | | | Active vs Inactive | | |
| Variables | Est. | SE | 95 % CI | Est. | SE | 95 % CI | Est. | SE | 95 % CI |
| Age | **-0.030** | **0.007** | **(-0.047, -0.022)** | **0.288** | **0.198** | **(0.017, 0.784)** | **0.236** | **0.179** | **(0.009, 0.657)** |
| BMI | **0.019** | **0.006** | **(0.006, 0.026)** | -0.390 | 0.251 | (-0.829, 0.114) | -0.319 | 0.292 | (-0.830, 0.306) |
| Education | 0.010 | 0.010 | (-0.013, 0.026) | 0.208 | 0.423 | (-0.640, 0.855) | -0.008 | 0.451 | (-0.785, 0.868) |
| Poverty ratio | 0.010 | 0.029 | (-0.044, 0.068) | 0.208 | 0.630 | (-0.918, 0.983) | 0.212 | 0.691 | (-0.990, 0.984) |
| Gender (ref: Male) | **0.229** | **0.085** | **(0.061, 0.399)** | -0.082 | 0.109 | (-0.282, 0.132) | -0.033 | 0.109 | (-0.245, 0.175) |
| RUC (ref: No metro) | 0.074 | 0.087 | (-0.109, 0.261) | -0.111 | 0.103 | (-0.315, 0.096) | -0.106 | 0.114 | (-0.335, 0.120) |
| Non-Hispanic Black (ref: No) | **-0.256** | **0.087** | **(-0.442, -0.063)** | 0.113 | 0.124 | (-0.130, 0.365) | **0.228** | **0.130** | **(0.045, 0.536)** |
| Comorbidity (ref: $\leq$1) | **-0.256** | **0.092** | **(-0.460, -0.071)** | -0.012 | 0.115 | (-0.237, 0.219) | 0.029 | 0.120 | (-0.209, 0.275) |

Table 2.7: Illustration of optimal treatment assignment for patients with similar characteristics.

| Comparison | Gender | Non-Hispanic Black | RUC | Comorbidity | Age | BMI | Education | Poverty ratio | Observed Trt | Observed outcome | Optimal Trt | Optimal outcome (spline) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Set 1 | Female | Yes | Metro | $\geq 2$ | 62 | 24.5 | 6 | 2.46 | Inactive | 5.19 | Active | 5.24 |
| | Female | Yes | Metro | $\geq 2$ | 62 | 24.0 | 12 | 3.48 | Insufficient | 5.28 | Active | 5.42 |
| | Female | Yes | Metro | $\geq 2$ | 62 | 25.0 | 12 | 0.54 | Active | 5.28 | Active | 5.34 |
| Set 2 | Female | No | No metro | $\geq 2$ | 63 | 22.4 | 12 | 2.20 | Inactive | 4.64 | Insufficient | 5.56 |
| | Female | No | No metro | $\geq 2$ | 63 | 23.5 | 9 | 2.47 | Insufficient | 5.42 | Insufficient | 5.42 |
| | Female | No | No metro | $\geq 2$ | 63 | 21.1 | 12 | 0.45 | Active | 5.15 | Insufficient | 5.57 |

collected. According to the World Health Organization (WHO) 2020 guidelines, the physical activities are categorized as light-intensity physical activity (LPA) which is performed between 1.5 and 3 METs, moderate-intensity physical activity (MPA) which is performed between 3 and 6 METs, and vigorous-intensity physical activity (VPA) with 6 or more METs (Bull et al., 2020). Combining frequency and intensity, the active physical exercise group is defined as having MPA five or more times per week or VPA at least three times per week. Inactive group is defined as if no physical activity is reported. Insufficient active group is defined as if there is reported physical activities but do not meet the criteria of active group (Beddhu et al., 2009). Demographic variables, laboratory measures and comorbidity are collected and considered as confounding variables. Age, education and poverty income ratio are obtained from the questionnaire data, and body mass index (BMI) is obtained from the examinationdata. Gender, ethnicity, and Rural-Urban commuting (RUC) are obtained from questionnaire taken as categorical variables. We also consider the following comorbidities which are summarized as binary variable. Hypertension is defined as an average systolic blood pressure $\geq$ 140 mmHg and/or an average diastolic blood pres-

sure $\geq$ 90 mmHg, or a self-reported physician diagnosis of hypertension. Diabetes is defined as a glycated hemoglobin (HbAlc) $\geq$ 6.5%, or a self-reported physician diagnosis of diabetes. Dyslipidemia is defined as a total cholestrol $\geq$ 240 mg/dL, or triglyceridemia $\geq$ 150 mg/dL, or a low-density lipoprotein cholestrol (LDL-C) $\geq$ 130 mg/dL, or a high-density lipoprotein cholestrol (HDL-C) $<$ 40 mg/dL for men and $<$50 mg/dL for women. Chronic kidney disease (CKD) is defined as the estimated glomerular filtration rate (eGFR) $<$ 60 mL/min/1.73$m^2$ and/or a urinary ACR $>$ 30 mg/g (Liu et al., 2013). A diagnosis of stroke, congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), cancer, and coronary heart disease (CHD) are from self-reported diagnosis. The final comorbidity condition is summarized as a binary variable according to whether the subject has 2 or more the above mentioned comorbidity conditions. We exclude subjects who have missing information in mortality, physical activity, or other characteristics, and we also exclude subjects who died from accidents. Totally, 3084 subjects are included in the current study.

Table 2.4 and 2.5 show the distribution of these confounding variables stratified by treatment assignment and outcome respectively. The significant difference is shown when the p-value is less than 0.05. Table 2.6 shows the estimates of parameters in the outcome model and contrast function using GPLSIM where the standard error and 95% CI are obtained using 1000 bootstrap samples. In both the outcome under control and contrast functions, age is significant variable. Race is significant in the outcome under control and the comparison between active and inactive group. Other variables are not significant in the model. Figure 2.13 shows the outcome profile for subjects without physical activity. Based on Figure 2.13, we can see (1) for subjects without physical activity, their survival time declines as subjects are aged; (2) the survival time for non-Hispanic black is significantly shorter than other races. Figure 2.14 shows the estimated contrast functions for insufficient active physical exercise group and active physical exercise group, comparing with inactive physical exercise

group. From Figure 2.14, we can see that (1) both insufficient physical exercise and active physical exercise are beneficial for longer survival time for patients aged 55 and older, comparing with inactive physical exercise group; and (2) non-Hispanic black benefits significantly from active physical exercise.

To illustrate the predicted optimal treatment group versus observed outcome under different physical exercise levels, we select two sets of subjects. In each set, the characteristics of the subjects are similar but with different physical exercise level (see Table 2.7). In the first set, the three subjects are 62-year old female, Non-Hispanic black, from metro area, and have 2 or more comorbidities. However, their physical exercise levels are different; one with inactive exercise (treatment 1), one with insufficient active exercise (treatment 2), and one with active exercise (treatment 3). According to their characteristics, the estimated optimal treatment is active exercise. Subject who is in the optimal treatment would have a longer survival time. In the estimated optimal outcome, the estimates with GPLSIM and CBPS appears to be close to the observed outcome. From the set 2, we see that with optimal treatment has the longest survival time.

## 2.5 Conclusion

In this project, we propose to use structural AFT mean model to select optimal treatment. The frame work of structural AFT mean model is used to estimate the benefit from a treatment. GPLSIM is used to estimate the outcome function under control and contrast functions. The proposed model provides more flexibility in modeling contrast functions. We incorporate the inverse probability of censoring weights and propensity score to reduce the impact due to censoring and confounding variables.

We present the accuracy and MSE of proposed method and show a consistent estimation when either the propensity score model or the outcome model under con-

trol is correctly specified. Furthermore, even when both models are mis-specified, the proposed method still selects optimal treatment with a good accuracy.

# REFERENCES


Bather, J. (2000). *Decision theory: An introduction to dynamic programming and sequential decisions*. John Wiley & Sons, Inc.

Beddhu, S., Baird, B. C., Zitterkoph, J., Neilson, J., and Greene, T. (2009). Physical activity and mortality in chronic kidney disease (nhanes iii). *Clinical Journal of the American Society of Nephrology*, 4(12):1901–1906.

Blair, S. N., Kohl, H. W., Paffenbarger, R. S., Clark, D. G., Cooper, K. H., and Gibbons, L. W. (1989). Physical fitness and all-cause mortality: a prospective study of healthy men and women. *Jama*, 262(17):2395–2401.

Boehm, W. (1980). Inserting new knots into b-spline curves. *Computer-Aided Design*, 12(4):199–201.

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., Carty, C., Chaput, J.-P., Chastin, S., Chou, R., et al. (2020). World health organization 2020 guidelines on physical activity and sedentary behaviour. *British journal of sports medicine*, 54(24):1451–1462.

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92(438):477–489.

Charig, C. R., Webb, D. R., Payne, S. R., and Wickham, J. E. (1986). Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *Br Med J (Clin Res Ed)*, 292(6524):879–882.

Chatton, A., Le Borgne, F., Leyrat, C., Gillaizeau, F., Rousseau, C., Barbin, L., Laplaud, D., Léger, M., Giraudeau, B., and Foucher, Y. (2020). G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. *Scientific reports*, 10(1):9219.

Chattopadhyay, A., Hase, C. H., and Zubizarreta, J. R. (2020). Balancing vs modeling approaches to weighting in practice. *Statistics in Medicine*, 39(24):3227–3254.

Chen, P.-Y. and Tsiatis, A. A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics*, 57(4):1030–1038.

Cole, S. R. and Frangakis, C. E. (2009). The consistency statement in causal inference: a definition or an assumption? *Epidemiology*, 20(1):3–5.

Deb, S., Austin, P. C., Tu, J. V., Ko, D. T., Mazer, C. D., Kiss, A., and Fremes, S. E. (2016). A review of propensity-score methods and their use in cardiovascular research. *Canadian Journal of Cardiology*, 32(2):259–265.

Esser, M. B., Sherk, A., Liu, Y., Naimi, T. S., Stockwell, T., Stahre, M., Kanny, D., Landen, M., Saitz, R., and Brewer, R. D. (2020). Deaths and years of potential life lost from excessive alcohol use—united states, 2011–2015. *Morbidity and Mortality Weekly Report*, 69(39):1428.

Geng, Y., Zhang, H. H., and Lu, W. (2015). On optimal treatment regimes selection for mean survival time. *Statistics in medicine*, 34(7):1169–1184.

Guo, S. and Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications*, volume 11. SAGE publications.

Guo, W., Zhou, X.-H., and Ma, S. (2021). Estimation of optimal individualized treat-

ment rules using a covariate-specific treatment effect curve with high-dimensional covariates. *Journal of the American Statistical Association*, 116(533):309–321.

Hall, W. H., Ramachandran, R., Narayan, S., Jani, A. B., and Vijayakumar, S. (2004). An electronic application for rapidly calculating charlson comorbidity score. *BMC cancer*, 4(1):1–8.

Hernan, M. and Robins, J. (2018). Causal inference. boca ration, florida.

Hernán, M. A. (2012). Beyond exchangeability: the other conditions for causal inference in medical research.

Horowitz, J. L. (2001). The bootstrap. In *Handbook of econometrics*, volume 5, pages 3159–3228. Elsevier.

Hu, H., Mitra, R., Han, Y., Pal, S., Huang, H., Vatsalya, V., et al. (2022). Prevalence and treatment for alcohol use disorders based on kentucky medicaid 2012-2019 datasets. *J Alcohol Drug Depend*, 10(4).

Imai, K. and Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):243–263.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

Ishigooka, J., Murasaki, M., Miura, S., and Group, T. O. L.-P. I. S. (2000). Olanzapine optimal dose: Results of an open-label multicenter study in schizophrenic patients. *Psychiatry and clinical neurosciences*, 54(4):467–478.

Jain, K. K. (2002). Personalized medicine. *Current opinion in molecular therapeutics*, 4(6):548–558.

Jiang, R., Lu, W., Song, R., Hudgens, M. G., and Naprvavnik, S. (2017). Doubly robust estimation of optimal treatment regimes for survival data—with application to an hiv/aids study. *The annals of applied statistics*, 11(3):1763.

Julious, S. A. and Mullee, M. A. (1994). Confounding and simpson's paradox. *Bmj*, 309(6967):1480–1481.

Kranzler, H. R. and Soyka, M. (2018). Diagnosis and pharmacotherapy of alcohol use disorder: a review. *Jama*, 320(8):815–824.

Li, J., Handorf, E., Bekelman, J., and Mitra, N. (2016). Propensity score and doubly robust methods for estimating the effect of treatment on censored cost. *Statistics in medicine*, 35(12):1985–1999.

Li, Y. and Li, L. (2021). Propensity score analysis methods with balancing constraints: A monte carlo study. *Statistical methods in medical research*, 30(4):1119–1142.

Litten, R. Z., Wilford, B. B., Falk, D. E., Ryan, M. L., and Fertig, J. B. (2016). Potential medications for the treatment of alcohol use disorder: an evaluation of clinical efficacy and safety. *Substance abuse*, 37(2):286–298.

Liu, X., Wang, Y., Wang, C., Shi, C., Cheng, C., Chen, J., Ma, H., Lv, L., Li, L., and Lou, T. (2013). A new equation to estimate glomerular filtration rate in chinese elderly population. *PloS one*, 8(11):e79675.

Lu, J. (2018). On the partial identification of a new causal measure for ordinal outcomes. *Statistics & Probability Letters*, 137:1–7.

Murphy, K. P. (2002). *Dynamic bayesian networks: representation, inference and learning*. University of California, Berkeley.

National Institutes of Health (2014). Treatment for alcohol problems: Finding and getting help. Technical report, NIH Publication.

O'shea, R. S., Dasarathy, S., and McCullough, A. J. (2010). Alcoholic liver disease. *Official journal of the American College of Gastroenterology— ACG*, 105(1):14–32.

Pan, Y. and Zhao, Y.-Q. (2021). Improved doubly robust estimation in learning optimal individualized treatment rules. *Journal of the American Statistical Association*, 116(533):283–294.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146.

Piano, M. R. (2002). Alcoholic cardiomyopathy: incidence, clinical characteristics, and pathophysiology. *Chest*, 121(5):1638–1650.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2):1180.

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.-C., Saunders, L. D., Beck, C. A., Feasby, T. E., and Ghali, W. A. (2005). Coding algorithms for defining comorbidities in icd-9-cm and icd-10 administrative data. *Medical care*, pages 1130–1139.

Rogal, S., Youk, A., Zhang, H., Gellad, W. F., Fine, M. J., Good, C. B., Chartier, M., DiMartini, A., Morgan, T., Bataller, R., et al. (2020). Impact of alcohol use disorder treatment on clinical outcomes among patients with cirrhosis. *Hepatology*, 71(6):2080–2092.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Saint-Maurice, P. F., Troiano, R. P., Matthews, C. E., and Kraus, W. E. (2018). Moderate-to-vigorous physical activity and all-cause mortality: do bouts matter? *Journal of the American Heart Association*, 7(6):e007678.

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: a review journal of the Institute of Mathematical Statistics*, 29(4):640.

Simoneau, G., Moodie, E. E., Nijjar, J. S., Platt, R. W., Investigators, S. E. R. A. I. C., et al. (2020). Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association*, 115(531):1531–1539.

Smothers, B. A., Yahr, H. T., and Ruhl, C. E. (2004). Detection of alcohol use disorders in general hospital admissions in the united states. *Archives of internal medicine*, 164(7):749–756.

Song, R., Luo, S., Zeng, D., Zhang, H. H., Lu, W., and Li, Z. (2017). Semiparametric single-index model for estimating optimal individualized treatment strategy. *Electronic journal of statistics*, 11(1):364.

Sugihara, M. (2010). Survival analysis using inverse probability of treatment weighted methods based on the generalized propensity score. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, 9(1):21–34.

VanderWeele, T. J., Luedtke, A. R., van der Laan, M. J., and Kessler, R. C. (2019). Selecting optimal subgroups for treatment using many covariates. *Epidemiology (Cambridge, Mass.)*, 30(3):334.

Vittadini, G., Buonocore, M., Colli, G., Terzi, M., Fonte, R., and Biscaldi, G. (2001). Alcoholic polyneuropathy: a clinical and epidemiological study. *Alcohol and alcoholism*, 36(5):393–400.

Wallace, M. P., Moodie, E. E., and Stephens, D. A. (2019). Model selection for g-estimation of dynamic treatment regimes. *Biometrics*, 75(4):1205–1215.

Wang, J.-L., Xue, L., Zhu, L., and Chong, Y. S. (2010). Estimation for a partial-linear single-index model.

Weisner, C., Ray, G. T., Mertens, J. R., Satre, D. D., and Moore, C. (2003). Short-term alcohol and drug treatment outcomes predict long-term outcome. *Drug and alcohol dependence*, 71(3):281–294.

Witkiewitz, K., Litten, R., and Leggio, L. (2019). Advances in the science and treatment of alcohol use disorder. *Science advances*, 5(9):eaax4043.

Yao, L. and Tarpey, T. (2022). A single index model for longitudinal outcomes to optimize individual treatment decision rules. *Stat*, page e493.

Zhang, R., Huang, Z., and Lv, Y. (2010). Statistical inference for the index parameter in single-index models. *Journal of Multivariate Analysis*, 101(4):1026–1041.

Zhao, Y., Kosorok, M. R., and Zeng, D. (2009). Reinforcement learning design for cancer clinical trials. *Statistics in medicine*, 28(26):3294–3315.

Zhao, Y., Zeng, D., Socinski, M. A., and Kosorok, M. R. (2011). Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics*, 67(4):1422–1433.

# APPENDIX

## Appendix 1: Derivation for equation

Derivation for Equation (1.5)

Let us denote $\pi_i^{(a)} = Pr(Y^{(a)} = i)$ for $i = 1, 2, \cdots, c$ and $a = 0, 1, \cdots, K$. Let us denote $\gamma_i^{(a)} = Pr(Y^{(a)} \leq i)$ for $i = 1, 2, \cdots, c$ and $a = 0, 1, \cdots, K-1$. It is clear that $\gamma_c^{(a)} = 1$. Then we have:

$\pi_1^{(a)} = Pr(Y^{(a)} = 1) = Pr(Y^{(a)} \leq 1) = \gamma_1^{(0)}$ for $a = 0, 1, \cdots, K-1$,

$\pi_i^{(a)} = Pr(Y^{(a)} = i) = Pr(Y^{(a)} \leq i) - Pr(Y^{(a)} \leq i-1) = \gamma_i^{(a)} - \gamma_{i-1}^{(a)}$, for $i = 2, 3, \cdots, c$, and $a = 0, 1, \cdots, K-1$.

We can obtain:

$$\theta_a = Pr(Y^{(0)} < Y^{(a)}) + 0.5Pr(Y^{(0)} = Y^{(a)})$$

$$= \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} Pr(Y^{(0)} = i)Pr(Y^{(a)} = j) + 0.5 \sum_{i=1}^{c} Pr(Y^{(0)} = i)Pr(Y^{(a)} = i)$$

$$= \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} \pi_i^{(0)} \pi_j^{(a)} + 0.5 \sum_{i=1}^{c} \pi_i^{(0)} \pi_i^{(a)}$$

$$= [\pi_1^{(a)}, \pi_2^{(a)}, \cdots, \pi_{c-1}^{(a)}, \pi_c^{(a)}] \times \begin{bmatrix} 0.5 & 0 & \cdots & 0 & 0 \\ 1 & 0.5 & \cdots & 0 & 0 \\ & \cdots & \cdots & \cdots & \\ 1 & 1 & \cdots & 0.5 & 0 \\ 1 & 1 & \cdots & 1 & 0.5 \end{bmatrix} \times \begin{bmatrix} \pi_1^{(0)} \\ \pi_2^{(0)} \\ \vdots \\ \pi_{c-1}^{(0)} \\ \pi_c^{(0)} \end{bmatrix}$$

$$= [\gamma_1^{(a)}, \gamma_2^{(a)} - \gamma_1^{(a)}, \cdots, \gamma_{c-1}^{(a)} - \gamma_{c-2}^{(a)}, 1 - \gamma_{c-1}^{(a)}] \times \begin{bmatrix} 0.5 & 0 & \cdots & 0 & 0 \\ 1 & 0.5 & \cdots & 0 & 0 \\ & \cdots & \cdots & \cdots & \\ 1 & 1 & \cdots & 0.5 & 0 \\ 1 & 1 & \cdots & 1 & 0.5 \end{bmatrix} \times \begin{bmatrix} \gamma_1^{(0)} \\ \gamma_2^{(0)} - \gamma_1^{(0)} \\ \vdots \\ \gamma_{c-1}^{(0)} - \gamma_{c-2}^{(0)} \\ 1 - \gamma_{c-1}^{(0)} \end{bmatrix}$$

$$= \begin{bmatrix} 0.5\gamma_1^{(a)} + \gamma_2^{(a)} - \gamma_1^{(a)} + \cdots + \gamma_{c-1}^{(a)} - \gamma_{c-2}^{(a)} + 1 - \gamma_{c-1}^{(a)} \\ 0.5(\gamma_2^{(a)} - \gamma_1^{(a)}) + \gamma_3^{(a)} - \gamma_2^{(a)} + \cdots + \gamma_{c-1}^{(a)} - \gamma_{c-2}^{(a)} + 1 - \gamma_{c-1}^{(a)} \\ \vdots \\ 0.5(\gamma_{c-1}^{(a)} - \gamma_{c-2}^{(a)}) + 1 - \gamma_{c-1}^{(a)} \\ 0.5(1 - \gamma_{c-1}^{(a)}) \end{bmatrix} \times \begin{bmatrix} \gamma_1^{(0)} \\ \gamma_2^{(0)} - \gamma_1^{(0)} \\ \vdots \\ \gamma_{c-1}^{(0)} - \gamma_{c-2}^{(0)} \\ 1 - \gamma_{c-1}^{(0)} \end{bmatrix}$$

$$= \gamma_1^{(0)}[1 - 0.5\gamma_1^{(a)}] + (\gamma_2^{(0)} - \gamma_1^{(0)})[1 - 0.5\gamma_1^{(a)} - 0.5\gamma_2^{(a)}] + \cdots +$$

$$(\gamma_{c-1}^{(0)} - \gamma_{c-2}^{(0)})[1 - 0.5\gamma_{c-1}^{(a)} - 0.5\gamma_{c-2}^{(a)}] + (1 - \gamma_{c-1}^{(0)})[0.5 - 0.5\gamma_{c-1}^{(a)}]$$

$$= [0.5 + 0.5\gamma_{c-1}^{(0)} - 0.5\gamma_{c-1}^{(a)}] + [0.5\gamma_1^{(0)}\gamma_2^{(a)} - 0.5\gamma_2^{(0)}\gamma_1^{(a)}] + \cdots +$$

$$[0.5\gamma_{c-3}^{(0)}\gamma_{c-2}^{(a)} - 0.5\gamma_{c-2}^{(0)}\gamma_{c-3}^{(a)}] + [0.5\gamma_{c-2}^{(0)}\gamma_{c-1}^{(a)} - 0.5\gamma_{c-1}^{(0)}\gamma_{c-2}^{(a)}]$$

$$= [0.5 + 0.5\gamma_{c-1}^{(0)} - 0.5\gamma_{c-1}^{(a)}] + \left[\gamma_1^{(0)}, \gamma_2^{(0)}, \cdots, \gamma_{c-2}^{(0)}, \gamma_{c-1}^{(0)}\right] \times \begin{bmatrix} 0 & 0.5 & \cdots & 0 & 0 \\ -0.5 & 0 & \cdots & 0 & 0 \\ & \cdots & \cdots & \cdots & \\ 0 & 0 & \cdots & 0 & 0.5 \\ 0 & 0 & \cdots & -0.5 & 0 \end{bmatrix} \times \begin{bmatrix} \gamma_1^{(a)} \\ \gamma_2^{(a)} \\ \vdots \\ \gamma_{c-2}^{(a)} \\ \gamma_{c-1}^{(a)} \end{bmatrix}$$

Recall that $\gamma^{(0)} = (\gamma_1^0, \gamma_2^{(0)}, \cdots, \gamma_{c-1}^{(0)})^T, \gamma^{(a)} = (\gamma_1^a, \gamma_2^{(a)}, \cdots, \gamma_{c-1}^{(a)})^T$. Thus, Equation (1.5) follows.

Derivation for Equation (1.6)

Note that the marginal structural model: $logit\ Pr(Y^{(a)} \leq j) = \alpha_j - \sum_{k=1}^{3} \tau_k I_{a=k}$, $(j = 1, 2, a = 0, 1, 2, 3)$, which implies that:

$logit\ Pr(Y^{(0)} \leq j) = \alpha_j$, for potential outcome under control, and

$logit\ Pr(Y^{(a)} \leq j) = \alpha_j - \tau_a$, for potential outcome under treatment $a$.

Thus, we have:

$\gamma_j^{(0)} = Pr(Y^{(0)} \leq j) = \frac{exp(\alpha_j)}{1+exp(\alpha_j)}$,

$\gamma_j^{(a)} = Pr(Y^{(a)} \leq j) = \frac{exp(\alpha_j - \tau_a)}{1+exp(\alpha_j - \tau_a)}$.

Note that $\gamma^{(0)} = [\gamma_1^{(0)}, \gamma_2^{(0)}]$, $\gamma^{(a)} = [\gamma_1^{(a)}, \gamma_2^{(a)}]$, where:

$\gamma_1^{(0)} = \frac{exp(\alpha_1)}{1+exp(\alpha_1)}$, $\gamma_2^{(0)} = \frac{exp(\alpha_2)}{1+exp(\alpha_2)}$, $\gamma_1^{(a)} = \frac{exp(\alpha_1 - \tau_a)}{1+exp(\alpha_1 - \tau_a)}$, and $\gamma_2^{(a)} = \frac{exp(\alpha_2 - \tau_a)}{1+exp(\alpha_2 - \tau_a)}$,

We have:

$\frac{\partial \gamma_1^{(0)}}{\partial \alpha_1} = \frac{exp(\alpha_1)}{[1+exp(\alpha_1)]^2}$, $\frac{\partial \gamma_1^{(a)}}{\partial \alpha_1} = \frac{exp(\alpha_1 - \tau_a)}{[1+exp(\alpha_1 - \tau_a)]^2}$, $\frac{\partial \gamma_2^{(0)}}{\partial \alpha_1} = \frac{\partial \gamma_2^{(a)}}{\partial \alpha_1} = 0$,

$\frac{\partial \gamma_1^{(0)}}{\partial \alpha_2} = \frac{\partial \gamma_1^{(a)}}{\partial \alpha_2} = 0$, $\frac{\partial \gamma_2^{(0)}}{\partial \alpha_2} = \frac{exp(\alpha_2)}{[1+exp(\alpha_2)]^2}$, $\frac{\partial \gamma_2^{(a)}}{\partial \alpha_2} = \frac{exp(\alpha_2 - \tau_a)}{[1+exp(\alpha_2 - \tau_a)]^2}$,

$\frac{\partial \gamma_1^{(0)}}{\partial \tau_a} = 0$, $\frac{\partial \gamma_1^{(a)}}{\partial \tau_a} = \frac{-exp(\alpha_1 - \tau_a)}{[1+exp(\alpha_1 - \tau_a)]^2}$,

$\frac{\partial \gamma_2^{(0)}}{\partial \tau_a} = 0$, $\frac{\partial \gamma_2^{(a)}}{\partial \tau_a} = \frac{-exp(\alpha_2 - \tau_a)}{[1+exp(\alpha_2 - \tau_a)]^2}$.

According to Equation (1.5), $\theta_a = \gamma^{(0)T} D \gamma^{(a)} + 0.5(1 + \gamma_2^{(0)} - \gamma_2^{(a)})$,

where $D = \begin{bmatrix} 0 & 0.5 \\ -0.5 & 0 \end{bmatrix}$. Let $\phi(x) = \frac{e^x}{(1+e^x)^2}$, and $\Phi(x) = \frac{e^x}{1+e^x}$, then we have

$$\frac{\partial \theta_a}{\partial \alpha_1} = \left[\frac{\partial \gamma_1^{(0)}}{\partial \alpha_1}, \frac{\partial \gamma_2^{(0)}}{\partial \alpha_1}\right] D \begin{bmatrix} \gamma_1^{(a)} \\ \gamma_2^{(a)} \end{bmatrix} + \left[\gamma_1^{(0)}, \gamma_2^{(0)}\right] D \begin{bmatrix} \frac{\partial \gamma_1^{(a)}}{\partial \alpha_1} \\ \frac{\partial \gamma_2^{(a)}}{\partial \alpha_1} \end{bmatrix}$$

$$= \left[\frac{exp(\alpha_1)}{[1+exp(\alpha_1)]^2}, 0\right] D \begin{bmatrix} \frac{exp(\alpha_1-\tau_a)}{1+exp(\alpha_1-\tau_a)} \\ \frac{exp(\alpha_2-\tau_a)}{1+exp(\alpha_2-\tau_a)} \end{bmatrix} + \left[\frac{exp(\alpha_1)}{1+exp(\alpha_1)}, \frac{exp(\alpha_2)}{1+exp(\alpha_2)}\right] D \begin{bmatrix} \frac{exp(\alpha_1-\tau)}{[1+exp(\alpha_1-\tau)]^2} \\ 0 \end{bmatrix}$$

$$= \frac{0.5exp(\alpha_1)}{[1+exp(\alpha_1)]^2} * \frac{exp(\alpha_2-\tau_a)}{1+exp(\alpha_2-\tau_a)} + \frac{-0.5exp(\alpha_2)}{1+exp(\alpha_2)} * \frac{exp(\alpha_1-\tau_a)}{[1+exp(\alpha_1-\tau_a)]^2}$$

$$= 0.5\phi(\alpha_1)\Phi(\alpha_2-\tau_a) - 0.5\phi(\alpha_1-\tau_a)\Phi(\alpha_2)$$

$$\frac{\partial \theta_a}{\partial \alpha_2} = \left[\frac{\partial \gamma_1^{(0)}}{\partial \alpha_2}, \frac{\partial \gamma_2^{(0)}}{\partial \alpha_2}\right] D \begin{bmatrix} \gamma_1^{(a)} \\ \gamma_2^{(a)} \end{bmatrix} + \left[\gamma_1^{(0)}, \gamma_2^{(0)}\right] D \begin{bmatrix} \frac{\partial \gamma_1^{(a)}}{\partial \alpha_2} \\ \frac{\partial \gamma_2^{(a)}}{\partial \alpha_2} \end{bmatrix} + 0.5\left(\frac{\partial \gamma_2^{(0)}}{\partial \alpha_2} - \frac{\partial \gamma_2^{(a)}}{\partial \alpha_2}\right)$$

$$= \left[0, \frac{exp(\alpha_2)}{[1+exp(\alpha_2)]^2}\right] D \begin{bmatrix} \frac{exp(\alpha_1-\tau_a)}{1+exp(\alpha_1-\tau_a)} \\ \frac{exp(\alpha_2-\tau_a)}{1+exp(\alpha_2-\tau_a)} \end{bmatrix} + \left[\frac{exp(\alpha_1)}{1+exp(\alpha_1)}, \frac{exp(\alpha_2)}{1+exp(\alpha_2)}\right] D \begin{bmatrix} 0 \\ \frac{exp(\alpha_2-\tau_a)}{[1+exp(\alpha_2-\tau_a)]^2} \end{bmatrix}$$

$$+ 0.5\left(\frac{exp(\alpha_2)}{[1+exp(\alpha_2)]^2} - \frac{exp(\alpha_2-\tau_a)}{[1+exp(\alpha_2-\tau_a)]^2}\right)$$

$$= \frac{-exp(\alpha_2)}{[1+exp(\alpha_2)]^2} * \frac{exp(\alpha_1-\tau_a)}{1+exp(\alpha_1-\tau_a)} + \frac{0.5exp(\alpha_1)}{1+exp(\alpha_1)} * \frac{exp(\alpha_2-\tau_a)}{[1+exp(\alpha_2-\tau_a)]^2}$$

$$+ \frac{0.5exp(\alpha_2)}{[1+exp(\alpha_2)]^2} - \frac{0.5exp(\alpha_2-\tau_a)}{[1+exp(\alpha_2-\tau_a)]^2}$$

$$= 0.5\phi(\alpha_2-\tau_a)[\Phi(\alpha_1)-1] - 0.5\phi(\alpha_2)[\Phi(\alpha_1-\tau_a)-1]$$

$$\frac{\partial \theta_a}{\partial \tau_a} = \left[\frac{\partial \gamma_1^{(0)}}{\partial \tau_a}, \frac{\partial \gamma_2^{(0)}}{\partial \tau_a}\right] D \begin{bmatrix} \gamma_1^{(a)} \\ \gamma_2^{(a)} \end{bmatrix} + \left[\gamma_1^{(0)}, \gamma_2^{(0)}\right] D \begin{bmatrix} \frac{\partial \gamma_1^{(a)}}{\partial \tau_a} \\ \frac{\partial \gamma_2^{(a)}}{\partial \tau_a} \end{bmatrix} + 0.5\left(\frac{\partial \gamma_2^{(a)}}{\partial \tau_a} - \frac{\partial \gamma_2^{(a)}}{\partial \tau_a}\right)$$

$$= \left[\frac{exp(\alpha_1)}{1+exp(\alpha_1)}, \frac{exp(\alpha_2)}{1+exp(\alpha_2)}\right] D \begin{bmatrix} \frac{-exp(\alpha_1-\tau_a)}{[1+exp(\alpha_1-\tau_a)]^2} \\ \frac{-exp(\alpha_2-\tau_a)}{[1+exp(\alpha_2-\tau_a)]^2} \end{bmatrix} + \frac{0.5exp(\alpha_2-\tau_a)}{[1+exp(\alpha_2-\tau_a)]^2}$$

$$= \frac{0.5exp(\alpha_2)}{1+exp(\alpha_2)} * \frac{exp(\alpha_1-\tau_a)}{[1+exp(\alpha_1-\tau_a)]^2} - \frac{0.5exp(\alpha_1)}{1+exp(\alpha_1)} * \frac{exp(\alpha_2-\tau_a)}{[1+exp(\alpha_2-\tau_a)]^2}$$

$$+ \frac{0.5exp(\alpha_2-\tau_a)}{[1+exp(\alpha_2-\tau_a)]^2}$$

$$= 0.5\Phi(\alpha_2)\phi(\alpha_1-\tau_a) - 0.5\phi(\alpha_2-\tau_a)[\Phi(\alpha_1)-1]$$

# Appendix 2: Supplementary tables

Table S1.1: International Classification of Disease, Ninth and Tenth Revision (ICD-9 and ICD-10) diagnosis code for patients with alcohol abuse or dependence

| Description | ICD-9 codes | ICD-10 codes |
|---|---|---|
| Alcohol dependent | 303.00, 303.01, 303.02, 303.9, 303.9X | F10.20, F10.22, F10.22X, F10.23, F10.23X, F10.29 |
| Alcohol abuse | 305.00, 305.01, 305.02 | F10.10, F10.12, F10.12X, F10.13, F10.13X, F10.19 |
| Alcohol related organ disease | 571.0, 571.1, 571.2, 571.3, 357.5, 425.5, 535.3 | K70.XX, G62.1, I42.6, K29.2X |
| Remission | 305.03, 303.03 | F10.11, F10.21 |

Table S1.2: Healthcare common procedure coding system (HCPCS) and current procedural terminology (CPT) codes related to behaviour treatments for patients with alcohol abuse or dependence

| Description | HCPCS and CPT codes |
|---|---|
| Alcohol consultation | H0005, H0015, T1006, G0396, G0397 |
| No Specific consultation | H0004, 90804, 90805, 90806, 90807, 90808, 90809, 90810, 90811, 90812, 90813, 90814, 90815, 90816, 90817, 90818, 90819, 90821, 90822, 90823, 90824, 90826, 90827, 90828, 90829, 90847, 90849, 90853, 90857, 90875, 90876, 0074T |
| Other alcohol related behaviour treatment | H0001, H0003, H0006, H0007, H0021, H0022, H0026, H0027, H0028, H0029, H0047, H0050, H2034, H2035, H2036, T1007, T1011, T1012 |

Table S1.3: National drug codes (NDC) which could be used to treat patients with alcohol abuse or dependence

| Drug | Drug codes |
|---|---|
| Naltrexone | 00406117001, 00406117003, 16729008101, 16729008110, 47335032683, 47335032688, 51224020630, 51224020650, 68084029121, 68094085362, 68084029111, 00555090201, 00555090202, 00185003901, 51927437700, 38779088703, 38779088704 |
| Acamprosate | 00093535286, 00378633380, 60687012125, 68382056928, 68462043518 00258400060, 51079024106 |
| Disulfiram | 00054035613, 00054035625, 00054035713, 00054035725, 00093503501, 00093403601, 47781060730, 64980017103, 64980017203, 64980017101, 64980071202, 00378414001, 00378414101, 00603343221 |
| Topiramate | 00093015506, 00093015510, 00093721906, 00093721910, 00093722006, 00093733506 00093733506, 00093733606, 00093754006, 00093754010, 00378610105, 00378610191, 00378610205, 00378610291, 00378610305, 00378610391, 00378610591, 00781227660, 13668003105, 13668003160, 13668003205, 13668003260, 13668003305, 13668003360, 13668003405, 13668003460, 16252056860, 16252056960, 31722027805, 31722027810, 31722027860, 31722027905, 31722027910, 31722027960, 31722028005, 31722028010, 31722028060, 31722028105, 31722028110, 31722028160, 51079072620, 31079072720, 51079072820, 59762103001, 59762103101, 59762103201, 59762103301, 60429077010, 60505276006, 60505276106, 60505276206, 60505276306, 62756070713, 62756070786, 62756071013, 62756071086, 62756071113, 62756071186, 62756071213, 62756071286, 64376012101, 64376012110, 68084034211, 68084034401, 68084034411, 68084034521, 68382000414, 68382000514, 68382013805, 68382013814, 68382013905, 68382013914, 68382014005, 68382014014, 68382014105, 68382014114, 68462010810, 68462010860, 68462010910, 68462010960, 68462011010, 68462011060, 68462015310, 68462015360, 60505276005, 60505276105, 65862017360, 65862017460, 00832107130, 00832107430, 00832107530, 60505276205, 60505276305, 68084034201, 00832107230, 00832107315, 00832107330, 00832107415, 29300011710, 38779244308, 47335070713, 47335071013, 47335071013, 47335071086, 47335071113, 47335071186, 47335071213, 47335071286, 51079072601, 51927467100, 69097012203, 69097012212, 69097012215, 69097012303, 69097012312, 69097012315, 69097012403, 69097012412, 69097012415, 69097012503, 69097012512, 69097012515, 29300011610, 51552120605, 69097081603, 69097081615, 69097081703, 69097081715, 69097081803, 69097081815, 69097081903, 29300011505, 29300011510, 29300011605, 29300011616, 29300011705, 29300011805, 29300011816, 29300011516, 29300011716, 00395815156, 29300011810, 38779244305, 69097081915 |

# CURRICULUM VITA

**NAME** Huirong Hu

**EDUCATION**

**Ph.D. Candidate**, *University of Louisville*, Louisville KY. 2019-2023

**M.S. in Biostatistics**, *University of Louisville*, Louisville KY, 2017-2019

**B.S. in Applied Psychology**, *Anhui University of Chinese Medicine*, Anhui, 2010-2014

**EMPLOYMENT**

**Intern**, Medpace, Cincinnati, Ohio, 2022

**Research Assistant**, Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, 2020-2022.

**PUBLICATIONS**

H Hu, R Mitra, M Kong (2023).**Statistical Methods for Assessing Treatment Effects on Ordinal Outcomes Using Observational Data.** *(In preparation)*

H Hu, KB Kulasekera, Q Zheng, M Kong (2023). **Selecting optimal treatment for survival outcomes using observational data.** *(In preparation)*

H Hu, R Mitra, Y Han, S Pal, H Huang, V Vatsalya, J McClain, KB Kulasekera, M Kong (2022). **Prevalence and Treatment for Alcohol Use Disorders Based on Kentucky Medicaid 2012-2019 Datasets.** *Journal of Alcoholism and Drug Dependence. 10(4).*

M Sagaram, AJ Royer, H Hu, A Rajhans, R Parthasarathy, SS Krishnasamy, SP Mokshagundam, M Kong, ML Schwandt, D Parajuli, MC Cave and V Vatsalya. (2022). **Illustration of Gut–Thyroid Axis in Alcohol Use Disorder: Interplay of Gut Dysfunction, Pro-Inflammatory Responses, and Thyroid Function.** *Cells 11(19): 3100.*

JB Warner, KH Zirnheld, H Hu, A Floyd, M Kong, CJ McClain, and IA Kirpich (2022). **Analysis of Alcohol Use, Consumption of Micro-and Macronutrients, and Liver Health in the 2017-2018 National Health and Nutrition Examination Survey**. *Alcoholism: Clinical and Experimental Research* 46(11): 2025-2040.

Y Han, H Huang, R Mitra, H Hu, S Pal and CJ McClain, KB Kulasekera, M Kong (2022). **Prevalence and Treatment Utilization of Patients Diagnosed with Depression and Anxiety Disorders Based on Kentucky Medicaid 2012-2019 Datasets.** *Journal of Depression and Anxiety 11: 459.*

**PRESENTATIONS**

Assessing Treatment Effects For Patients With Alcohol Abuse Or Dependence Using Kentucky Medicaid Data, *Kentucky Chapter of the American Statistical Association Spring Meeting*, April, 2022, Lexington, Kentucky.

Statistical Methods For Assessing Treatment Effects On Ordinal Outcomes Using Observational Data, *ENAR Spring 2023*, March, 2023, Nashville, Tennessee.

Statistical Methods For Assessing Treatment Effects On Ordinal Outcomes Using Observational Data, *University of Louisville Biostatistics Seminar*, March, 2023, Louisville, Kentucky.

## HONORS AND AWARDS

University Fellowship, *Graduate School, University of Louisville*, August 2019, - June, 2021 and August, 2022 - July, 2023

Graduate Research Assistantship, *Department of Bioinformatics and Biostatistics, University of Louisville*, July, 2020- June, 2021

## CERTIFICATES

SAS Certified Base Programmer for SAS 9.

SAS Certified Advanced Programmer for SAS 9.