ON DEVELOPMENT OF STATISTICAL LEARNING METHODS IN PRECISION MEDICINE

Siyeon Kim

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2023

Approved by:

Michael R. Kosorok

Kelli Allen

Eric B. Laber

Feng-Chang Lin

Donglin Zeng

ABSTRACT

Siyeon Kim: On Development of Statistical Learning Methods
in Precision Medicine
(Under the direction of Michael R. Kosorok)

Precision medicine is an area that seeks to maximize clinical effectiveness by assigning treatment regimes tailored to individuals. In this dissertation, we present three topics that advance the methods and applications in the field of precision medicine.

The first topic introduces a novel methodology termed random forest informed tree-based learning to discover underlying patient characteristics associated with differential improvement in knee osteoarthritis (OA) symptoms and to identify the individualized treatment regime (ITR) among three available treatments. The proposed algorithm suggests decision rules that divide participants into subgroups based on their characteristics. In our analysis, the estimated treatment rule yielded greater improvements in OA symptoms that could ultimately guide patients toward suitable treatment strategies.

In the second topic, we propose a doubly robust estimator for patient-specific utilities and ITRs based on the inverse reinforcement framework from Luckett et al. (2021). This framework optimizes patient-utility for two outcomes by leveraging experts' decisions on observational data. The suggested doubly robust estimator guarantees consistency even when incorrect outcome models or incorrect propensity score models are applied, alleviating the need for exact formulation of the outcome model and improving the previous estimator. We also present asymptotic distributions for the estimators of boundary and utility functions using the newly developed indexed argmax theorem, which can be used for deriving weak convergence of M-estimators with multiple layers.

Lastly, we suggest an estimator for utilities when there are more than two treatments. Specifically, we utilize stabilized direct learning to estimate ITRs. Subsequently, we apply the inverse reinforcement framework once again to obtain an estimator for a composite outcome and the balance of the two outcomes. Also, the proposed estimator for utilities considers the heterogeneity in the variance of patients, leveraging the benefits of stabilized direct learning.

To my family.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to everyone who has supported my Ph.D. journey. First and foremost, I am deeply thankful to my advisor, Dr. Michael Kosorok, for his support and guidance throughout the entire process. I am also grateful to my committee members, Dr. Donglin Zeng, Dr. Feng-chang Lin, and Dr. Eric Laber, for their insightful and invaluable feedback, which significantly contributed to the improvement of my dissertation.

I would like to extend my appreciation to the collaborators at the Thurston Arthritis Center. Especially, I would like to thank Dr. Kelli Allen, who collaborated with me on my first project and provided invaluable guidance. I am also grateful to Dr. Becki Cleveland, Liubov Arbeeva, Carolina Alverez, and Dr. Amanda Nelson. Furthermore, I would like to thank the members of The Precision Health and AI Research Lab for their constant encouragement.

Lastly, I want to express my gratitude to my family for their unconditional support. I would also like to thank Sangyoon Yi and Miso Kim for their support and kindness throughout this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIPWE	Augmented Inverse Probability Weighted Estimator
BMI	Body Mass Index
CART	Classification and Regression Trees
EARL	Efficient Augmentation and Relaxation Learning
GUIDE	Generalized, Unbiased, Interaction Detection and Estimation
HTE	Heterogeneous Treatment Effects
ITR	Individualized Treatment Rule
OA	Osteroarthritis
PATH-IN	Physical Therapy vs Internet-Based Exercise Training for Knee osteoarthritis
PT	Physical Therapy
QUINT	Qualitative INteration Trees
RF	Random Forest
SD	Standard Deviation
SD-learning	Stabilized Direct Learning
VF	Value Function
WT	Waitlist
ZOM	Zero-Order Model

INTRODUCTION

Precision medicine aims to assign the optimal treatment customized for each patient, leveraging the heterogeneity in each patient's characteristics. Due to its effectiveness, precision medicine research has undergone substantial development, incorporating various machine learning methodologies since its advent. Additionally, the announcement of the Precision Medicine Initiative by President Obama in 2015 played a significant role in advancing precision medicine research, elevating it to an important national agenda and accelerating its progress. In this paper, we introduce methods for estimating optimal treatment regimes suited for different settings, which may also potentially contribute to the advance of precision medicine research.

In the first topic, we explore characteristics underlying differential improvement among participants in the Physical Therapy vs Internet-Based Exercise Training for Knee osteoarthritis (PATH-IN) trial, which compared standard physical therapy (PT) with internet-based exercise training (IBET), both relative to a usual care/ waitlist control group (WT). While current machine learning methods do allow for individual features to be taken into account when determining a therapy, the treatment assignment rules are not always immediately interpretable in terms of demonstrating which characteristics lead to specific treatments. To resolve this issue, we develop a unique machine learning approach to obtain the optimal treatment rule in the context of the PATH-IN study. The new algorithm, Random Forest (RF) informed Tree-based Learning, which obtains split points by random forests, improves the interpretability by revealing inherent mechanisms of treatment and patients' characteristics, enabling clinicians to understand the mechanisms easily.

In the second topic, we explore the approach that could represent a patient-specific composite outcome for identifying individualized treatment regimes (ITRs) when multiple outcomes are present. We implement an inverse reinforcement learning framework introduced in Luckett et al. (2021). However, the outcome model, which contains the utility, includes a complex formulation, hampering the accuracy of the estimator. To address this problem, we employ a doubly robust estimator from efficient augmentation and relaxation learning (Zhao et al., 2019) for the estimation of ITRs in order to protect the consistency of the utility estimator from the misspecification. We also prove the doubly robust consistency and limiting distribution of the estimator. We present simulation studies that support these theories.

In the third topic, we extend the estimation of composite outcomes to the multi-armed setting. We implement stabilized direct learning to estimate the boundary function when there are more than two outcomes. We provide a detailed formulation of the estimator for the boundary when composite outcomes are used.

The remaining chapters of the dissertation are structured as follows. We first provide a literature review of the methods used in Chapters 1-3. Topics 1 (Random forest informed tree-based learning), 2 (Doubly robust estimation and inference of patient-specific utility functions), and 3 (Estimation of composite outcomes in multi-treatment setting) are discussed in detail in Chapters 1–3. We conclude this paper with the technical details of each chapter.

LITERATURE REVIEW

This chapter provides an overview of the background and development of statistical methods and machine learning techniques that are pertinent to the methods described in Chapters , 1.4, and 2.6. In addition, the definitions and concepts that are necessary to comprehend the proposed methods described in the following chapters are presented here. Random-forest informed tree-based learning, Doubly-robust estimation and inference of utility functions in a two-outcome setting, and estimation of composite outcome in the multi-treatment setting are the three subsections of this review that correspond to each of the three precision medicine, deep learning, and survival modeling, respectively. In addition, we provided a more in-depth description of the pseudo-likelihood framework, as well as efficient augmentation and relaxation learning, which are two of the most important themes covered in Chapter 2.

Random Forest Informed Tree-based Learning

Ever since their introduction, tree-based approaches have found widespread use in the fields of classifications and regressions due to the ease of interpretation they offer, despite the presence of nonlinearity in the data they analyze. Classification and Regression Trees, also known as CART, are one of the earliest and most well-known algorithms. This algorithm iteratively divides the data into binary regions that are disjoint from one another. CART has attracted attention due to its straightforward and understandable structure. (Breiman et al., 1984). Along with its functionality, the CART has expanded its range of coverage to include a variety of data types. Some examples include survival data (Davis and Anderson, 1989; Gordon and Olshen, 1985; LeBlanc and Crowley, 1992; Therneau et al., 1990), longitudinal data (Abdolell et al., 2002; Segal, 1992), data for generalized linear model (Ciampi, 1991), and multiresponse outcomes

(Zhang, 1998). In addition to CART, there are a variety of algorithms that can be used to construct a tree. Some examples of these algorithms include CHi-squared Automatic Interaction Detector (Kass, 1980), C4.5 (Quinlan, 1993), Fast and Accurate Classification Tree Loh and Vanichsetakul (1988), Classification Rule with Unbiased Interaction Selection and Estimation (Kim and Loh, 2001), Qualitative INteraction Trees (QUINT, Dusseldorp and Van Mechelen (2014)) and Generalized, Unbiased, Interaction Detection and Estimation (Loh, 2009).

However, due to the fact that splits in a single decision tree might be affected by a peripheral disturbance in the data, a single tree could cause overfitting, which would result in the model being unstable and producing mediocre predictions. As a result, there have been efforts made to gather multiple trees and grow a forest in the hopes that this may solve these issues. Bagging is a technique that aggregates many classifiers that have been constructed using bootstrapped data and combines them into one in order to decrease the variance, resulting in an enhanced prediction compared to using a single tree (Breiman, 1996). Boosting is a concept that trains weak learners sequentially in the direction of lowering bias by assigning weights adaptively to each of the trained trees (Schapire, 1990). AdaBoost (Freund and Schapire, 1997), gradient boosting (Friedman, 2001), XGBoost (Chen and Guestrin, 2016), and LightGBM (Ke et al., 2017) are all variations of boosting. Super learners are an ensemble that uses cross-validation to determine the optimal weights for each individual learner (Van der Laan et al., 2007).

Random forests may be the most popular ensemble of trees among the ensemble algorithms because they prevent overfitting by growing multiple trees randomly based on CART (Breiman, 2001). A random forest is composed of trees that have been constructed from bootstrapped pseudo data (bagging) that is recursively split with randomly chosen variables to minimize the impurity in each node. Predictions are made with a new data point for every tree in the forest, and the forest determines the final prediction by majority voting (classification) or averaging (regression), which results in robustness to noise. Random forests have been actively employed in the field of survival analysis and precision medicine (Zhu and Kosorok, 2012; Cui et al., 2017; Cho et al., 2020, 2021).

However, along with other ensemble methods, one major drawback of random forests is its lack of interpretability which makes it clinicians difficult to reveal the exact effects of certain variables in outcomes, impeding applications in biomedical data. Nevertheless, there have been attempts to develop methods that could identify crucial factors that influence the outcomes and predictions in random forest literature. Meinshausen (2010) introduces Node harvest that aggregates nodes from a random forest and determines the right node with weights calculated by quadratic programming. Bénard et al. (2021) builds a rule by frequency of appearance of variables in the process of random forest modeling. In respect of causal inference, heterogeneous treatment effects (HTE) have been estimated using random forest. Wager and Athey (2018) developed a causal forest, built by causal trees (Athey and Imbens, 2016). It estimates HTE by allowing the data to adaptively determine the nearest neighbor and provide asymptotic normality of the estimator of (HTE). In the line of the causal forest, Athey et al. (2019) suggested a generalized random forest to estimate HTE with instrumental variables. Also, Oprescu et al. (2019) proposed the orthogonal random forest, which leverages Neyman-orthogonality to reduce estimation error in generalized random forests. Cui et al. (2020) designs a causal survival forest for estimating HTE in right-censored data and longitudinal data.

In precision medicine, interpretability could be more crucial for understanding the underlying mechanism of treatment interacting with certain characteristics of patients and subgrouping patients with certain treatments. Hence, there have been efforts to develop a model that is understandable in assigning optimal personalized treatments to individuals. Kallus (2017) recursively partition the observational data by introducing a new impurity measure for personalization and building a personalization tree. There have also been approaches that provide interpretable treatment rules or by leveraging nonparametric algorithms, which yield good predictions. A decision list of “if-then” statements of treatment rules is another formulation of interpretable ITRs (Zhang et al., 2015, 2018). Building these list-based rules includes maximizing Q-functions which have freedom in modeling, allowing nonparametric models such as kernel ridge regression or random forests to be helpful in accuracy. Yadlowsky et al. (2021) suggested an estimator for the conditional average treatment effect as the ratio of expected potential outcomes. The estimator borrows the advantage of nonparametric methods to obtain the correct conditional expectation by the doubly robust property and then obtain the coefficients by refitting the regression.

In this sense of leveraging nonparametric models to identify factors that affect outcomes, we developed a methodology that builds a tree by utilizing random forests for splitting rules in the first part of this dissertation. Also, we applied it to the patients with knee osteoarthritis (OA) who have differential effects in treatments and obtained treatment rules that maximize the outcomes of patients.

Doubly-robust estimation and inference of utility functions in the two-outcome setting

In the field of precision medicine, there are generally two distinctive approaches to methods for estimating ITRs. (Kosorok and Laber, 2019). One approach is the model-based approach which estimates ITRs in a two-step process that first estimates an outcome model of treatments and covariates and then infers a personalized treatment that delivers the best outcome for each patient. Examples of in this approach include g-estimation (Robins, 1989, 1997), Q-learning (Murphy, 2005; Qian and Murphy, 2011; Zhao et al., 2011; Goldberg and Kosorok, 2012; Laber et al., 2014; Schulte et al., 2014), and A-learning (Murphy, 2003; Blatt et al., 2004; Robins, 2004; Moodie et al., 2007; Fan et al., 2016; Shi et al., 2018). Nonetheless, one weakness of this approach is that the performance of estimators highly depends on the accuracy of the postulated outcome model. This significant dependency typically results in significant discrepancies between estimated ITRs and optimal ITRs, particularly when the true outcome models are complicated. An alternative approach is the classification-based approach or the direct-search approach, which focuses on obtaining ITRs that maximize the expectation of potential outcomes themselves, reducing the need to accurately specify outcome models. Examples of this approach include outcome weighted learning, which formulates estimating ITR as a weighted classification problem and employs a convex loss as a surrogate to an 0-1 loss (Zhao et al., 2012, 2015; Zhou et al., 2017; Liu et al., 2018), V-learning for infinite horizon (Luckett et al., 2019), and robust value-search estimator (Zhang et al. (2012, 2013)).

However, the aforementioned methods mostly focus on the single-outcome scenario, and in a multiple-outcome setting, a utility is needed that summarizes multiple outcomes to a scalar outcome. There have been efforts to build patient utilities in various ways. Murray et al. (2016) devised a randomized trial design that is based on a physician-derived utility.

We examine inverse reinforcement learning employed in further detail. The goal of inverse reinforcement learning is to derive reward functions by applying observed optimal policy (Ng et al., 2000). In Luckett et al. (2021), the decisions of clinicians are assumed to be optimal, and the personalized utility function of two outcomes and the accompanying ITRs are then estimated. In the suggested method, Efficient Augmentation and Relaxation Learning (EARL, Zhao et al. (2019)) is employed to estimate ITRs instead of the Q-function in Luckett et al. (2021). EARL searches for the boundary function of ITRs on an augmented inverse probability weighted estimator (AIPWE) by replacing 0-1 loss to convex surrogates to reduce the computational burden (Freund and Schapire, 1998; Bartlett et al., 2006). Additionally, EARL benefits from having the doubly-robustness.

An estimator is doubly robust if it is guaranteed to be consistent when at least one of a propensity score model or an outcome model is correctly specified. Robins et al. (1994) introduced an augmented inverse probability weighted estimator useful in missing data. It was shown by Scharfstein et al. (1999) that this estimator is doubly robust. Further extension and investigation of doubly robust estimator were given by Robins and Rotnitzky (2001); Lunceford and Davidian (2004); Bang and Robins (2005); Neugebauer and van der Laan (2005); Kang and Schafer (2007). In the suggested method, the estimators for the utility and the probability of the correct treatment assignment achieve the doubly robust property, which is transferred from EARL.

From a theoretical point of view, the second topic proposes a new advancement in M-estimation theory. M-estimators are defined as data-dependent functions that nearly maximize objective functions which are calculated from data. The *argmax* theorem, which is the center of the M-estimation theory, states that limits of M-estimators converge weakly to the argmax of the limiting process (Kosorok, 2008). There has been extensive literature on M-estimation theory and its expansions. Kim and Pollard (1990) provides cube-root asymptotic results for statistics with certain sufficient conditions. Ma and Kosorok (2005) and Kosorok and Song (2007) present weak convergence results in infinite dimensional settings. Seijo and Sen (2011) introduces an *argmax* theorem when objective functions converge to a limiting process that maximizes at multiple locations with some assumptions.

Next, we provide detailed reviews of the pseudo-likelihood framework in Luckett et al. (2021) and EARL in Zhao et al. (2019).

The pseudo-likelihood estimation for the utility and the probability of correct treatment assignment

In order to estimate the utility function of two outcomes, the pseudo-likelihood approach was introduced in Luckett et al. (2021). Let $(\boldsymbol{X}_i, A_i, Y_i, Z_i),\ i = 1, \cdots, n$ be the independent and identically distributed realizations of $(\boldsymbol{X}, A, Y, Z)$, where $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^p$ are patient covariates, $A \in \mathcal{A} = \{-1, 1\}$ is the assigned intervention, and $Y, Z \in \mathcal{Y} \times \mathcal{Z} \subseteq \mathbb{R}^2$ are scalar outcomes, each coded so that higher values are better. For a treatment assignment function, let $d : \mathcal{X} \mapsto \{1, -1\},\ d \in \mathcal{D}$, which allocates $d(\boldsymbol{x})$ to patients who have $\boldsymbol{X} = \boldsymbol{x}$ as covariates. Let a function f be a measurable function where $f : \mathcal{X} \mapsto \mathbb{R}$ and $\text{sgn}(f(\boldsymbol{X})) = d(\boldsymbol{X})$, where $\text{sgn}(t) = 1$ when $t \geq 0$ and $\text{sgn}(t) = -1$ when $t < 0$.

$u(y, z; \boldsymbol{x}, w, \theta) = \omega_\theta(\boldsymbol{x})y + \{1 - \omega_\theta(\boldsymbol{x})\}z$ is defined as a utility function where $\omega_\theta : \mathcal{X} \mapsto [0, 1]$ for each utility parameter $\theta \in \Theta$. $\omega_\theta(\boldsymbol{x}) = \text{expit}(\boldsymbol{x}^T\theta)$ is assumed, where $\text{expit}(t) = e^t/(1 + e^t)$. Also, $d^*_\theta(\boldsymbol{X})$ is defined as the optimal treatment for each $\theta \in \Theta$. Let the probability of assigning $d^*_\theta(\boldsymbol{X})$ to each patient be $\Pr\big\{A = d^*_\theta(\boldsymbol{X})|\boldsymbol{X}\big\} = \text{expit}(\boldsymbol{X}^T\beta)$. It is assumed that there exist the densities $f(Y, Z|\boldsymbol{X}, A)$ and $f(\boldsymbol{X})$ so that we can factor the likelihood for (θ, β) into

$$f(\boldsymbol{X}, A, Y, Z) = f(Y, Z|\boldsymbol{X}, A)f(\boldsymbol{X})\frac{\exp\big[\boldsymbol{X}^T\beta 1\big\{A = d^*_\theta(\boldsymbol{X})\big\}\big]}{1 + \exp(\boldsymbol{X}^T\beta)}.$$

This leads to the pseudo logistic regression likelihood

$$\hat{\mathcal{L}}_n(\theta, \beta) \propto \prod_{i=1}^{n} \frac{\exp\big[\boldsymbol{X}_i^T\beta 1\big\{A_i = \hat{d}_{n,\theta}(\boldsymbol{X}_i)\big\}\big]}{1 + \exp(\boldsymbol{X}_i^T\beta)},$$

where $\hat{d}_{n,\theta}$ is an estimator for the optimal treatment regime d^*_θ.

The pseudo-likelihood incorporates the parameters for the utility and the parameters for the probability of assigning optimal treatment into one likelihood for the inverse reinforcement learning framework. For fixed θ, $\hat{\beta}$ is estimated using the logistic regression after obtaining $\hat{d}_{n,\theta}$ for each individual.

Efficient augmentation and relaxation learning

Let $Q_Y(\boldsymbol{X}, a) = \mathbb{E}[Y|\boldsymbol{X}, A = a]$. The paper assumes the causal assumptions (Hernán and Robins, 2010). Then, the optimal treatment regime is defined as $d_Y^*(\boldsymbol{X}) = \max_{a \in \mathcal{A}} Q_Y(\boldsymbol{X}, a)$.

In this approach, the objective is to estimate d_Y^* by searching $d \in \mathcal{D}$ that maximizes $V(d)$. This could be achieved by expressing using inverse probability,

$$V(d) \equiv \mathbb{E}\big[\frac{YI(A = a)}{\pi(a; \boldsymbol{X})}\big],$$

where $\pi(a; \boldsymbol{X}) \equiv P(A = a|\boldsymbol{X})$ is the propensity score. However, since the estimation of $V(d)$ only includes a subset of the data, the estimator of $V(d)$ results in a potentially large variance. Therefore, an alternative approach to estimate an optimal treatment using AIPWE was used in Zhao et al. (2019). Specifically, if there exists a function f in a Hilbert space $\mathcal{F}$ that satisfies $d(\boldsymbol{X}) = \text{sgn}(f(\boldsymbol{X}))$, the value, which is expected outcome when assumed that individuals assumed treatments by the regime d, is

$$V^{\text{AIPWE}}(d) = \mathbb{E}\big[\frac{YI\{A = d(\boldsymbol{X})\}}{\pi(d(\boldsymbol{X}); \boldsymbol{X})} - \frac{I\{A = d(\boldsymbol{X})\} - \pi(d(\boldsymbol{X}); \boldsymbol{X})}{\pi(d(\boldsymbol{X}); \boldsymbol{X})} Q\{\boldsymbol{X}, d(\boldsymbol{X})\}\big].$$

Denote $\mathbb{E}_n g = n^{-1} \sum_{i=1}^n g(X_i)$. Then, the estimator of $V^{\text{AIPWE}}(d)$ is $\hat{V}^{\text{AIPWE}}(d) = \mathbb{E}_n\big[\frac{YI\{A=d(\boldsymbol{X})\}}{\hat{\pi}(d(\boldsymbol{X}); \boldsymbol{X})} - \frac{I\{A=d(\boldsymbol{X})\} - \hat{\pi}(d(\boldsymbol{X}); \boldsymbol{X})}{\hat{\pi}(d(\boldsymbol{X}); \boldsymbol{X})} \hat{Q}\{\boldsymbol{X}, d(\boldsymbol{X})\}\big]$, where $\hat{\pi}(a; \boldsymbol{X})$ and $\hat{Q}(\boldsymbol{X}, a)$ are estimators of $\pi(a; \boldsymbol{X})$ and $Q(\boldsymbol{X}, a)$, respectively.

EARL optimizes the boundary function $f^*(\boldsymbol{X}) \in \mathcal{F}$ such that $d^*(\boldsymbol{X}) = \text{sgn}(f^*(\boldsymbol{X}))$ by maximizing the V^{AIPWE}. Since maximizing the value is equivalent to minimizing the risk, EARL minimizes a sum of weighted misspecification rates. However, in order to avoid the discontinuity of 0-1 loss, EARL replaces an indicator function with one of the following convex surrogates; $\phi(t) = \max(1 - t)$ for hinge loss; $\phi(t) = e^{-t}$ for exponential loss; $\phi(t) = \log(1 + e^{-t})$ for logistic loss; or $\phi(t) = \max(1 - t, 0)^2$ for squared hinge loss.

Therefore, when replaced by one of the suggested surrogates, it is proposed that

$$\tilde{f}^{\lambda_n}_{n,\phi} = \arg\inf_{f\in\mathcal{F}} \mathbb{E}_n\Big[|\hat{W}_1|\phi\big\{\text{sgn}(\hat{W}_1)f(\boldsymbol{X})\big\} + |\hat{W}_{-1}|\phi\big\{-\text{sgn}(\hat{W}_{-1})f(\boldsymbol{X})\big\} + \lambda_n\|f\|^2\Big],$$

where $\hat{W}_{a,\theta}$ is the estimator of

$$W_a(Y) = \frac{YI(A=a)}{\pi(a;\boldsymbol{X})} - \frac{I(A=a)-\pi(a;\boldsymbol{X})}{\pi(a;\boldsymbol{X})}Q(\boldsymbol{X},a)$$

for $a \in \{-1,1\}$ which substitutes $\hat{\pi}(a;\boldsymbol{X})$ for $\pi(a;\boldsymbol{X})$ and $\hat{Q}(\boldsymbol{X},a)$ for $Q(\boldsymbol{X},a)$, and $\lambda_n \to 0$ as $n \to \infty$.

In Zhao et al. (2019), the sample splitting technique is applied, which ensures that samples used for $\hat{\pi}(\boldsymbol{X})$ and $\hat{Q}(\boldsymbol{X})$ are not reused to optimize $f^*(\boldsymbol{X})$ at the same time. Let n be the number of samples, and assume that the samples are partitioned evenly at random into J disjoint groups. Let $I_1,\cdots,I_J$ be the sets of indices of the samples in each of J groups, and let $\{n_j : j = 1,\cdots,J\}$ be the set of the numbers of the samples in each group I_j. If there are remaining observations, we randomly distribute them to some of the J groups so that $\frac{n_j}{n}$ converges to a fixed constant n^*, as n increases. For each group I_j, we estimate $\hat{\pi}_j(a;\boldsymbol{X})$ for $\pi(a;\boldsymbol{X})$, and $\hat{Q}_j(\boldsymbol{X},a)$ for $Q(\boldsymbol{X},a)$ using the samples $\{(\boldsymbol{X}_i,A_i,Y_i) : i \in I_j\}$, and calculate $\hat{f}^{\lambda_n,(j)}_{n,\phi}$ use the samples in $I_{(-j)}$, where $I_{(-j)} = \{1,\cdots n\} \setminus I_j$. When this sample splitting technique is used the estimator for $f^*(\boldsymbol{X})$ is

$$\hat{f}^{\lambda_n}_{n,\phi} = \frac{1}{J}\sum_{j=1}^{J}\hat{f}^{\lambda_n,(j)}_{n}, \text{ and}$$

$$\hat{f}^{\lambda_n,(j)}_{n,\phi} = \arg\inf_{f\in\mathcal{F}} \mathbb{E}^{(-j)}_n\Big[|\hat{W}_{1j}|\phi\big\{\text{sgn}(\hat{W}_{1j})f(\boldsymbol{X})\big\} + |\hat{W}_{-1j}|\phi\big\{-\text{sgn}(\hat{W}_{-1j})f(\boldsymbol{X})\big\} + \lambda_{nj}\|f\|^2\Big],$$

where $\hat{W}_{aj} = \frac{YI(A=a)}{\hat{\pi}_j(a;\boldsymbol{X})} - \frac{I(A=a)-\hat{\pi}_j(a;\boldsymbol{X})}{\hat{\pi}_j(a;\boldsymbol{X})}\hat{Q}_j(\boldsymbol{X},a)$ for $a \in \{1,-1\}$, and $\mathbb{E}^{(-j)}_n g = \frac{1}{n-n_j}\sum_{i\in I_{-(j)}} g(X_i)$. $\lambda_{nj}\|f\|^2$ provides L_2 penalization.

Estimation of composite outcome in multi-treatment setting

In addition to the model-based approach and policy-search approach, Tian et al. (2014) proposed a novel approach for estimating ITRs by employing a modified covariate method that employs regression to directly estimate an interaction of treatment and covariate. In other words, it directly estimates the boundary function and causal treatment effect by regressing the outcome on modified covariates multiplied by one-half of the treatment assignment. Qi and Liu (2018) named this approach D-learning and expanded to estimating ITRs with K categories ($K > 2$) by pairwise decision functions. Qi et al. (2020) suggested angle-based direct learning (AD-learning) borrowing the angle-based approach from Zhang and Liu (2014) to construct the boundary function of optimal treatments in K-treatment setting, which could be utilized to the variety of outcomes including survival, or binary outcome with theoretic guarantees. Meng and Qiao (2020) proposed robust direct learning (RD-Learning), which satisfies doubly robust consistency by using residuals instead of the outcomes in D-Learning. Lastly, Shah et al. (2022) introduced stabilized direct learning (SD-Learning) that leverages the heteroscedasticity possibly residing in the error of treatment and covariates. It improves the efficiency of the estimator by obtaining the estimator for the residual variance with nonparametric machine learning algorithms and re-estimate the boundary function after adjusting for the weights with the estimated residual variances. Also, in the K treatment setting, it improved AD-learning by suggesting analogous residual reweighting and proposed the estimator for the boundary of multi-category ITRs.

We provide some more context about the classification technique for multiple categories used for the multi-armed optimal ITRs in Qi et al. (2020) and Shah et al. (2022). For a binary classification problem, there has been a vast volume of literature using large-margin classifiers. Support vector machines (Vapnik, 1999), AdaBoost (Freund and Schapire, 1997), LogitBoost (Friedman et al., 2000), and import vector machines (Zhu and Hastie, 2001) are examples of the margin-based binary classifiers. Zhang and Liu (2014) introduced a large-margin approach in solving multi-category problems, in contrast to literature that has added a constraint that states that elements of K dimensional maps $f(x) \in \mathbb{R}^K$ sum to zero (Wang and Shen, 2007; Liu and Yuan, 2011; Zhang and Liu, 2013). The method by Zhang and Liu (2014) implicitly satisfied this inefficient constraint by employing a K simplex vertices in $\mathbb{R}^{K-1}$, investigated in Lange and Tong Wu (2008). Then, it predicts the label that minimizes the angle of a function and the vector of K vertices, which is equivalent to maximizing the margin.

In the third topic of the dissertation, we apply the SD-Learning from Shah et al. (2022) to the inverse reinforcement learning framework in Luckett et al. (2021) to obtain the boundary for ITRs with multiple treatments and optimal utilities with two outcomes.

CHAPTER 1: RANDOM FOREST INFORMED TREE-BASED LEARNING

1.1 Introduction

Knee osteoarthritis (OA) is one of the most common causes of pain and disability (United States Bone and Joint Initiative, 2020). Exercise-based therapies, including physical therapy (PT), are considered core treatments for patients with knee OA (Kolasinski et al., 2020; Bannuru et al., 2019). However, patients vary considerably in their level of improvement following exercise-based interventions, and very little is known about drivers of this variability. This limits our ability to make patient-centered recommendations about specific exercise interventions.

To date there has been little application of precision medicine-based machine learning in the context of OA management. One recent study found that in the context of a clinical trial comparing exercise, dietary weight loss and their combination, the combination intervention was optimal for most participants, but further improvement could be obtained through assignment to diet only for a subgroup of participants characterized by high baseline weight or low waist circumference, without a history of myocardial infarction Jiang et al. (2020). In this research, we add to this literature by exploring characteristics underlying differential improvement among participants in the Physical Therapy vs Internet-Based Exercise Training for Knee osteoarthritis (PATH-IN) trial, which compared standard physical therapy (PT) with internet-based exercise training (IBET), both relative to a usual care/ wait list control group (WT). Previously, there have been applications with QUINT, a sequential partitioning method, and GUIDE, a regression tree approach to evaluate heterogeneity of treatment effects in at the short-term follow-up time point (4-months) in PATH-IN (Coffman et al., 2021). We now extend this work by focusing on longer-term (12-month) outcomes, which is important for understanding maintenance of treatment effects.

Another way in which we extend prior work in this area is through development of a novel machine learning approach to obtain the optimal treatment rule in the context of the PATH-IN study. Although existing machine learning approaches enable individual characteristics to be reflected in the assignment of a treatment, they often lack interpretability as mentioned in the literature review. Specifically, many precision medicine approaches do not reveal the mechanism underlying differential improvement, since they focus on prediction and often result in decision rules generated by complicated interactions between factors.

To address these limitations, we developed a new machine learning algorithm that produces mechanistic decision rules that distinguish between subgroups of patients. This new algorithm, Random Forest (RF) informed Tree-based Learning, enables the final decision rule (regarding optimal treatment assignment) to determine the patient characteristics that most strongly influence the outcome and identify the thresholds of those characteristics to split the patients for assignment. In an iterative fashion, the algorithm identifies a subgroup of patients that could most benefit from a specific treatment and searches for more detailed rules consisting of successively finer subgroups of patients, in pursuit of the largest average benefit for the target population. We applied this methodology to data from the PATH-IN trial and obtained decision rules regarding the treatment from which each patient may expect the greatest improvement in OA symptoms and function at 12-month follow-up. We also compare the performance of the suggested method with LIST-based approach, another interpretable precision medicine tool.

1.2 Methods

1.2.1 PATH-IN Trial

The PATH-IN trial (Trial Registration: NCT02312713) included 350 participants with symptomatic knee OA; details of the participant eligibility criteria and other trial methods, as well as main trial outcomes, have been published previously (Williams et al., 2015; Allen et al., 2018). Briefly, participants were randomized to standard PT, IBET, or WT, in a 2:2:1 ratio, respectively. Participants in the PT group received up to 8 individual in-person treatment sessions within 4 months. Participants in the IBET arm received access to the online program for the full 12-month intervention period. Participants in the WT group did not receive PT or IBET during the study but were offered two PT visits and access to IBET following the 12-month assessments. For the fully study sample at 12-month follow-up (the time point of interest for this study), IBET was non-inferior to PT but neither PT nor IBET were superior to WT for the primary outcome, the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (Allen et al., 2018). This study was approved by the Institutional Review Boards of the University of North Carolina at Chapel Hill (UNC; #14-1331) and Duke University Medical Center (#00055318). Recruitment for the trial occurred from November 2014 to February 2016, and follow-up assessments were completed in February 2017.

1.2.2 Overview of Machine Learning Approach Used for the Estimation of Treatment Regimes

These exploratory analyses aimed to discover the features of patients that resulted in differential improvement within the PT, IBET, and WT study arms, particularly at 12 months. We first considered other established machine learning methods to address this question, including RF (Breiman, 2001) and LIST-based methods embedded with both kernel ridge regression and RF (Zhang et al., 2018), which are complementary approaches with different advantages. We obtained the average outcomes that would have been produced if all patients had assigned to treatments by each of the machine learning methods, respectively. Then, the average outcomes were compared using Z-tests to the average outcomes that would have been achieved if all patients had received a single treatment. However, we were not able to identify a decision rule among these methods that yielded significantly greater improvement for patients than simply assigning all patients to one overall best treatment. Hence, we developed and applied a new tree-based approach, which judiciously splits the data set into multiple disjoint subgroups sequentially in order to maximize a leave-one-out cross-validated estimate of the average outcome resulting from each split. This new method is distinct from the aforementioned

Table 1.1: Patient characteristics included in analyses

	Characteristic		N (%) or Mean (SD)
Demographic characteristics	Age at baseline		65.1(10.8)
	Education status	Grade school/ junior high	2 (0.7%)
		Some high school	9 (3.0%)
		High school or equivalent	28 (9.2%)
		Trade/technical/vocational school	19 (6.3%)
		Some college credit	31 (10.2%)
		Associate's degree	31 (10.2%)
		Bachelor's degree	61 (20.1%)
		Post graduate work or graduate degree	122 (40.3%)
Clinical and OA-related characteristics	Body mass index (kg/m^2)		31.2 (7.9)
	Baseline WOMAC pain score (Range: 0-20)		5.9 (3.7)
	Baseline WOMAC stiffness score (Range: 0-8)		3.4 (2.1)
	Baseline WOMAC function score (Range: 0-68)		21.9 (12.8)
	Baseline WOMAC total score (Range: 0-96)		31.2 (17.6)
Other Symptoms and Psychosocial Variables	Brief fear of movement score (Range: 6-24)		14.1 (3.4)
	Self-efficacy for exercise scale (Range: 0-90)		56.4 (20.3)
	Family support for exercise (Range: 10-50)		28.4 (11.4)
	Satisfaction with physical function (Range: -3-3)		0.1 (1.7)
	PROMIS fatigue score (Range: 33.1-77.8)		51.2 (9.0)
	Depressive symptoms (PHQ-8) (Range: 0-24)		3.8 (4.2)

machine learning methods in that it determines each mechanistic subgroup by the average outcome (value function), incorporating the advantages of RF. The algorithm is detailed in the Technical Details for Chapter 1.

1.2.3 PATH-IN Data

PATH-IN participants completed outcome assessments at baseline, 4-month follow-up and 12-month follow-up; these analyses focus on 12-month follow-up. The primary outcome was the WOMAC, which is a well-established self-reported measure of pain (5 items), stiffness (2 items) and function (17 items) (Bellamy, 2002). All items are measured on a 5-point Likert scale, with a total scale range of 0-96; higher scores indicate worse symptoms and function. For this analysis, we included participants in all 3 study arms. There were 47 covariates measured at baseline, including demographic, clinical, OA-related and physical activity related variables (Table 1.1). Out of 350 participants, 47 had missing values in their WOMAC total score at 12-month follow-up. We removed these participants, resulting in 303 participants for our analysis. Since the proportion of missingness was uniformly less than 15% for missing values in baseline covariates, we imputed these values using MissForest, implemented in the missForest R package (Stekhoven and Bühlmann, 2012). The advantage of MissForest is that it can simultaneously handle categorical variables and continuous variables of unequal scales, which aligns with the PATH-IN data, and it is suited for data sets with high dimensional and potentially non-linear interactions.

1.2.4 Application of the Value Function

In this study, we used the value function (VF), V, as a measure of treatment effectiveness based on assignments generated from the RF informed Tree-based Learning algorithm, and we denote $\hat{V}$ as an estimate of V. The VF is the expectation of an outcome if future patients followed the estimated decision rule derived from the data; see Supplement for details. Higher value functions indicate greater quality of the decision rule and greater effectiveness of the regime. We chose a jackknife estimator for estimating the value function, as recommended in Jiang et al. (2020). This approach is equivalent to leave-one-out cross-validation, and it is approximately unbiased (i.e., consistent). Using this estimator, we can compare how well each machine learning method performs and determine statistical significance for the differences between treatment regimes. As in Jiang et al. (2020), the value function of a Zero-Order Model (ZOM), i.e., the regime where the estimated best single treatment is given to everyone, was used as a tool for comparing how well estimated regimes performed. That is, the VFs for three ZOMs (IBET, PT, and WT) were estimated, and the IBET ZOM, which produced the largest VF estimate out of three VF estimates was compared with a candidate treatment regime. Z-tests were applied to evaluate whether the method returned statistically significantly better treatment

Method used for estimating a treatment regime	**Value function**
Random Forest Informed Tree-Based Learning	75.5
Random Forest	71.3
LIST embedded with Random Forest	69.9
LIST embedded with Kernel Ridge Regression	71.1
PT ZOM	69.8
IBET ZOM	71.1
WT ZOM	67.1

Table 1.2: Value function estimates for outcome at 12-month visit

regimes than simply assigning the estimated single best treatment to all patients. The value function estimates for each method are displayed in Table 1.2.

1.2.5 Random Forest Informed Tree-based Learning

For computational feasibility, we chose the 13 candidates of the patient characteristics for the analysis, based on Variable Importance from RF prior to running the algorithm, since these 13 covariates were selected at least once by the cross-validation to be the most important covariates. In Section 2 of Supplement, the detailed strategy for the selected 13 covariates is described. The algorithm then begins dividing the data set into two subgroups, followed by iteratively splitting these subgroups into finer subgroups. In each iteration, the VF is calculated, and the algorithm determines whether the partition at that iteration is beneficial. The iterations continue until the splitting does not statistically significantly improve the VF. More detailed explanations for the variable selection and the algorithm are included in the Supplement.

1.3 Results

According to Table 1.2, RF informed Tree-based learning and RF were the two methods that yielded higher VF than the VF of ZOMs. The p-value for RF informed Tree-based learning was 0.0125 and the p-value for RF was 0.9. The significantly greater VF for RF informed Tree-based learning indicates that subgroups of patients would achieve greater improvement from the assigned treatments estimated by the new method than from receiving IBET (the best

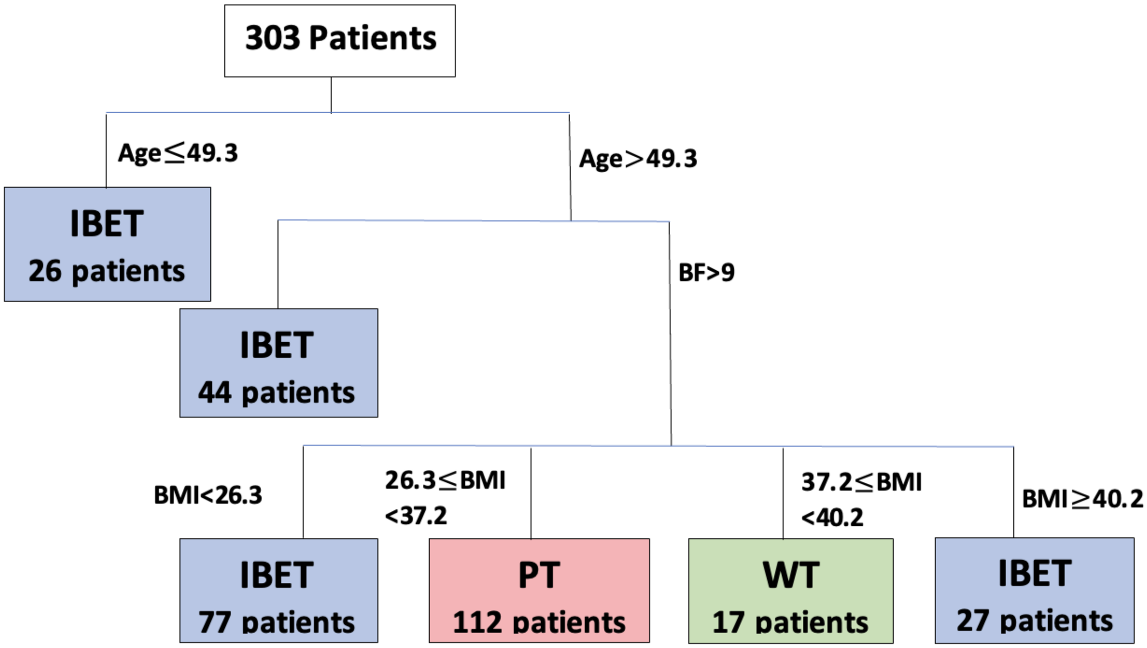

Figure 1.1: Final Rule for the data set with the outcome at month 12

overall treatment based on 12-month WOMAC change) uniformly. Figure 1.1 displays the final rule determined for the total WOMAC outcome at 12-month follow-up. Although the rule has five split points, the fourth and fifth split points have been combined for improve interpretability since thresholds for both nodes are defined using BMI. Notable features of this decision rule include: 1) IBET was the optimal treatment for more than half of patients overall (n=174); 2) For a subgroup of younger individuals (age ≤ 49.3 years), IBET was the optimal treatment; 3) The subgroup for whom PT was the optimal treatment was characterized by age > 49.3 years, high BF (> 9), and BMI between 26.3 and 37.2 kg/m^2. 4) For 17 patients, WT was the optimal treatment.

1.4 Discussion

In this study, we applied a new machine learning algorithm, Random Forest informed Tree-based Learning, to discover optimal treatment regimes for subgroups of patients in a trial of two exercise-based interventions for knee OA. The method addresses limitations of established machine learning methods (RF and LIST-based methods with kernel ridge regression and RF), which did not produce regimes that were significantly better than the ZOM in this study. The new algorithm successfully identified distinct subgroups for whom PT, IBET, or WT was the best treatment at the 12-month visit. Specifically, assignment of the optimal treatment regime resulted in a significant improvement over the ZOM; this is strong evidence that the proposed treatment regimes would deliver more beneficial results to patients than assigning a single best treatment to all individuals. Hence, tailoring referrals to specific exercise-based interventions, based on patient characteristics, could result in greater impacts on OA symptoms. These findings are particularly interesting in the context of the overall findings of the trial, which showed that mean improvements in WOMAC were similar across the 3 study arms, including the wait list, at 12-months. This further suggests that exercise-based interventions may be most effective when they are selected based on patient characteristics.

Subgroups identified by the algorithm were characterized by differences in age, BMI and fear of movement, which are all feasible to evaluate in clinical settings. IBET was the optimal treatment for 57$\%$ of patients in these analyses. This is of interest, as it suggests that this lower resource intervention (relative to PT) may be more favorable for about half of patients with OA, when considering 12-month outcomes. Participants younger than 49.3 years old and those at least 49.3 years old with low fear of movement were subgroups for whom IBET was the optimal treatment regime; clinically, this suggests that patients with these characteristics may be better able to sustain behaviors and impacts of a self-guided exercise program. There was one relatively large subgroup (n = 112) for whom PT was the optimal treatment; this group was characterized by age > 49.3 years, high fear of movement and BMI ranging between 26.3 and 37.2 kg/m^2. The next largest group (n = 77) assigns IBET to patients with age $\leq\ 49.3$, high fear of movement, and BMI less than 26.4. However, for other subgroups, results are more challenging to interpret clinically due to their involvement of combinations of variables and their identified thresholds. For a fairly small number of participants (n=17), the wait list condition was the optimal treatment regime. This indicates that for the majority of individuals in the trial, one of the two active treatments (IBET or PT) was superior to no treatment.

Although this algorithm addresses some shortcomings of other machine learning methods, it also has some limitations. Since the algorithm exhaustively searches for one split point out of all the distinct points from every important variable in the list until the third variable is chosen, it can be computationally burdensome. Moreover, it is not guaranteed that the VF estimate from the final rule is the maximum of all possible VF estimates. The reason is that once a subgroup in a particular iteration has been decided, the algorithm in the next iteration searches for the subsequent finer subgroup only in the subgroup identified in the previous iteration. Although this process does not necessarily lead to the maximum VF estimate, it is designed to obtain a decision rule that produces a VF estimate as statistically significant as possible while also providing mechanistic parsimonious rules. For future studies, we suggest developing a tool for discovering the maximum VF estimate with its corresponding decision rule with factors that identify distinct subgroups.

In summary, these secondary analyses from the PATH-IN trial successfully identified meaningful subgroups of patients for whom PT, IBET and WT was the optimal treatment. Because these results are exploratory, further analyses are needed to evaluate whether these patterns are also observed in other cohorts and contexts. However, we believe these results offer some practical guidance for patients with knee OA, as well as clinicians who refer these patients to exercise-based interventions. First, results suggest that younger patients ($leq49$ years) and those who are older but have low fear of movement may be able to sustain benefits (over a 12-month period) from a supported home-based exercise intervention. Second, patients > 49 years of age who have greater fear of movement may be good candidates for a referral to PT and may particularly benefit from this higher level of support and guidance, with respect to sustaining improvements after 1 year.

CHAPTER 2: DOUBLY ROBUST ESTIMATION AND INFERENCE OF UTILITY FUNCTIONS

2.1 Introduction

In this chapter, we study a method for estimating the utility that is customized to each patient. By introducing the utility to combine multiple outcomes and estimate ITRs, it is possible to use the numerous estimators previously developed for ITRs for a single scalar outcome. However, since the outcome model now incorporates the utility whose true model is complicated, the concern of misspecification of the outcome model still remains. Hence, an alternative estimator that is robust to misspecification of the outcome model is required.

Therefore, we propose doubly robust estimators for the utility of two outcomes and the probability of assigning the correct treatments in observational data. In addition, we suggest the estimator for ITRs that corresponds to the optimal utility that would yield the best improvements in the outcomes of patients. According to the literature review, we employ the inverse reinforcement learning framework suggested in Luckett et al. (2021). During the estimation process, we use EARL (Zhao et al., 2019) for estimating ITRs, transferring the doubly robust property to the suggested estimators.

We introduce our method in Section 2.2. In Section 2.3, we present the theoretical properties of the utility estimator and the boundary function of the estimated ITRs. In section 2.4, we study the performance of our method by simulations. In section 2.5, we present an illustrative application based on data for a bipolar disorder study. Lastly in section 2.6, we summarize our method with possible future topics.

2.2 Methods

2.2.1 Setting

Let $U_i = (\boldsymbol{X}_i, A_i, Y_i, Z_i),\ i = 1, \cdots, n$ be independent and identically distributed realizations of $U = (\boldsymbol{X}, A, Y, Z)$. We assume that we have two available treatments, $A \in \mathcal{A} = \{1, -1\}$, and we have covariates $\boldsymbol{X} \in \mathcal{X} \subset \mathbb{R}^p$. We have two scalar outcomes, Y, and Z, where higher values are most desirable. In order to express the two outcomes Y and Z into one scalar to formally quantify the two outcomes simultaneously for coherent treatment regime estimation, we introduce a utility function $U_\theta = u(Y, Z; \theta)$ where $u : \mathbb{R}^2 \to \mathbb{R}$, and θ is in the parameter space Θ. The utility functions are dependent on covariates $\boldsymbol{X}$, and in this paper, we define the utility as a convex combination of Y and Z, $U_\theta = u(Y, Z; \boldsymbol{X}, \theta) = \omega_\theta(X)Y + \{1 - \omega_\theta(\boldsymbol{X})\}Z$, where $\omega_\theta(\boldsymbol{X}) : \mathcal{X} \to \mathbb{R}$ is a smooth function. For a treatment assignment function, let $d : \mathcal{X} \mapsto \{1, -1\}$ be the decision which allocates $d(x)$ to patients who have $\boldsymbol{X} = \boldsymbol{x}$ as covariates, and when d is assumed to be in a known class of decision $\mathcal{D}$. We assume there is a measurable function $f : \mathcal{X} \mapsto \mathbb{R}$ for which $d(\boldsymbol{X}) = \text{sgn}(f(\boldsymbol{X}))$.

In this paper, we take the potential outcome framework (Rubin, 1974; Splawa-Neyman et al., 1990). Let's denote $Y^*(a)$ the potential outcome of Y which would have been produced if the treatment $a \in \mathcal{A}$ had been assigned. In the same context, $Y^*(d) = \sum_{a \in \mathcal{A}} Y^*(a) I\{d(X) = a\}$ indicates the outcome that would have been produced under a treatment regime d, and the utility of counterfactual outcomes is defined as $U_\theta^*(d) = u\{Y^*(d), Z^*(d); \theta\}$. The expectation of the potential outcome under a regime d, which is called the value, is defined as $V_Y(d) = E[Y^*(d)]$, and similarly, $V_\theta(d) = E[U_\theta^*(d)] = E\big[u\{Y^*(d), Z^*(d); \theta\}\big]$. The values are used to evaluate performance of a treatment regime d, since it is the expectation of the outcome that the population would have produced if the treatment regime d had been assigned. The treatment regime that results in the largest value with respect to Y is denoted as an optimal treatment regime d_Y^*, i.e. $V_Y(d_Y^*) = \max_{d \in \mathcal{D}} V_Y(d)$. The optimal treatment regime for U_θ is d_θ^* where $V_\theta(d_\theta^*) = E\big[u\{Y^*(d_\theta^*), Z^*(d_\theta^*)\}\big] \geq E\big[u\{Y^*(d), Z^*(d)\}\big]$ for all $d \in \mathcal{D}$. However, in order to identify values and optimal treatment regimes using observed data, we need to make the following assumptions.

Assumption 2.1. *We assume the following causal assumptions:*

1. *Consistency,* $Y = Y^*(A)$ *and* $Z = Z^*(A)$.

2. *Positivity,* $\forall(\boldsymbol{x}, a) \in \mathcal{X} \times \mathcal{A},\ \Pr(A = a|\boldsymbol{X} = \boldsymbol{x}) \geq c > 0$.

3. *No unmeasured confounders,* $\{Y^*(-1), Y^*(1)\} \perp A|\boldsymbol{X}$ *and* $\{Z^*(-1), Z^*(1)\} \perp A|\boldsymbol{X}$.

We also assume the stable unit treatment value assumption: there is no interference between subjects, and there is only one treatment set (Rubin, 1980).
Additionally, we make an assumption that experts are making optimal decisions with nonzero probability. We model the probability that the experts based observed decisions are actually matching the true optimal decision, we reversely make inference on the optimal utility function for each patient, adopting an inverse reinforcement learning approach. We will introduce the detail in the next section.

2.2.2 The pseudo-likelihood estimation

For the doubly robust estimator of the utility, we employ the pseudo likelihood suggested in Luckett et al. (2021), which we recall,

$$\hat{\mathcal{L}}_n(\theta, \beta) \propto \prod_{i=1}^{n} \frac{\exp\left[\boldsymbol{X}_i^T \beta 1\{A_i = \hat{d}_{n,\theta}(\boldsymbol{X}_i)\}\right]}{1 + \exp(\boldsymbol{X}_i^T \beta)}, \tag{2.1}$$

where $\hat{d}_{n,\theta}$ is an estimator for the optimal treatment regime d^*_θ, and θ is unknown. In addition, we assume the following Assumption 2.2. Then, according to Theorem 5 in Luckett et al. (2021), identifiability of the model holds.

Assumption 2.2 (Identifiability)**.** *The following conditions hold.*

1. $\beta \in \mathcal{B} \subset \mathbb{R}^p$ *and* $\theta \in \Theta \subset \mathbb{R}^q$*, where* $\mathcal{B}$ *and* Θ *are compact.*

2. *$\beta_0 \neq 0$.*

3. *$\mathcal{X}$ is bounded ($\boldsymbol{X} \in \mathcal{X} \subset \mathbb{R}^p$ $a.s.$).*

4. *Let $\mathcal{X}_S$ be the collection of subsets of $\mathcal{X}$ consisting of sets of the form $\{\boldsymbol{x} \in \mathcal{X} : d_\theta(\boldsymbol{x}) \neq d_{\theta_0}(\boldsymbol{x})\}$ for $\theta \in \Theta \setminus \{\theta_0\}$, together with the complements of these sets. Then:*

   (a) For all $X_S \in \mathcal{X}_S$, $0 < \Pr(\boldsymbol{X} \in X_S) < 1$, and

   (b) $E\left(\boldsymbol{X}\boldsymbol{X}^T | \boldsymbol{X} \in X_S\right)$ is full rank $\forall X_S \in \mathcal{X}_S$.

In the next sections, we describe the doubly robust approach for estimating $\hat{d}_{n,\theta}$ and the detailed algorithm for $(\hat{\theta}_n, \hat{\beta}_n)$ to optimize the true parameters (θ_0, β_0).

2.2.3 Estimation of individualized treatment regimes (ITR)

Prior to applying logistic regression, an estimated individualized treatment regime (ITR) is required. In Luckett et al. (2021), this is achieved by Q-learning. Let $Q_Y(\boldsymbol{X}, a) = \mathbb{E}[Y|\boldsymbol{X}, A = a]$ and $Q_Z(\boldsymbol{X}, a) = \mathbb{E}[Z|\boldsymbol{X}, A = a]$. Then, due to Assumption 2.1, the optimal treatment regime for outcome Y is $d_Y^*(\boldsymbol{X}) = \max_{a \in \mathcal{A}} Q_Y(\boldsymbol{X}, a)$, and for outcome Z is $d_Z^*(\boldsymbol{X}) = \max_{a \in \mathcal{A}} Q_Z(\boldsymbol{X}, a)$ (Qian and Murphy, 2011). For the composite utility, let $Q_\theta(\boldsymbol{X}, a) = E[U_\theta|\boldsymbol{X}, A = a] = \omega_\theta(\boldsymbol{X})Q_Y(\boldsymbol{X}, a) + (1 - \omega_\theta(\boldsymbol{X}))Q_Z(\boldsymbol{X}, a)$, and then the optimal treatment regime is $d_\theta^*(\boldsymbol{X}) = \max_{a \in \mathcal{A}} Q_\theta(\boldsymbol{X}, a)$. Then, for fixed θ, an estimated ITR is $\hat{d}_{n,\theta}$ by Q-learning. However, the performance of Q-learning depends heavily on the correct specification of Q-functions, especially when the true relationships of outcomes with covariates and treatments are complicated. For these reasons, we utilize a direct approach for estimating an optimal treatment regime for two outcomes.

In this doubly robust approach, the objective is to estimate d_θ^* by searching $d \in \mathcal{D}$ that maximizes $V_\theta(d)$. Specifically, we use AIPWE for estimating d_θ^* in a manner analogous to the single outcome scenario of EARL described in the literature review. Let

$$V_\theta^{\text{AIPWE}}(d) = \mathbb{E}\big[\frac{U_\theta I\{A = d(\boldsymbol{X})\}}{\pi(d(\boldsymbol{X}); \boldsymbol{X})} - \frac{I\{A = d(\boldsymbol{X})\} - \pi(d(\boldsymbol{X}); \boldsymbol{X})}{\pi(d(\boldsymbol{X}); \boldsymbol{X})} Q\{\boldsymbol{X}, d(\boldsymbol{X})\}\big]. \quad (2.2)$$

The estimator of $V_\theta^{\text{AIPWE}}(d)$ is $\hat{V}_\theta^{\text{AIPWE}}(d) = \mathbb{E}_n\big[\frac{U_\theta I\{A=d(\boldsymbol{X})\}}{\hat{\pi}(d(\boldsymbol{X});\boldsymbol{X})} - \frac{I\{A=d(\boldsymbol{X})\}-\hat{\pi}(d(\boldsymbol{X});\boldsymbol{X})}{\hat{\pi}(d(\boldsymbol{X});\boldsymbol{X})}\hat{Q}\{\boldsymbol{X}, d(\boldsymbol{X})\}\big]$, where $\hat{Q}_\theta(\boldsymbol{X}, a)$ are estimators of $Q_\theta(\boldsymbol{X}, a)$. Next, for the doubly robust property, we assume consistent estimators for the propensity score model $\pi(a; \boldsymbol{X})$ and the outcome model $Q_\theta(\boldsymbol{X}, a)$ in the following assumption allowing that the limiting quantities may not be correct. The theorem verifying doubly robustness is presented in Section 2.3.

Assumption 2.3. *For each $\theta \in \Theta$, there exist $\pi^m(a; \boldsymbol{X})$ and $Q_\theta^m(\boldsymbol{X}, a)$ such that $\hat{\pi}(a; \boldsymbol{X}) \xrightarrow{P} \pi^m(\boldsymbol{X}; a)$ and $\hat{Q}_\theta(\boldsymbol{X}, a) \xrightarrow{P} Q_\theta^m(\boldsymbol{X}, a)$.*

For each $\theta \in \Theta$, we optimize $f_\theta^*(\boldsymbol{X}) \in \mathcal{F}$ such that $d_\theta^*(\boldsymbol{X}) = \text{sgn}(f_\theta^*(\boldsymbol{X}))$ for a functional space $\mathcal{F}$, and hence $\mathcal{D} = \{\text{sgn}(f(\boldsymbol{X})) : f \in \mathcal{F}\}$. Let

$$\begin{aligned} W_{a,\theta}(U) &= \frac{U_\theta I(A = a)}{\pi(a; \boldsymbol{X})} - \frac{I(A = a) - \pi(a; \boldsymbol{X})}{\pi(a; \boldsymbol{X})} Q_\theta(\boldsymbol{X}, a), \text{ and} \\ \hat{W}_{a,\theta}(U) &= \frac{U_\theta I(A = a)}{\hat{\pi}(a; \boldsymbol{X})} - \frac{I(A = a) - \hat{\pi}(a; \boldsymbol{X})}{\hat{\pi}(a; \boldsymbol{X})} \hat{Q}_\theta(\boldsymbol{X}, a), \end{aligned}$$

Analogous to EARL in a single outcome setting, for each $\theta \in \Theta$,

$$\tilde{f}_{n,\phi,\theta}^{\lambda_n} = \arg\inf_{f \in \mathcal{F}} \mathbb{E}_n\Big[|\hat{W}_{1,\theta}|\phi\big\{\text{sgn}(\hat{W}_{1,\theta}) f(\boldsymbol{X})\big\} + |\hat{W}_{-1,\theta}|\phi\big\{-\text{sgn}(\hat{W}_{-1,\theta}) f(\boldsymbol{X})\big\} + \lambda_n \|f\|^2\Big], \quad (2.3)$$

where $n\lambda_n \to 0$ as $n \to \infty$. We also apply the sample splitting technique used to estimate EARL. Additionally, assume $T_{1,n} \equiv \sqrt{n}\max_{1\leq j\leq J}\Big|\frac{(J-1)n}{J(n-n_j)} - 1\Big| \to 0$, and $T_{2,n} \equiv \sqrt{n}\max_{1\leq j\leq J}\Big|\frac{n}{Jn_j} -$

$1\Big| \to 0$ as $n \to \infty$. We suggest the estimator for the boundary function $f^*_\theta(\boldsymbol{X})$ as

$$\hat{f}^{\lambda_n}_{n,\phi,\theta} = \frac{1}{J}\sum_{j=1}^{J} \hat{f}^{\lambda_n,(j)}_{n,\theta}, \text{ and} \tag{2.4}$$

$$\begin{aligned}\hat{f}^{\lambda_n,(j)}_{n,\phi,\theta} = \arg\inf_{f\in\mathcal{F}} \mathbb{E}^{(-j)}_n \Big[|\hat{W}_{1j,\theta}|\phi\big\{\text{sgn}(\hat{W}_{1j,\theta})f(\boldsymbol{X})\big\} + |\hat{W}_{-1j,\theta}|\phi\big\{-\text{sgn}(\hat{W}_{-1j,\theta})f(\boldsymbol{X})\big\} \\ + \lambda_{nj}\|f\|^2\Big],\end{aligned} \tag{2.5}$$

for each $\theta \in \Theta$ where $\hat{W}_{aj,\theta} = \frac{U_\theta I(A=a)}{\hat{\pi}_j(a;\boldsymbol{X})} - \frac{I(A=a)-\hat{\pi}_j(a;\boldsymbol{X})}{\hat{\pi}_j(\boldsymbol{X};a)}\hat{Q}_{\theta,j}(\boldsymbol{X},a)$ for $a \in \{1,-1\}$, $n\lambda_{nj} \to 0$.

2.2.4 Overview of the algorithm

An estimated ITR $\hat{d}_{n,\theta} = \text{sgn}(\hat{f}^{\lambda_n}_{n,\phi,\theta})$ is plugged in to the pseudo-likelihood (2.1) to estimate the parameters for the utility function and the probabilities of correct treatment recommendation. However, as mentioned in Luckett et al. (2021), an optimizing method that uses gradients is not applicable since (2.1) is not smooth in θ. Therefore, we employ a profile pseudo-likelihood $\widetilde{L}_n(\theta;\hat{d}_{n,\theta}) = \max_{\beta\in\mathbb{R}^p} \hat{L}_n\big\{\theta,\beta;\hat{d}_{n,\theta}\big\}$ as in Luckett et al. (2021). In order to optimize a multi-dimensional parameter θ, we take advantage of the Metropolis algorithm. We generate a chain from a random walk, $(\theta^1,\cdots,\theta^B)$, and obtain $\hat{\theta}_n$ that yields the largest profile pseudo-likelihood $\tilde{L}_n(\theta;\hat{d}_{n,\theta})$. The high-level description of the algorithm is below.

Another advantage of the suggested method is that asymptotic consistency is guaranteed when at least one of the Q-function or the propensity score function is correctly specified by utilizing EARL, which achieves a doubly robust property. By plugging in the EARL estimator to the pseudo-likelihood, we can enjoy the flexible characteristic of EARL, which provides us protection if the true Q-function or propensity score function is hard to be formulated. In addition to the doubly robustness of the estimated ITR, we expect the robustness of estimators for utilities and the probability of the correct assignment of optimal treatments.

Algorithm 1: Pseudo-likelihood estimation of utility function

```
Set a chain length, $B$, fix $\Sigma \succcurlyeq 0$, and initialize $\theta^1$ to a starting value in $\mathbb{R}^p$;
for $b = 2, \ldots, B$ do
    Generate $\mathbf{e} \sim N(0, \Sigma)$;
    Set $\widetilde{\theta}^{b+1} = \theta^b + \mathbf{e}$;
    Obtain $\hat{f}_{n,\widetilde{\theta}^{b+1}} = \hat{f}^{\lambda_n}_{n,\widetilde{\theta}^{b+1}}$ by (2.4) and (2.5);
    Estimate $\hat{d}_{n,\widetilde{\theta}^{b+1}} = \text{sgn}(\hat{f}_{n,\widetilde{\theta}^{b+1}})$;
    Compute $p = \min\left\{\widetilde{L}_n\left(\widetilde{\theta}^{b+1}; \hat{d}_{n,\widetilde{\theta}^{b+1}}\right) / \widetilde{L}_n\left(\widetilde{\theta}^{b}; \hat{d}_{n,\widetilde{\theta}^{b}}\right), 1\right\}$;
    Generate $U \sim U(0,1)$; if $U \leq p$, set $\theta^{b+1} = \widetilde{\theta}^{b+1}$; otherwise, set $\theta^{b+1} = \theta^b$;
end
Select $\hat{\theta}_n = \arg\max_{\theta \in \{\theta^1, \cdots, \theta^B\}} \widetilde{L}_n\left(\theta; \hat{d}_{n,\theta}\right)$;
Estimate $\hat{\beta}_n = \arg\max_{\beta \in \mathbb{R}^P} \hat{L}_n\left(\hat{\theta}_n, \beta; \hat{d}_{n,\hat{\theta}_n}\right)$;
```

2.3 Theoretic Results

In this section, we state the theorems regarding the consistency and the asymptotic distribution of the proposed estimators. All proofs in this section are deferred to the Technical Details for Chapter 2.

We assume that $f \in \mathcal{F} \subset \mathcal{M}$, where $\mathcal{M}$ is a space of measurable functions. For each $\theta \in \Theta$, let the risk of a function f be

$$\mathcal{R}_\theta(f) = \mathbb{E}\big[\frac{\{\omega_\theta(\boldsymbol{X})Y + (1 - \omega_\theta(\boldsymbol{X}))Z\} I(A \neq \text{sgn}(f(\boldsymbol{X})))}{\pi(a; \boldsymbol{X})}\big].$$

Note that $f^*_\theta(\boldsymbol{X}) = \arg\inf_{f \in \mathcal{F}} \mathcal{R}_\theta(f)$ such that $d^*_\theta(\boldsymbol{X}) = \text{sgn}(f^*_\theta(\boldsymbol{X}))$ for each $\theta \in \Theta$. Accordingly, we consider the ϕ-risk as

$$\mathcal{R}_{\theta,\phi}(f) = \mathbb{E}\Big[|W_{1,\theta}(U)|\phi\big\{\text{sgn}(W_{1,\theta}(U))f\big\} + |W_{-1,\theta}(U)|\phi\big\{-\text{sgn}(W_{-1,\theta}(U))f\big\}\Big],$$

where $\phi(\cdot)$ is a convex surrogate similar to Zhao et al. (2019). Denote $f^*_{\theta,\phi}(\boldsymbol{X}) = \arg\inf_{f \in \mathcal{F}}$ $\mathcal{R}_{\theta,\phi}(f)$. The following lemma states the Fisher consistency for each $\theta \in \Theta$, where the θ-optimal rule $d^*_\theta(\boldsymbol{X})$ is equivalent to the sign of $f^*_{\theta,\phi}(\boldsymbol{X})$, which is obtained from minimizing ϕ-risk. For

Fisher consistency of the optimal ITR and consistency of estimators in the following theorem, assume $\pi^m(a;\boldsymbol{x}) = \pi(a;\boldsymbol{x})$ or $Q^m_\theta(\boldsymbol{x}, a) = Q_\theta(\boldsymbol{x}, a)$, i.e., either the propensity score model or the Q-function is correct.

Lemma 2.1 (Fisher Consistency). *For each $\theta \in \Theta$, let $d^*_\theta(\boldsymbol{x})$ be the optimal ITR that satisfies $d^*_\theta(\boldsymbol{x}) = \arg\max_{a\in\{-1,1\}} Q_\theta(\boldsymbol{x}, a)$. Then, $d^*_\theta(\boldsymbol{x}) = sgn\{f^*_{\theta,\phi}(\boldsymbol{x})\}$.*

Let θ_0 be the true parameter for the utility function. Also, let β_0 be the true parameter for the probability of correct recommendation of treatments, i.e., $P\big\{A = d^*_{\theta_0}(\boldsymbol{X})|\boldsymbol{X} = \boldsymbol{x}\big\} = \text{expit}(\boldsymbol{x}^T\beta_0)$. Then, Lemma 2.1 leads to the conclusion that the optimal treatment $d^*_{\theta_0}(\boldsymbol{x})$, which is the true optimal decision and is equivalent to the sign of $f^*_{\theta_0,\phi}(\boldsymbol{x})$. Also, Lemma 2.1 enables us to estimate the optimal treatment $d^*_\theta(\boldsymbol{x})$ by $\text{sgn}(\hat{f}^{\lambda_n}_{n,\phi,\theta}(\boldsymbol{X})))$ for each θ, and further estimate $(\hat{\theta}_n, \hat{\beta}_n)$. Therefore, for the remainder of this section, we replace $f^*_{\theta,\phi}$ with f^*_θ. We denote $f^*_\theta \equiv f^*_{\phi,\theta}$ and $\hat{f}_{n,\theta} \equiv \hat{f}^{\lambda_n}_{n,\phi,\theta}$ for simplicity. Next, we state the doubly robust consistency of the estimators, and present the additional needed assumptions in advance:

Assumption 2.4. *The following hold.*

1. *For $\theta_0, \theta \in \Theta$ such that $d^*_{\theta_0}(\boldsymbol{X}) = d^*_\theta(\boldsymbol{X})$, $\theta_0 = \theta$ almost surely.*

2. *β, β_0 are in the interior of a compact set $\mathcal{B}$.*

3. *Assume $\|E\boldsymbol{X}\| < \infty$ where $\|\cdot\|$ is a euclidean norm.*

Assumption 2.5. *We assume the following conditions on functions.*

1. *The collection $\mathcal{F} = \{\gamma^T\xi(x) : \gamma \in \mathbb{R}^q\}$, where $\xi(\cdot) = \{\xi_1(\cdot), \ldots, \xi_q(\cdot)\}^T$ is a q-dimensional vector basis, where $\xi_j : \mathcal{X} \mapsto \mathbb{R}$, for $j = 1, \ldots, q$ and where $P(\xi(X)\xi(X)^T) > 0$.*

2. *For each $\theta \in \Theta$, the set of utility functions $\omega_\theta(\boldsymbol{X}) : \mathcal{X} \to \mathbb{R}$ is contained in a VC class, where $\omega_\theta(\boldsymbol{X})$ has a first order derivative $\dot{\omega}_\theta(\boldsymbol{X}) : \mathcal{X} \to \mathbb{R}$.*

Assumption 2.6. *The convex surrogate $\phi(\cdot)$ is differentiable except for a finite set C, where for any $c \in C$, $\Pr(\gamma^T \xi(\boldsymbol{X}) = c) + \Pr(-\gamma^T \xi(\boldsymbol{X}) = c) = 0$. Denote $\dot{\phi}(\cdot)$ and $\ddot{\phi}(\cdot)$ as the first and second derivative of $\phi(\cdot)$, respectively.*

The choice of the convex surrogates in this paper, which include hinge loss, exponential loss, logistic loss, and squared hinge loss, satisfies Assumption 2.6. Assumption 2.5 puts a restriction to the functional forms for the boundary function and the utility function, and we let $\omega_\theta(\boldsymbol{X}) = \text{expit}(\boldsymbol{X}^T\theta)$ for the utility function in this paper.

Theorem 2.1 (Doubly robust consistency)**.** *Assume $\pi^m(a; X) = \pi(a; X)$, or $Q^m_\theta(X, a) = Q_\theta(X, a)$ uniformly for $\theta \in \Theta$. Then the following results are achieved.*

(a) Let the pseudo likelihood estimators be $(\hat{\theta}_n, \hat{\beta}_n) = \arg\max_{\theta\in\Theta,\beta\in\mathcal{B}} \hat{L}_n(\theta, \beta)$. Then, $\|\hat{\theta}_n - \theta_0\| \xrightarrow{P} 0$ and $\|\hat{\beta}_n - \beta_0\| \xrightarrow{P} 0$.

(b) $\sup_{\theta\in\Theta} \mathbb{E}[\|\hat{f}_{n,\theta}(\boldsymbol{X}) - f^*_\theta(\boldsymbol{X})\|] \xrightarrow{P} 0$ and $\sup_{\theta\in\Theta} \mathbb{E}[\|\hat{d}_{n,\theta}(\boldsymbol{X}) - d^*_\theta(\boldsymbol{X})\|] \xrightarrow{P} 0$, where $\hat{d}_{n,\theta}(\boldsymbol{X}) = sgn\{\hat{f}_{n,\theta}(\boldsymbol{X})\}$.

(c) Denote $\hat{V}_\theta(d) \equiv \hat{V}^{\text{AIPWE}}_\theta(d)$. Then, $\left|\hat{V}_{\hat{\theta}_n}\left(\hat{d}_{n,\hat{\theta}_n}\right) - V_{\theta_0}\left(d^*_{\theta_0}\right)\right| \xrightarrow{P} 0$.

For the asymptotic distribution of $(\hat{\theta}_n, \hat{\beta}_n)$, we first need to show the asymptotic distribution of $\hat{f}_{n,\hat{\theta}_n}$ which is the estimator for the boundary function $f^*_{\theta_0,\phi}$, since the behavior of the utilities is considerably affected by the boundary. However, obtaining $f^*_{\theta_0,\phi}$ consists of minimizing $\mathcal{R}_{\theta_0,\phi}$, where θ_0 is the argmax of the likelihood $\mathcal{L}$. In order to disentangle this two-staged maximization problem, we firstly present the following new argmax and rate of convergence theorems where a maximizing function converges weakly to another maximizing function uniformly in the indexing parameters.

Theorem 2.2 (Indexed Argmax Theorem)**.** *Let (T, d_1) and $(\mathcal{H}, d_2)$ be metric spaces, with T compact and $\mathcal{H}$ complete. Let $(t, h) \mapsto M_{n,t}(h)$ and $(t, h) \mapsto M_t(h)$ be stochastic processes in $l^\infty(T \times \mathcal{H})$. For any $A \subset \mathcal{H}$, let $\tilde{A}$ be the space of maps from T to A. Then, $\tilde{\mathcal{H}}$ is a complete metric space with metric $d(h_1, h_2) = \sup_{t\in T} d_2(h_{1,t}, h_{2,t})$. Assume that in each*

$t \in T$, $h \mapsto M_{n,t}(h)$ has a unique maximum at $\hat{h}_{n,t} \in \mathcal{H}$ and $h \mapsto M_t(h)$ has a unique maximum at $\hat{h}_t \in \mathcal{H}$, where $\hat{h}_n \in \tilde{\mathcal{H}}$ and $\hat{h} \in \tilde{\mathcal{H}}$ almost surely. Assume that $\forall \epsilon > 0$, there exists compact $K \subset \mathcal{H}$ such that $\liminf_{n\to\infty} P(\hat{h}_n \in \tilde{K}) \geq 1 - \epsilon$ and $P(\hat{h} \in \tilde{K}) \geq 1 - \epsilon$. Also, assume that for every compact set $K \subset \mathcal{H}$, $M_n \rightsquigarrow M$ in $l^\infty(T \times K)$ and that $\forall t_1 \in T$, $\lim_{\delta\downarrow 0} \sup_{t\in T, d(t,t_1)<\delta} \sup_{h\in K} |M_{t_1}(h) - M_t(h)| = 0$. Then, $\hat{h}_n \rightsquigarrow \hat{h}$ in $\tilde{\mathcal{H}}$, $t \mapsto \hat{h}_t$ is uniformly equicontinuous over T, and $\hat{h}$ is separable.

Theorem 2.3 (Indexed Rate of Convergence)**.** *Let $M_{n,t}$ be a sequence of stochastic processes indexed by a semimetric space $(\mathcal{H}, d)$, and $M_t : \mathcal{H} \to \mathbb{R}$ a deterministic function such that for every $h \in N_t$ where $N_t = \{h \in \mathcal{H} : d(h, h_t^*) \leq \delta\}$ for some $\delta > 0$, there exists a $c_1 > 0$ such that*

$$\sup_{t\in T} \left[M_t(h) - M_t(h_t^*)\right] \leq -\sup_{t\in T} c_1 d^2(h, h_t^*) \tag{2.6}$$

Suppose that for all n large enough and sufficiently small δ, the centered process $M_{n,t} - M_t$ satisfies

$$\mathbb{E}^* \sup_{t\in T, d(h,h_t^*)<\delta} \sqrt{n}\left|M_{n,t}(h) - M_t(h) - M_{n,t}(h_t^*) + M_t(h_t^*)\right| \leq c_2\phi_n(\delta), \tag{2.7}$$

for $c_2 < \infty$ and functions ϕ_n such that $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ not depending on n, where $\mathbb{E}^*$ is an outer expectation. Let

$$r_n^2\phi_n(r_n^{-1}) \leq c_3\sqrt{n}, for\ every\ n\ and\ some\ c_3 < \infty. \tag{2.8}$$

If the sequence $\hat{h}_{n,t}$ satisfies $\inf_{t\in T}\left(M_{n,t}(\hat{h}_{n,t}) - \sup_{h\in N_t} M_{n,t}(h)\right) \geq -O_P(r_n^{-2})$, then $r_n \sup_{t\in T} d(\hat{h}_{n,t}, h_t^*) = O_P(1)$.

We make use of Theorem 2.2 for investigating the asymptotic behavior of $\hat{f}_{n,\theta}$ near f_θ^*, $\forall\theta \in \Theta$. Let the rate of convergence for $\hat{f}_{n,\theta}$ be a nondecreasing, positive sequence r_n. Provided that $\hat{f}_{n,\theta}$ is the argmax of $M_{n,\theta}(f)$, $\hat{h}_{n,\theta} = r_n(\hat{f}_{n,\theta} - f_\theta^*)$ is the argmax of $h \mapsto \tilde{M}_{n,\theta}(h) \equiv$

$r_n\big[M_n(f_\theta^* + h/r_n) - M_n(f_\theta^*)\big]$ as in Chapter 14 of Kosorok (2008). Then, if $\tilde{M}_{n,\theta} \rightsquigarrow M_\theta$, $r_n(\hat{f}_{n,\theta} - f_\theta^*)$ converges weakly to the argmax of M_θ for $\forall\theta \in \Theta$. In Theorem 2.3, we introduced the theorem for determining rate of convergence r_n.

Before presenting the limiting distribution of $\hat{f}_{n,\hat{\theta}_n}$, we recall that the consistency of $\hat{f}_{n,\hat{\theta}}$, which is the condition for weak convergence, is satisfied by previously stated assumptions. Also, we restrict our index set to Θ_ϵ where $\Theta_\epsilon = \{\theta : ||\theta - \theta_0|| < \epsilon, \theta \in \Theta\}$. In addition, we need stronger (but reasonable) assumptions for the weak convergence of $r_n(\hat{f}_{n,\hat{\theta}_n} - f_{\theta_0}^*)$. Therefore, we present the further regularity conditions.

Assumption 2.7. *There exist constants $K_1, K_2 > 0$ such that $|Y| < K_1$ and $|Z| < K_2$.*

Assumption 2.8. *For the estimator $\hat{Q}(\boldsymbol{X}, a)$ of the outcome model $Q(\boldsymbol{X}, a)$, and $\hat{\pi}(a; \boldsymbol{X})$ of the propensity score function $\pi(a; \boldsymbol{X})$, we assume the following.*

1. *Assume $\hat{Q}(X, a)$ and $\hat{\pi}(X; a)$ are determined by a finite number of unknown parameters.*

2. *Assume that $L_\Pi < \pi(a; \boldsymbol{X}) < U_\Pi$ for some $0 < L_\Pi < U_\Pi < 1$.*

3. *$E\|\hat{\pi}(a; x) - \pi(a; x)\|_{P,2}^2 = O(n^{-1})$ and $E\sup_{\theta\in\Theta}\|\hat{Q}_\theta(x, a) - Q_\theta(x, a)\|_{P,2}^2 = O(n^{-1})$, where $\|g\|_{P,r} \equiv [\int_{\mathcal{X}} |g(x)|^r dP(x)]^{1/r}$.*

Assumption 2.9. *Define $V = (V_1, V_2, V_3)$, where $V_1, V_2, V_3 \in \mathbb{R}^{d_j},\ j = 1, 2, 3$ are the unique parameters for $\pi(a; \boldsymbol{X}), Q_Y(\boldsymbol{X}, a), Q_Z(\boldsymbol{X}, a)$, respectively, and let $\hat{V}_n$ be an estimator of V and V_0 be the true value. Assume the following conditions.*

1. *For U_i, independent and identical distributed random variable of $U = (\boldsymbol{X}, Y, Z, A)$, assume that there exists an influence vector $\psi_{i,V} \equiv \psi_V(U_i)$ such that*

$$\sqrt{n}(\hat{V}_n - V_0) = n^{-1/2}\sum_{i=1}^{n}\psi_{i,V} + o_P(1),$$

where $E(\|\psi_V(U)\|^2) < \infty$, and $\psi_V(u) \in \mathbb{R}^p$ for some $p \leq d_1 + d_2 + d_3$.

2. *There exist vectors* $D_\pi(a, \boldsymbol{X}), D_Y(a, \boldsymbol{X}), D_Z(a, \boldsymbol{X}) \in \mathbb{R}^{d_j},\ j = 1, 2, 3$*, such that*

$$\sup_{a,\boldsymbol{x}} \left\|\sqrt{n}\{\hat{\pi}_n(a;\boldsymbol{x}) - \pi(a;\boldsymbol{x})\} - \sqrt{n}(\hat{V}_n - V_0)^T D_\pi(a, \boldsymbol{x})\right\| = o_P(1),$$

$$\sup_{a,\boldsymbol{x}} \left\|\sqrt{n}\{\hat{Q}_Y(\boldsymbol{x}, a) - Q_Y(\boldsymbol{x}, a)\} - \sqrt{n}(\hat{V}_n - V_0)^T D_Y(a, \boldsymbol{x})\right\| = o_P(1),\ and$$

$$\sup_{a,\boldsymbol{x}} \left\|\sqrt{n}\{\hat{Q}_Z(\boldsymbol{x}, a) - Q_Z(\boldsymbol{x}, a)\} - \sqrt{n}(\hat{V}_n - V_0)^T D_Z(a, \boldsymbol{x})\right\| = o_P(1).$$

3. *Let*

$$\tilde{D}^\theta_{\tilde{a},1}(u) = \Big(- \frac{\{\omega_\theta(\boldsymbol{x})y + (1 - \omega_\theta(\boldsymbol{x}))z\}1(a = \tilde{a})}{\pi^2(a;\boldsymbol{x})} + \frac{1(a = \tilde{a})}{\pi^2(a;\boldsymbol{x})} Q^2_\theta(\boldsymbol{x}, a)\Big) D_\pi(a, \boldsymbol{x}),$$

$$\tilde{D}^\theta_{\tilde{a},2}(u) = -\frac{1(a = \tilde{a}) - \pi(a;\boldsymbol{x})}{\pi(a;\boldsymbol{x})} \omega_\theta(\boldsymbol{x}) D_Y(a, \boldsymbol{x}),$$

and $$\tilde{D}^\theta_{\tilde{a},3}(u) = -\frac{1(a = \tilde{a}) - \pi(a;\boldsymbol{x})}{\pi(a;\boldsymbol{x})}(1 - \omega_\theta(\boldsymbol{x})) D_Z(a, \boldsymbol{x}).$$

Also, define $D^\theta_{\tilde{a}}(U) = \sum_{j=1}^3 \tilde{D}^\theta_{\tilde{a},j}(U)$ *such that*

$$\sup_{u,\tilde{a},\theta} \left|\sqrt{n}\big(\hat{W}_{\tilde{a},\theta}(u) - W_{a,\theta}(u)\big) - \sqrt{n}(\hat{V}_n - V_0)^T D^\theta_{\tilde{a}}(u)\right| = o_P(1),$$

where $\sum_{\tilde{a}=-1,1} E\|D^\theta_{\tilde{a}}(u)\|^2 < \infty$*.*

Theorem 2.4 (Asymptotic distribution of the boundary function)**.** *Define* $N_\theta = \mathbb{E}[\sum_a D^\theta_a(U)\dot{\phi}(a \cdot \operatorname{sgn}(W_{a,\theta}(U))\gamma^{*T}_\theta \xi(\boldsymbol{X}))a\xi(\boldsymbol{X})^T]$ *for* $f^*_\theta(\boldsymbol{X}) = \gamma^{*T}_\theta \xi(\boldsymbol{X})$*. Also, define* $B_0 = -A^{-1}_{1,0}A_{2,0}$*, a* $q \times p$

matrix, where

$$
\begin{aligned}
A_{1,0} &= -\mathbb{E}\big[\sum_{a\in\{1,-1\}} |W_{a,\theta_0}(U)|\ddot{\phi}\{a\cdot \operatorname{sgn}(W_{a,\theta}(U)){\gamma_{\theta_0}^*}^T\xi(\boldsymbol{X})\}\xi(\boldsymbol{X})\xi(\boldsymbol{X})^T\big], \text{ and} \\
A_{2,0} &= -\mathbb{E}\Big[\sum_a a\cdot\Big(\frac{I(A=a)}{\pi(a;\boldsymbol{X})}(Y-Z) - \frac{I(A=a)-\pi(a;\boldsymbol{X})}{\pi(a;\boldsymbol{X})}(Q_Y(\boldsymbol{X})-Q_Z(\boldsymbol{X}))\Big) \\
&\qquad\quad \cdot\omega_{\theta_0}(\boldsymbol{X})(1-\omega_{\theta_0}(\boldsymbol{X}))\dot{\phi}\{a\cdot \mathit{sgn}(W_{a,\theta_0}(U)){\gamma_{\theta_0}^*}^T\xi(\boldsymbol{X})\}\xi(\boldsymbol{X})\boldsymbol{X}^T\Big] \\
&\quad - 2\mathbb{E}\Big[\sum_a W_{a,\theta_0}(U)\Big(\frac{I(A=a)}{\pi(a;\boldsymbol{X})}(Y-Z) - \frac{I(A=a)-\pi(a;\boldsymbol{X})}{\pi(a;\boldsymbol{X})}(Q_Y(\boldsymbol{X}) \\
&\qquad\quad - Q_Z(\boldsymbol{X}))\Big)\cdot\omega_{\theta_0}(\boldsymbol{X})(1-\omega_{\theta_0}(\boldsymbol{X}))\ddot{\phi}\big\{a\cdot\operatorname{sgn}(W_{a,\theta_0}(U)){\gamma_{\theta_0}^*}^T\xi(\boldsymbol{X})\big\}{\gamma_{\theta_0}^*}^T\xi(\boldsymbol{X}) \\
&\qquad\quad \cdot\xi(\boldsymbol{X})\boldsymbol{X}^T\Big| W_{a,\theta_0}(U)=0\Big].
\end{aligned}
$$

For $f_\theta(X) \in \mathcal{F} = \{f(\boldsymbol{X}) : f(\boldsymbol{X}) = \gamma^T\xi(\boldsymbol{X})\}$, *denote* $m_\gamma^\theta(U) = -\sum_a |W_{a,\theta}(U)|\cdot \phi\big\{\mathit{sgn}\big(W_{a,\theta}(U)\big)a\gamma^T\xi(\boldsymbol{X})\big\}$, *and* $m_{\gamma_\theta^*}^{(j),\theta}(U) = \big(\frac{\partial}{\partial\gamma}\big)^j m_\gamma^\theta(U)\Big|_{\gamma=\gamma_\theta^*}$ *for* $j=1,2$. *Let* $\hat{f}_{n,\hat{\theta}_n}(\boldsymbol{X})$ *be the estimator from (2.4) where* $\hat{\theta}_n = \arg\max_{\theta\in\Theta}\widetilde{L}_n(\theta)$, *and denote* $\hat{f}_{n,\theta}(\boldsymbol{X}) = \hat{\gamma}_{n,\theta}^T\xi(\boldsymbol{X})$. *Also, let* $\tilde{U}_{1h} = \sqrt{n}(\hat{\theta}_n - \theta_0)$. *Then,*

$$
\sqrt{n}\big(\hat{\gamma}_{n,\hat{\theta}_n} - \gamma_{\theta_0}^*\big) - B_0\tilde{U}_{1h} \rightsquigarrow -V_{\theta_0}^{-1}\tilde{Z}, \tag{2.9}
$$

where $\tilde{Z}$ *is mean zero Gaussian process with covariance*

$$
\begin{aligned}
A_{\theta_0} = \mathbb{E}\big[m_{{\gamma_{\theta_0}}^*}^{(1),\theta_0}(U)m_{\gamma_{\theta_0}^*}^{(1),\theta_0}(U)^T\big] + N_{\theta_0}^T\mathbb{E}\big[\psi_V(U)\psi_V^T(U)\big]N_{\theta_0} \\
+ \mathbb{E}\big[m_{\gamma_{\theta_0}^*}^{(1),\theta_0}(U)\psi_V^T(U)\big]N_{\theta_0} + N_{\theta_0}^T\mathbb{E}\big[\psi_V(U)m_{\gamma_{\theta_0}^*}^{(1),\theta_0}(U)^T\big],
\end{aligned}
$$

and

$$
V_{\theta_0} = \mathbb{E}\Big[\sum_a -|W_{a,\theta_0}(U)|\ddot{\phi}\big(a\cdot \mathit{sgn}(W_{a,\theta_0}(U)){\gamma_{\theta_0}^*}^T\xi(\boldsymbol{X})\big)\xi(\boldsymbol{X})\xi(\boldsymbol{X})^T\Big].
$$

With the limiting distribution for the estimator of the boundary function $\hat{f}_{n,\hat{\theta}_n}(\cdot)$, we can now derive the limiting distribution of $(\hat{\theta}_n, \hat{\beta}_n)$. Recall that $P_\beta(x) = \text{expit}(x^T\beta)$. Also, define $I_n(\beta) = \mathbb{E}_n\big[P_\beta(\boldsymbol{X}\{1 - P_\beta(\boldsymbol{X})\boldsymbol{X}\boldsymbol{X}^T\big]$, and $I_0 = \mathbb{E}\big[P_{\beta_0}(\boldsymbol{X}\{1 - P_{\beta_0}(\boldsymbol{X})\boldsymbol{X}\boldsymbol{X}^T\big]$. We let $Z_{A,n} = n^{-1/2}\sum_{i=1}^n \psi_A(U_i)$, where $\psi_A(U_i) = \big[1\{A_i = d^*_{\theta_0}(\boldsymbol{X}_i)\} - P_{\beta_0}(\boldsymbol{X}_i)\big]\boldsymbol{X}_i$ is an independent and identically distributed influence vector for the unique parameters of $\pi(a;\boldsymbol{X})$ and $\mathbb{E}[\psi_A\psi_A^T] < \infty$. Further, we assume the following conditions.

Assumption 2.10. *For $\tilde{Z} \in \mathbb{R}^q$, $\tilde{U} \in \mathbb{R}^p$, and a $q \times p$ matrix B, define*

$$k(\tilde{Z}, \tilde{U}) = \mathbb{E}\big(\boldsymbol{X}\{2P_{\beta_0}(\boldsymbol{X}) - 1\}\big|(-V_{\theta_0}^{-1}\tilde{Z} + B\tilde{U})^T\xi(\boldsymbol{X})\big|\big|f^*_{\theta_0}(\boldsymbol{X}) = 0\big).$$

Assume that $M(\tilde{U}) = \beta_0^T k(\tilde{Z}, \tilde{U})$ has a unique and finite minimum over $\mathbb{R}^p$ for all $\tilde{Z} \in \mathbb{R}^q$.

Assumption 2.11. *We assume the following conditions.*

(a) The random variable $f^*_\theta(X)$ has a continuous density function g in a neighborhood of 0 with $g_0 = g(0) \in (0, \infty)$.

(b) The conditional distribution of $\boldsymbol{X}$ given that $|f^*_{\theta_0}(\boldsymbol{X})| \leq \epsilon$ converges to a non-degenerate distribution as $\epsilon \downarrow 0$.

(c) There exist δ_1 and δ_2 such that

$$\lim_{\epsilon \downarrow 0} \inf_{t \in S^p} \Pr\big[|\boldsymbol{X}^T\beta_0| \geq \delta_1, |(-V_{\theta_0}^{-1}\tilde{Z} + B_0 t)^T\xi(\boldsymbol{X})| \geq \delta_1\big||f^*_{\theta_0}(\boldsymbol{X})| \leq \epsilon\big] \geq \delta_2,$$

where $\tilde{Z}$ is a tight mean zero Gaussian process with covariance A_{θ_0}, and S^p is the p-dimensional unit sphere.

Theorem 2.5. *Let*

$$\Sigma_0 = \begin{bmatrix} \Sigma_A & -\mathbb{E}\Big[\psi_A V_{\theta_0}^{-1}\{m^{(1),\theta}_{\gamma^*_{\theta_0}} + N_{\theta_0}^T\psi_V\}^T\Big] \\ -\mathbb{E}\Big[\{m^{(1),\theta}_{\gamma^*_{\theta_0}} + N_{\theta_0}^T\psi_V\}V_{\theta_0}^{-1}\psi_A^T\Big] & A_{\theta_0} \end{bmatrix},$$

where $\Sigma_A = \mathbb{E}[\psi_A \psi_A^T]$*. Then, under the aforementioned assumptions,*

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\beta}_n - \beta_0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} \tilde{U} \\ I_0^{-1}\{Z_A - k(\tilde{Z}, \tilde{U})\} \end{pmatrix}, \tag{2.10}$$

where $\tilde{U} = \arg\min_{u \in \mathbb{R}^p} M(u)$*, and* $(Z_A^T, \tilde{Z}^T)^T \sim N(0, \Sigma_0)$*.*

2.4 Simulation Studies

In this section, we examine the performance of the suggested estimator via simulation studies. For $n = 100, 200, 500,$ and 1000, we generated $\boldsymbol{X} = (X_1, X_2)$ where $X_p \sim N(0,1)$ for $p = 1, 2$ and $Y = X_1^2 + A(4X_1^2 - 2X_2) + \epsilon_Y,\ Z = -2X_1 + A(2X_1^2 - 4X_2 - 1) + \epsilon_Z$. We assume $\omega_\theta(\boldsymbol{X}) =$ expit$(0.5 - X_1)$ where $U_\theta(\boldsymbol{X}, Y, Z) = \omega_\theta(\boldsymbol{X})Y + \{1 - \omega_\theta(\boldsymbol{X})\}Z$. Also, for the propensity score model, we assume Pr$(A = d_\theta^*(\boldsymbol{X})|\boldsymbol{X}) =$ expit$(1.5 - X_1)$. We repeated generating a data set and estimating parameters 500 times.

For EARL estimation of the optimal treatment, the logistic loss was implemented with $\lambda_{nj} = 2^{-5}$. Also, since the outcome model now incorporates the utility function, i.e., $U_\theta(\boldsymbol{X}) = \omega_\theta(\boldsymbol{X})Y + \{1 - \omega_\theta(\boldsymbol{X})\}Z$, the formula for the outcome model heavily depends on the utility function. Therefore, we applied linear approximation and confirmed that the correct outcome model is nearly approximated by $X_1^3 + X_1^2 + X_1 * X_2 + X_1 + X_2$. Regarding the Metropolis algorithm, we built a Markov chain of $10,000$ length in each replication. The result for this setting is presented in Table 2.3. As the sample size increases, root mean squared errors (RMSE) of both $\hat{\theta}_n$ and $\hat{\beta}_n$ decrease.

n	RMSE of $\widehat{\theta}_n$	RMSE of $\widehat{\beta}_n$	Error rate	Median(25th, 75th)
100	0.40 (0.34)	1.01 (0.30)	0.13 (0.05)	0.12 (0.09, 0.16)
200	0.34 (0.26)	1.00 (0.22)	0.12 (0.04)	0.12 (0.10, 0.15)
500	0.26 (0.12)	0.95 (0.19)	0.13 (0.04)	0.12 (0.10, 0.15)
1000	0.24 (0.09)	0.88 (0.18)	0.14 (0.04)	0.13 (0.11, 0.18)

Table 2.3: Estimation results for simulations where both utility and probability of optimal treatment are variable

Table 2.4 summarizes value estimates of the optimal policies with θ_0, estimated policies with $\hat{\theta}_n$ by the suggested method, policies when only Y is considered in the outcome model ($\omega_\theta(\boldsymbol{X}) \approx 1$), and policies when only Z is considered in the outcome model ($\omega_\theta(\boldsymbol{X}) \approx 0$), respectively. It is reasonable to conclude that the estimated policy yields notable improvement over the policy in only Y is considered, the policy that only Z is considered, or the standard of care.

n	Optimal	Estimated	Y only	Z only	Standard of care
100	2.73 (0.16)	2.56 (0.23)	2.18 (0.25)	1.41 (0.31)	1.25 (0.68)
200	2.73 (0.17)	2.62 (0.20)	2.23 (0.20)	1.45 (0.21)	1.25 (0.44)
500	2.74 (0.16)	2.64 (0.17)	2.22 (0.17)	1.46 (0.19)	1.24 (0.30)
1000	2.74 (0.18)	2.65 (0.16)	2.24 (0.17)	1.47 (0.18)	1.24 (0.21)

Table 2.4: Value results for simulations where both utility and probability of optimal treatment are variable

Additionally, in order to check the doubly robust property of the suggested estimator, we compare the settings that assumed the incorrect outcome model or the incorrect propensity score model. The following are four different settings for the comparison.

- Correct specification of both the outcome model and the propensity score (CC): $U_\theta(\boldsymbol{X}, Y, Z) \sim X_1^3 + X_1^2 + X_1 * X_2 + X_1 + X_2 + A(X_1^3 + X_1^2 + X_1 * X_2 + X_1 + X_2),\ \pi(a; \boldsymbol{X}) \sim X_1$.
- Incorrect specification of the outcome model and the correct model for the propensity score (CI): $U_\theta(\boldsymbol{X}, Y, Z) \sim X_1 + X_2 + A(X_1 + X_2),\ \pi(a; \boldsymbol{X}) \sim X_1$.
- Correct specification of the outcome model and the incorrect model for the propensity score model (IC): $U_\theta(\boldsymbol{X}, Y, Z) \sim X_1^3 + X_1^2 + X_1 * X_2 + X_1 + X_2 + A(X_1^3 + X_1^2 + X_1 * X_2 + X_1 + X_2),\ \pi(a; \boldsymbol{X}) \sim 1$.
- Incorrect specification of both the outcome model and the propensity score model (II): $U_\theta(\boldsymbol{X}, Y, Z) \sim X_1 + X_2 + A(X_1 + X_2),\ \pi(a; \boldsymbol{X}) \sim 1$.

Figure 2.2 presents four boxplots of estimated values in different settings. It appears that as n increases, the variance of value estimates decreases in all n. Also, although the estimation could be worsened when the Q-function is wrongly assumed than when the propensity score model is incorrect, it is reasonable to conclude that the impact of misspecification of the Q-function also significantly decreases when n is greater than 500.

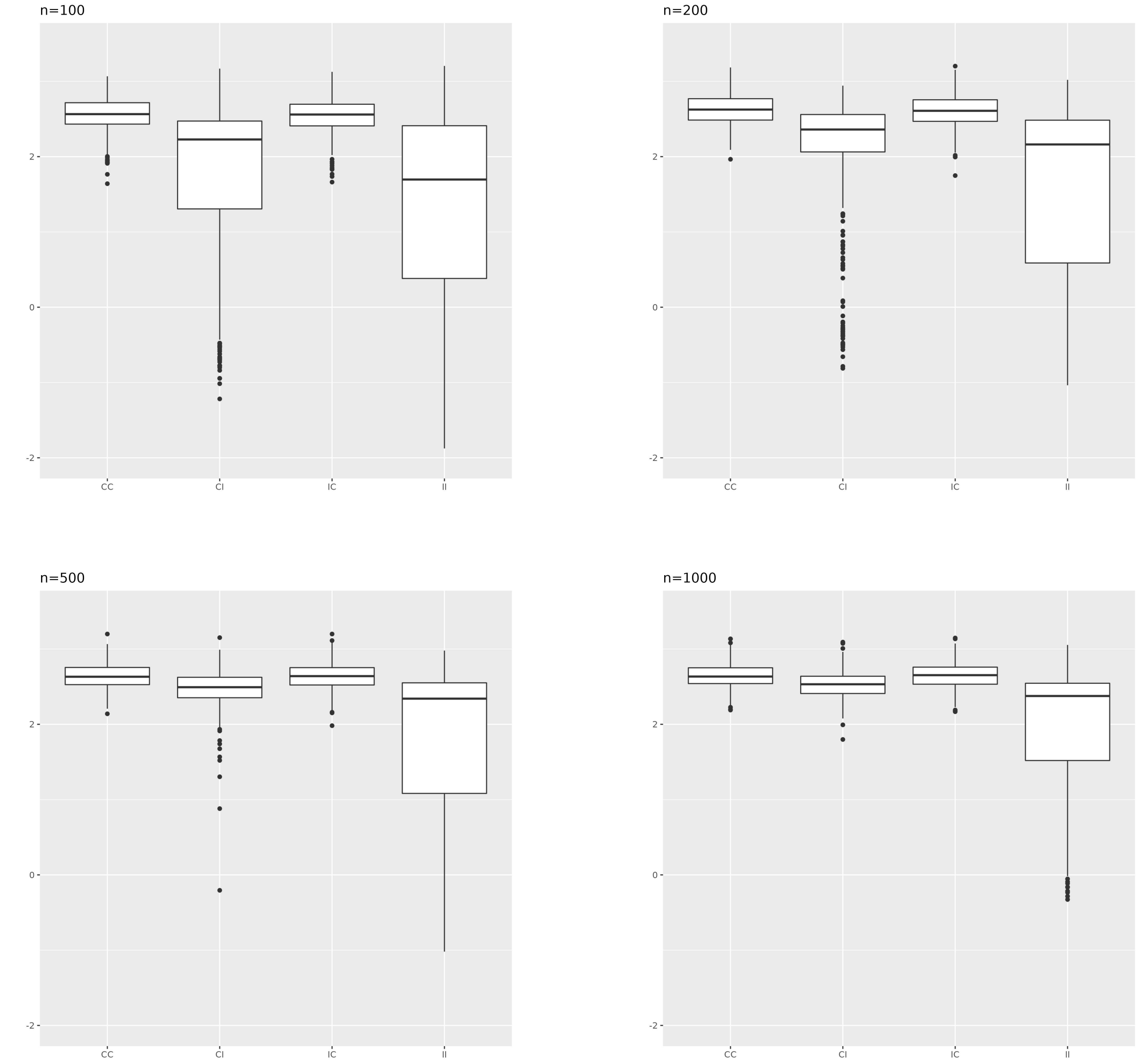

Figure 2.2: Boxplots of estimated values in four settings by $n = 100, 200, 500, 1000$, Y-axis: values

2.5 The Analysis of the STEP-BD Standard Care Pathway

Bipolar disorder is known for its two oppositing symptoms, depression and mania. In order to treat bipolar disorder, an antidepressant can be used; however, it has not been a standard treatment since there is a possibility of worsening the mania episode or triggering side effects in some patients. To reveal the relationship between antidepressants and these two symptoms, The Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), a project that includes a randomized trial and a large observational study, was established (Sachs et al., 2007). In this chapter, we applied the suggested method to the observational portion of the STEP-BD and investigated if the suggested precision medicine approach resulted in improvement in the symptoms of individuals similar to Luckett et al. (2021). For more detailed information on the STEP-BD and antidepressants in the study, we recommend the readers to see Section 5 of Luckett et al. (2021).

There are two outcomes that we consider in this study, the SUM-D score for depression episodes and the SUM-M score for mania episodes. Also, there are ten antidepressants (Deseryl, Serzone, Citalopram, Escitalopram, Oxalate, Prozac, Fluvoxamine, Paroxetine, Zoloft, Venlafaxine, Bupropion) all of which are considered as treatments. We used a logistic regression of the propensity score for the observational data to configure elements in our algorithm. Moreover, we utilized the randomized portion of STEP-BD by fitting a linear regression to each of the SUM-D and the SUM-M. As a result, we identified substance abuse and race at the significance level of 0.05 as potential confounders and used these variables to construct the outcome model, propensity score model, utility function, and the model for the probability of correct treatment assignment. We assumed that the lower SUM-D and SUM-M scores were desirable. Also, for the estimation of ITRs, we used the logistic loss as a surrogate for the indicator function.

Table 2.5 presents the improvement of value and estimates of the parameters. We used five-fold cross-validation. We computed the value by IPWE in the validation portion using the treatment regime that was estimated from the training portion, then averaged the five resulting value estimates. The value of standard care is calculated by IPWE using the estimated weights (utilities) and the observed outcomes. The estimated value is 6.6% greater than the value of standard care, implying that patients benefit 6.6% more from the estimated policy. $\mathbb{E}_n[\omega_{\hat{\theta}_n}(\boldsymbol{x})]$ indicates the average of the estimated weights in the utility functions. $\mathbb{E}_n[\hat{\rho}_n(\boldsymbol{x})]$ refers to the average of the estimated probabilities of the correct recommendation, i.e., $\rho(\boldsymbol{x}) = \text{expit}(\beta^T\boldsymbol{x})$. The resulting estimated $\hat{\theta}_n$ from the suggested method is presented in Table 2.6.

Policy	SUM-D	SUM-M	Value (% improvement)	$\widehat{\omega}_n$	$\widehat{\rho}_n$
Suggested method	2.338	0.843	6.6%	0.036	0.428
Standard of care	2.480	0.868	0.0%	·	·

Table 2.5: Results of analysis of STEP-BD data for SUM-D and SUM-M

	Intercept	Substance	Race
Estimate	-1.038	-1.677	-0.212

Table 2.6: Estimates of $\widehat{\theta}_n$ in the policy by the suggested method

2.6 Discussion

The philosophy of precision medicine and its usefulness has drawn attention, and methods for estimating dynamic treatment regimes have been extensively developed in various settings. Among the developed methods, Luckett et al. (2021) notably introduced a method for estimating the ITR when there are two outcomes to be considered, advancing from the single outcome case and pioneering to multiple outcome setting. However, due to the nonlinearity of the utility function in many cases, there is a need for some guarantee of robustness when the outcome model, which includes the utility function, is not correctly specified. Thus, it is reasonable to seek an improved approach that does not affect the estimation much under a misspecified outcome model. The suggested method achieves robustness of estimating the parameters for the patient-specific composite outcomes and further optimizes the ITR considering the heterogeneity of individuals.

One major advantage of the proposed method is that it alleviates the burden of determining the correct model for the outcome. To magnify this benefit, we suggest employing a doubly robust approach when there are more than two outcomes to be considered as an extension of this research, which opens up new possibilities for optimizing patient-specific outcomes and their ITRs when there are multiple entangled diagnostic results. Also, developing doubly-robust estimators for combining outcomes of various data types (e.g., survival outcomes) would be a huge advance in the study of composite outcomes. The development of doubly robust estimation in multiple outcome settings has considerable potential in precision medicine research, potentially advancing the widespread use of composite outcomes in clinical research.

CHAPTER 3: ESTIMATION OF COMPOSITE OUTCOMES IN THE MULTI-ARMED SETTING

3.1 Introduction

The methods suggested in Luckett et al. (2021) and Chapter 3 are developed specifically for binary treatment cases. Therefore, in this chapter, we extend the previous methods to accommodate a multi-armed setting, i.e., $A \in \mathcal{A} = \{1, \cdots, K\}$ where $K > 2$. In order to achieve this, we employ an estimator that could identify complicated boundaries of optimal treatments within the inverse reinforcement learning framework and subsequently obtain an estimator for a composite outcome.

To estimate the boundary function for multiple treatments, we first utilize AD-learning (Qi et al., 2020). AD-learning applies the angle-based approach by Zhang and Liu (2014), which projects each treatment to K simplex vertices. Additionally, we employ SD-learning by Shah et al. (2022), which uses a reweighing technique for a heterogeneous variance of outcomes of patients.

In this chapter, we present preliminary work on estimating the utilities of outcomes in the multi-armed setting. In Section 3.2, we introduce the algorithms for estimating the utilities, which include AD-learning and SD-learning in estimating treatment rules. In Section 3.3, we present the simulation results that validate the performance of the suggested estimators. Finally, in Section 3.4, we conclude this topic with a summary.

3.2 Overview of the methods

We assume the same setting as 2.2.1. We recall the pseudo likelihood framework from Luckett et al. (2021). For $\hat{d}_{n,\theta}$ an estimator for the optimal treatment regime d^*_θ, the pseudo logistic regression likelihood is

$$\hat{\mathcal{L}}_n(\theta, \beta) \propto \prod_{i=1}^{n} \frac{\exp\left[\boldsymbol{X}_i^T \beta 1\{A_i = \hat{d}_{n,\theta}(\boldsymbol{X}_i)\}\right]}{1 + \exp(\boldsymbol{X}_i^T \beta)}.$$

For estimation of the boundary function $\hat{f}_{n,\theta}(\boldsymbol{X})$ where $\hat{d}_{n,\theta}(\boldsymbol{X}) = \text{sgn}(\hat{f}_{n,\theta})$ from SD-learning (Shah et al., 2022) and AD-learning (Qi et al., 2020), the angle-based approach of Zhang and Liu (2014) should be preceded. Define e_i a $K-1$ dimensional zero vector where 1 is located at *i*th location. Let treatment A be expressed as the vector $u_A \in \mathbb{R}^{K-1}$ such that

$$u_A = \begin{cases} \frac{1}{\sqrt{K-1}} \mathbf{1}_{K-1}, & A = 1 \\ \sqrt{\frac{K}{K-1}} e_{A-1} - \frac{1+\sqrt{K}}{\sqrt{(K-1)^3}} \mathbf{1}_{K-1}, & 2 \leq A \leq K. \end{cases} \tag{3.11}$$

This representation allows treatment A to project into K simplex vertices in $\mathbb{R}^{K-1}$.

We use the following working model by Qi et al. (2020).

$$\frac{K}{K-1} U_\theta = U_A^T f(\boldsymbol{X}) + \epsilon,$$

where U_A is a random vector such that $U_A | (\boldsymbol{X}, A) \overset{\text{a.s.}}{=} u_A$. In order to use AD-learning and SD-learning in the observational data, we use the estimate $\hat{\pi}(a, \boldsymbol{x})$ for $\pi(a, \boldsymbol{x})$ calculated by machine learning techniques. Assume $\mathbb{E}[\epsilon | A, \boldsymbol{X}] = 0$ and $Var(\epsilon | A, \boldsymbol{X}) = \sigma_0^2(A, \boldsymbol{X})$. Also, assume $\mathcal{F} = \{f(\boldsymbol{X}) = B^T \boldsymbol{X} : B \in \mathbb{R}^{P \times (K-1)}\}$. For each $\theta \in \Theta$, the optimal ITR by AD-learning is

$$d_\theta^*(\boldsymbol{X}) = \arg \max_{k \in \{1, \cdots, K\}} u_k^T f_\theta^*(\boldsymbol{X}),$$

where $f_\theta^*(X)$ is a function maps from $\mathbb{R}^{p+1}$ to $\mathbb{R}^{K-1}$, and

$$f_\theta^* \in \arg \min_{f \in \mathbb{R}^{K-1}} \mathbb{E}\Big[\frac{1}{\pi(A; \boldsymbol{X})} \big\{\frac{K}{K-1} U_\theta - U_A^T f(\boldsymbol{X})\big\}^2\Big].$$

Also, for utilizing SD-learning, do the following suggested in Shah et al. (2022) for each $\theta \in \Theta$, We could utilize machine learning methods such as random forests, XGBoost, or SuperLearner suggested in Shah et al. (2022) to estimate the squared residuals for weights.

```
Algorithm 2: SD-learning for the estimator of the boundary function $f_\theta^*$
1 Obtain an AD-estimator $\hat{B}_{n,\theta}^{\text{AD}} = \arg\min_{B \in \mathbb{R}^{P \times (K-1)}} \mathbb{E}_n \big[ \frac{1}{\hat{\pi}(a,\boldsymbol{x})} \big( \frac{K}{K-1} U_\theta - u_a^T B^T x \big) \big]$
    where $\hat{\pi}(a, \boldsymbol{x})$ is an estimator of $\pi(a, \boldsymbol{x})$;
2 Obtain the squared residuals, $\big\{ \frac{K}{K-1} U_\theta - u_A^T (\hat{B}_{n,\theta})^T \boldsymbol{X} \big\}^2$;
3 Regress the squared residuals from 2 on $(A, \boldsymbol{X})$, and obtain prediction function
    $\hat{\sigma}_n^2(A, \boldsymbol{X})$;
4 Estimate $\hat{B}_{n,\theta}^{\text{SD}} = \arg\min_{B \in \mathbb{R}^{P \times (K-1)}} \mathbb{E}_n \frac{1}{\hat{\sigma}_n(a,\boldsymbol{x})} \big( \frac{K}{K-1} U_\theta - u_A^T B^T \boldsymbol{x} \big)$;
```

When the utilities are fixed among patients, we specify a grid from 0 to 1 for $\omega \equiv \text{expit}(\theta)$, obtain a profile estimator $\hat{\beta}_n(\omega) = \arg\max_{\beta \in \mathbb{R}^p} \hat{L}_n(\omega, \beta)$ by using each value in the pre-specified grid, and select the value in the grid that produces the largest profile pseudo-likelihood. For patient-specific utility, by $\hat{f}_{n,\theta}(\boldsymbol{X})$, we use Metropolis algorithm to estimate $\hat{\theta}_n$ that provides the largest $\tilde{L}_n(\theta; \hat{d}_{n,\theta})$ as in Algorithm 1. We build a chain $(\theta^1, \cdots, \theta^B)$ and estimate $\hat{f}_{n,\theta^b},\ b = 1, \cdots, B$. Since 2.2.2 use the agreement of the observed treatment and the estimated treatment as outcomes, not the treatment itself, the logistic likelihood is still valid in multiple treatment settings.

3.3 Simulation studies

3.3.1 Fixed utility with homogeneous variance

In this subsection, we present the results of simulation studies with fixed utilities. Firstly, we present the case when the variance of patients is equivalent. Let $X_1 \sim N(0, 1)$ and

$X_2 \sim N(0,1)$. Also, let $Y = Q_Y(\boldsymbol{X}, A) + \epsilon_Y$ and $Z = Q_Z(\boldsymbol{X}, A) + \epsilon_Z$, where

$$Q_Y(X,A) = \begin{cases} 0.5 + X_1 + X_2 & A = 1 \\ 0.5 - X_1 - X_2 & A = 2 \\ 0.5 + X_1 - X_2 & A = 3 \end{cases}$$

and

$$Q_Z(X,A) = \begin{cases} -0.5 - 0.5X_1 + 2X_2 & A = 1 \\ -0.5 + 0.5X_1 + 2X_2 & A = 2 \\ -0.5 - 0.5X_1 - 2X_2 & A = 3, \end{cases}$$

and $\epsilon_Y \sim N(0, 0.5^2)$, and $\epsilon_Z \sim N(0, 0.5^2)$. We assume that the utilities of outcomes are fixed, i.e., $\omega_\theta(\boldsymbol{X}) \equiv \omega$, where $\omega \in [0,1]$. In this context, we denote $d^*_\omega(\boldsymbol{X})$ as the optimal treatment regime that maximizes $Q_\omega(\boldsymbol{X}, a) = \omega Q_Y(\boldsymbol{X}, a) + (1-\omega) Q_Z(\boldsymbol{X}, a)$. Additionally, we assume that the optimal treatments are assigned to patients with a probability of $\Pr\{A = d^*_\omega(\boldsymbol{X}) | \boldsymbol{X} = \boldsymbol{x}\} = \rho$. We obtained $\hat{d}_{n,\omega}(\boldsymbol{X})$ using 2. To estimate $\hat{\pi}(a; \boldsymbol{x})$ and $\hat{\sigma}^2_n(A, \boldsymbol{X})$, we applied random forests. We performed 500 replications for each scenario.

Table 3.7 presents the estimates of ω and ρ when $\omega = 0.25$ or 0.75, $\rho = 0.7$, 0.8 or 0.9, and $n = 100, 200, 500,$ and 1000. For each row, $\hat{\omega}_n$ and $\hat{\rho}_n$ were calculated by averaging the estimates from 500 replications. The error rate was also averaged over the 500 replications. The standard errors are provided in parentheses.

Table 3.8 shows value estimates that are averaged over 500 values that would have been produced when treatments are assumed to be assigned by optimal policies with ω_0, estimated policies with $\hat{\omega}_n$, policies with only Y ($\omega_0 = 1$), policies with only Z ($\omega_0 = 0$), or standard of care. Standard errors of 500 replications are provided in the parentheses in each scenario.

n	ω	ρ	$\widehat{\omega}_n$	$\widehat{\rho}_n$	Error rate
100	0.25	0.7	0.14 (0.12)	0.66 (0.05)	0.08 (0.03)
		0.8	0.20 (0.17)	0.74 (0.04)	0.09 (0.04)
		0.9	0.23 (0.17)	0.84 (0.03)	0.08 (0.02)
	0.75	0.7	0.68 (0.13)	0.67 (0.06)	0.08 (0.06)
		0.8	0.72 (0.10)	0.75 (0.04)	0.07 (0.03)
		0.9	0.77 (0.12)	0.84 (0.04)	0.08 (0.04)
200	0.25	0.7	0.19 (0.16)	0.67 (0.03)	0.09 (0.05)
		0.8	0.20 (0.15)	0.76 (0.03)	0.08 (0.02)
		0.9	0.16 (0.10)	0.84 (0.03)	0.08 (0.02)
	0.75	0.7	0.67 (0.12)	0.68 (0.05)	0.08 (0.08)
		0.8	0.69 (0.09)	0.77 (0.03)	0.06 (0.03)
		0.9	0.75 (0.11)	0.85 (0.03)	0.07 (0.02)
500	0.25	0.7	0.14 (0.11)	0.66 (0.03)	0.08 (0.05)
		0.8	0.13 (0.12)	0.75 (0.02)	0.08 (0.02)
		0.9	0.14 (0.10)	0.84 (0.02)	0.08 (0.02)
	0.75	0.7	0.67 (0.05)	0.67 (0.02)	0.05 (0.02)
		0.8	0.68 (0.08)	0.77 (0.02)	0.05 (0.02)
		0.9	0.70 (0.08)	0.85 (0.02)	0.06 (0.02)
1000	0.25	0.7	0.11 (0.09)	0.67 (0.02)	0.07 (0.02)
		0.8	0.13 (0.11)	0.75 (0.02)	0.08 (0.02)
		0.9	0.15 (0.10)	0.83 (0.01)	0.08 (0.02)
	0.75	0.7	0.64 (0.02)	0.67 (0.02)	0.06 (0.02)
		0.8	0.65 (0.05)	0.77 (0.02)	0.05 (0.02)
		0.9	0.67 (0.06)	0.86 (0.02)	0.05 (0.01)

Table 3.7: Estimation results for simulations where utility and probability of optimal treatment are fixed with homogeneous variance

3.3.2 Fixed utility with heterogeneous variance

In addition to the case when Y and Z are generated with homogeneous variance, we present the simulation results that Y and Z are generated in heterogeneous variance in this subsection. We assume $\epsilon_Y \sim N(0, 0.\sigma_0^2(X))$ and $\epsilon_Z \sim N(0, 0.\sigma_0^2(X))$, where $\sigma_0^2(X) = 0.25 + (X_1 - 1)^2$. We estimated the utility and the probability of assigning the correct treatments by utilizing both AD-learning and SD-learning. Table 3.9 provides the mean estimates of ω and ρ calculated by AD-learning, and Table 3.10 contains the value estimates of true policy with ω_0, estimated policy with $\hat{\omega}_n$, policy that maximizes only Y, and policy that maximizes only Z. The results in

n	ω	ρ	Optimal	Estimated ω	Y only	Z only	Standard of care
100	0.25	0.7	1.17 (0.04)	1.15 (0.04)	0.76 (0.16)	1.14 (0.05)	0.50 (0.14)
		0.8	1.17 (0.03)	1.15 (0.04)	0.87 (0.14)	1.13 (0.05)	0.73 (0.15)
		0.9	1.15 (0.04)	1.15 (0.04)	1.01 (0.09)	1.14 (0.04)	0.95 (0.12)
	0.75	0.7	1.47 (0.02)	1.44 (0.03)	1.39 (0.06)	1.02 (0.17)	0.91 (0.14)
		0.8	1.47 (0.03)	1.44 (0.04)	1.41 (0.04)	1.06 (0.13)	1.08 (0.10)
		0.9	1.46 (0.03)	1.44 (0.04)	1.41 (0.08)	1.11 (0.14)	1.27 (0.10)
200	0.25	0.7	1.16 (0.04)	1.13 (0.04)	0.74 (0.18)	1.13 (0.04)	0.57 (0.12)
		0.8	1.16 (0.03)	1.15 (0.03)	0.89 (0.15)	1.13 (0.04)	0.78 (0.11)
		0.9	1.17 (0.04)	1.15 (0.03)	0.99 (0.12)	1.14 (0.03)	0.97 (0.10)
	0.75	0.7	1.45 (0.03)	1.42 (0.03)	1.39 (0.05)	1.05 (0.16)	0.94 (0.09)
		0.8	1.46 (0.03)	1.44 (0.04)	1.41 (0.05)	1.09 (0.12)	1.11 (0.08)
		0.9	1.46 (0.03)	1.44 (0.03)	1.40 (0.08)	1.11 (0.12)	1.27 (0.09)
500	0.25	0.7	1.16 (0.03)	1.15 (0.05)	0.65 (0.14)	1.14 (0.04)	0.53 (0.08)
		0.8	1.16 (0.04)	1.15 (0.04)	0.85 (0.13)	1.15 (0.04)	0.73 (0.06)
		0.9	1.16 (0.03)	1.14 (0.03)	1.00 (0.08)	1.14 (0.04)	0.95 (0.06)
	0.75	0.7	1.46 (0.03)	1.44 (0.04)	1.41 (0.04)	1.05 (0.15)	0.91 (0.06)
		0.8	1.46 (0.03)	1.44 (0.03)	1.42 (0.03)	1.15 (0.09)	1.10 (0.06)
		0.9	1.46 (0.03)	1.45 (0.03)	1.42 (0.04)	1.17 (0.06)	1.28 (0.04)
1000	0.25	0.7	1.14 (0.04)	1.14 (0.04)	0.67 (0.12)	1.14 (0.04)	0.53 (0.05)
		0.8	1.15 (0.03)	1.15 (0.03)	0.84 (0.09)	1.15 (0.03)	0.74 (0.05)
		0.9	1.16 (0.03)	1.15 (0.03)	1.00 (0.06)	1.14 (0.03)	0.95 (0.04)
	0.75	0.7	1.46 (0.02)	1.42 (0.02)	1.39 (0.04)	1.06 (0.11)	0.91 (0.04)
		0.8	1.46 (0.03)	1.43 (0.03)	1.41 (0.04)	1.16 (0.08)	1.10 (0.04)
		0.9	1.45 (0.03)	1.44 (0.03)	1.41 (0.04)	1.18 (0.05)	1.28 (0.03)

Table 3.8: Value results for simulations where utility and probability of optimal treatment are fixed with homogeneous variance

these tables imply that the estimation of ω and ρ, as well as value estimation, are significantly impacted by poor estimation of optimal ITR.

Table 3.11 and Table 3.12 provide estimates of $\hat{\omega}_n$ and $\hat{\rho}_n$, and the value estimates with treatment rules obtained by SD-learning. The results in these two tables imply that SD-learning enables us to obtain estimates of utilities that yield values close to the optimal values.

Figure 3.3 presents line plots of values, assuming that patients had followed the optimal treatment rule, the rule estimated by SD-learning, and the rule estimated by AD-learning with heterogeneous variances when $n = 100, 200, 500$ and 1000. Based on the figure, we can conclude that when the variances of outcomes differ among patients, SD-learning is more suitable for estimating the treatment rule and utilities compared to AD learning.

n	ω	ρ	$\widehat{\omega}_n$	$\widehat{\rho}_n$	Error rate
100	0.25	0.70	0.70 (0.29)	0.35 (0.13)	0.62 (0.21)
		0.80	0.67 (0.28)	0.36 (0.19)	0.63 (0.26)
		0.90	0.58 (0.31)	0.45 (0.26)	0.54 (0.29)
	0.75	0.70	0.52 (0.34)	0.47 (0.15)	0.41 (0.22)
		0.80	0.62 (0.28)	0.53 (0.18)	0.38 (0.22)
		0.90	0.68 (0.26)	0.62 (0.22)	0.34 (0.23)
200	0.25	0.70	0.78 (0.27)	0.35 (0.11)	0.62 (0.18)
		0.80	0.71 (0.30)	0.35 (0.16)	0.63 (0.21)
		0.90	0.63 (0.30)	0.42 (0.23)	0.57 (0.26)
	0.75	0.70	0.47 (0.34)	0.50 (0.14)	0.37 (0.22)
		0.80	0.61 (0.27)	0.58 (0.16)	0.32 (0.21)
		0.90	0.68 (0.22)	0.65 (0.18)	0.30 (0.21)
500	0.25	0.70	0.90 (0.21)	0.34 (0.06)	0.64 (0.09)
		0.80	0.84 (0.25)	0.35 (0.11)	0.64 (0.15)
		0.90	0.75 (0.30)	0.39 (0.17)	0.59 (0.20)
	0.75	0.70	0.39 (0.34)	0.51 (0.14)	0.35 (0.24)
		0.80	0.65 (0.20)	0.65 (0.11)	0.21 (0.15)
		0.90	0.70 (0.13)	0.74 (0.09)	0.19 (0.11)
1000	0.25	0.70	0.95 (0.15)	0.35 (0.05)	0.64 (0.08)
		0.80	0.95 (0.14)	0.34 (0.06)	0.65 (0.09)
		0.90	0.88 (0.23)	0.37 (0.13)	0.62 (0.15)
	0.75	0.70	0.37 (0.35)	0.51 (0.15)	0.35 (0.26)
		0.80	0.66 (0.16)	0.68 (0.08)	0.17 (0.12)
		0.90	0.69 (0.08)	0.77 (0.05)	0.15 (0.06)

Table 3.9: Estimation results with AD-learning for simulations where utility and probability of optimal treatment are fixed with heterogeneous variance

n	ω	ρ	Optimal	Estimated ω	Y only	Z only	Standard of care
100	0.25	0.70	1.15 (0.03)	0.85 (0.26)	-0.02 (0.27)	0.86 (0.25)	0.53 (0.16)
		0.80	1.16 (0.03)	0.80 (0.34)	-0.06 (0.30)	0.77 (0.36)	0.73 (0.16)
		0.90	1.16 (0.04)	0.80 (0.37)	0.03 (0.44)	0.77 (0.38)	0.94 (0.14)
	0.75	0.70	1.45 (0.03)	1.20 (0.31)	1.03 (0.33)	0.64 (0.21)	0.92 (0.12)
		0.80	1.46 (0.03)	1.12 (0.34)	0.92 (0.38)	0.65 (0.27)	1.08 (0.12)
		0.90	1.45 (0.03)	1.14 (0.37)	0.95 (0.41)	0.68 (0.30)	1.26 (0.11)
200	0.25	0.70	1.16 (0.04)	0.93 (0.13)	0.00 (0.24)	0.93 (0.14)	0.53 (0.12)
		0.80	1.16 (0.04)	0.90 (0.17)	-0.03 (0.25)	0.91 (0.15)	0.74 (0.11)
		0.90	1.16 (0.04)	0.86 (0.26)	-0.03 (0.31)	0.85 (0.26)	0.95 (0.10)
	0.75	0.70	1.45 (0.03)	1.31 (0.19)	1.12 (0.24)	0.66 (0.19)	0.91 (0.09)
		0.80	1.46 (0.03)	1.27 (0.19)	1.05 (0.30)	0.65 (0.20)	1.09 (0.08)
		0.90	1.46 (0.03)	1.21 (0.29)	0.99 (0.35)	0.66 (0.24)	1.27 (0.07)
500	0.25	0.70	1.16 (0.03)	0.95 (0.08)	0.03 (0.14)	0.96 (0.07)	0.53 (0.07)
		0.80	1.17 (0.04)	0.95 (0.08)	0.00 (0.16)	0.96 (0.07)	0.74 (0.07)
		0.90	1.17 (0.04)	0.94 (0.09)	0.01 (0.20)	0.96 (0.09)	0.95 (0.06)
500	0.75	0.70	1.46 (0.03)	1.38 (0.07)	1.22 (0.14)	0.61 (0.13)	0.92 (0.06)
		0.80	1.47 (0.03)	1.37 (0.08)	1.20 (0.15)	0.60 (0.12)	1.10 (0.05)
		0.90	1.46 (0.03)	1.35 (0.13)	1.18 (0.19)	0.65 (0.16)	1.28 (0.05)
1000	0.25	0.70	1.16 (0.04)	0.95 (0.07)	0.05 (0.11)	0.97 (0.06)	0.52 (0.05)
		0.80	1.17 (0.03)	0.95 (0.06)	0.04 (0.11)	0.97 (0.05)	0.74 (0.05)
	0.25	0.90	1.16 (0.04)	0.95 (0.07)	0.02 (0.13)	0.97 (0.06)	0.95 (0.05)
	0.75	0.70	1.45 (0.03)	1.40 (0.05)	1.25 (0.09)	0.57 (0.09)	0.91 (0.04)
		0.80	1.46 (0.03)	1.40 (0.05)	1.25 (0.08)	0.57 (0.09)	1.09 (0.04)
		0.90	1.45 (0.03)	1.38 (0.05)	1.24 (0.10)	0.61 (0.12)	1.28 (0.04)

Table 3.10: Value results with AD-learning for simulations where utility and probability of optimal treatment are fixed with heterogeneous variance

n	ω	ρ	$\widehat{\omega}_n$	$\widehat{\rho}_n$	Error rate
100	0.25	0.70	0.24 (0.21)	0.66 (0.06)	0.10 (0.06)
		0.80	0.21 (0.18)	0.75 (0.05)	0.09 (0.02)
		0.90	0.22 (0.20)	0.83 (0.04)	0.09 (0.03)
	0.75	0.70	0.75 (0.16)	0.67 (0.06)	0.09 (0.07)
		0.80	0.80 (0.11)	0.76 (0.04)	0.08 (0.04)
		0.90	0.83 (0.11)	0.84 (0.05)	0.09 (0.04)
200	0.25	0.70	0.17 (0.16)	0.66 (0.04)	0.09 (0.05)
		0.80	0.17 (0.15)	0.75 (0.03)	0.08 (0.03)
		0.90	0.18 (0.16)	0.84 (0.03)	0.08 (0.02)
	0.75	0.70	0.74 (0.09)	0.68 (0.04)	0.06 (0.04)
		0.80	0.77 (0.09)	0.77 (0.03)	0.06 (0.03)
		0.90	0.82 (0.10)	0.85 (0.03)	0.08 (0.03)
500	0.25	0.70	0.14 (0.11)	0.67 (0.02)	0.07 (0.02)
		0.80	0.15 (0.12)	0.75 (0.02)	0.08 (0.02)
		0.90	0.13 (0.11)	0.84 (0.02)	0.08 (0.02)
	0.75	0.70	0.72 (0.06)	0.68 (0.02)	0.04 (0.02)
		0.80	0.76 (0.07)	0.77 (0.02)	0.05 (0.02)
		0.90	0.83 (0.09)	0.86 (0.02)	0.06 (0.02)
1000	0.25	0.70	0.13 (0.09)	0.67 (0.02)	0.07 (0.02)
		0.80	0.12 (0.08)	0.75 (0.02)	0.07 (0.02)
		0.90	0.12 (0.08)	0.84 (0.01)	0.07 (0.02)
	0.75	0.70	0.71 (0.04)	0.69 (0.01)	0.03 (0.01)
		0.80	0.75 (0.06)	0.77 (0.02)	0.04 (0.02)
		0.90	0.82 (0.08)	0.86 (0.02)	0.05 (0.02)

Table 3.11: Estimation results with SD-learning for simulations where utility and probability of optimal treatment are fixed with heterogeneous variance

n	ω	ρ	Optimal	Estimated ω	Y only	Z only	Standard of care
100	0.25	0.70	1.17 (0.04)	1.14 (0.05)	0.76 (0.20)	1.13 (0.05)	0.52 (0.15)
		0.80	1.17 (0.04)	1.15 (0.04)	0.90 (0.17)	1.13 (0.05)	0.73 (0.16)
		0.90	1.17 (0.03)	1.15 (0.04)	1.02 (0.10)	1.14 (0.05)	0.94 (0.14)
	0.75	0.70	1.47 (0.03)	1.44 (0.03)	1.40 (0.06)	1.03 (0.16)	0.90 (0.13)
		0.80	1.47 (0.03)	1.44 (0.03)	1.42 (0.05)	1.09 (0.12)	1.09 (0.13)
		0.90	1.46 (0.03)	1.43 (0.04)	1.42 (0.05)	1.14 (0.10)	1.27 (0.11)
200	0.25	0.70	1.16 (0.04)	1.14 (0.04)	0.74 (0.17)	1.13 (0.04)	0.54 (0.11)
		0.80	1.17 (0.04)	1.15 (0.04)	0.88 (0.14)	1.14 (0.04)	0.74 (0.11)
		0.90	1.17 (0.04)	1.16 (0.04)	1.01 (0.09)	1.15 (0.04)	0.95 (0.10)
	0.75	0.70	1.46 (0.03)	1.44 (0.04)	1.40 (0.05)	1.05 (0.13)	0.91 (0.09)
		0.80	1.47 (0.03)	1.44 (0.04)	1.42 (0.05)	1.11 (0.10)	1.09 (0.08)
		0.90	1.47 (0.03)	1.45 (0.04)	1.43 (0.05)	1.15 (0.08)	1.28 (0.08)
500	0.25	0.70	1.16 (0.04)	1.15 (0.04)	0.70 (0.13)	1.14 (0.04)	0.53 (0.07)
		0.80	1.16 (0.04)	1.15 (0.04)	0.86 (0.11)	1.14 (0.04)	0.74 (0.07)
		0.90	1.16 (0.04)	1.15 (0.04)	1.00 (0.07)	1.14 (0.04)	0.95 (0.06)
	0.75	0.70	1.46 (0.03)	1.45 (0.04)	1.41 (0.05)	1.08 (0.10)	0.92 (0.06)
		0.80	1.46 (0.03)	1.45 (0.04)	1.43 (0.04)	1.13 (0.07)	1.10 (0.06)
		0.90	1.45 (0.03)	1.44 (0.03)	1.43 (0.04)	1.16 (0.05)	1.28 (0.05)
1000	0.25	0.70	1.16 (0.04)	1.15 (0.04)	0.68 (0.12)	1.14 (0.04)	0.53 (0.05)
		0.80	1.16 (0.04)	1.15 (0.04)	0.84 (0.10)	1.14 (0.04)	0.74 (0.05)
		0.90	1.16 (0.04)	1.15 (0.04)	1.00 (0.06)	1.15 (0.04)	0.95 (0.04)
	0.75	0.70	1.46 (0.03)	1.45 (0.03)	1.41 (0.04)	1.09 (0.08)	0.91 (0.04)
		0.80	1.46 (0.03)	1.45 (0.03)	1.42 (0.03)	1.14 (0.06)	1.09 (0.04)
		0.90	1.46 (0.03)	1.45 (0.03)	1.44 (0.04)	1.18 (0.05)	1.28 (0.03)

Table 3.12: Value results with SD-learning for simulations where utility and probability of optimal treatment are fixed with heterogeneous variance

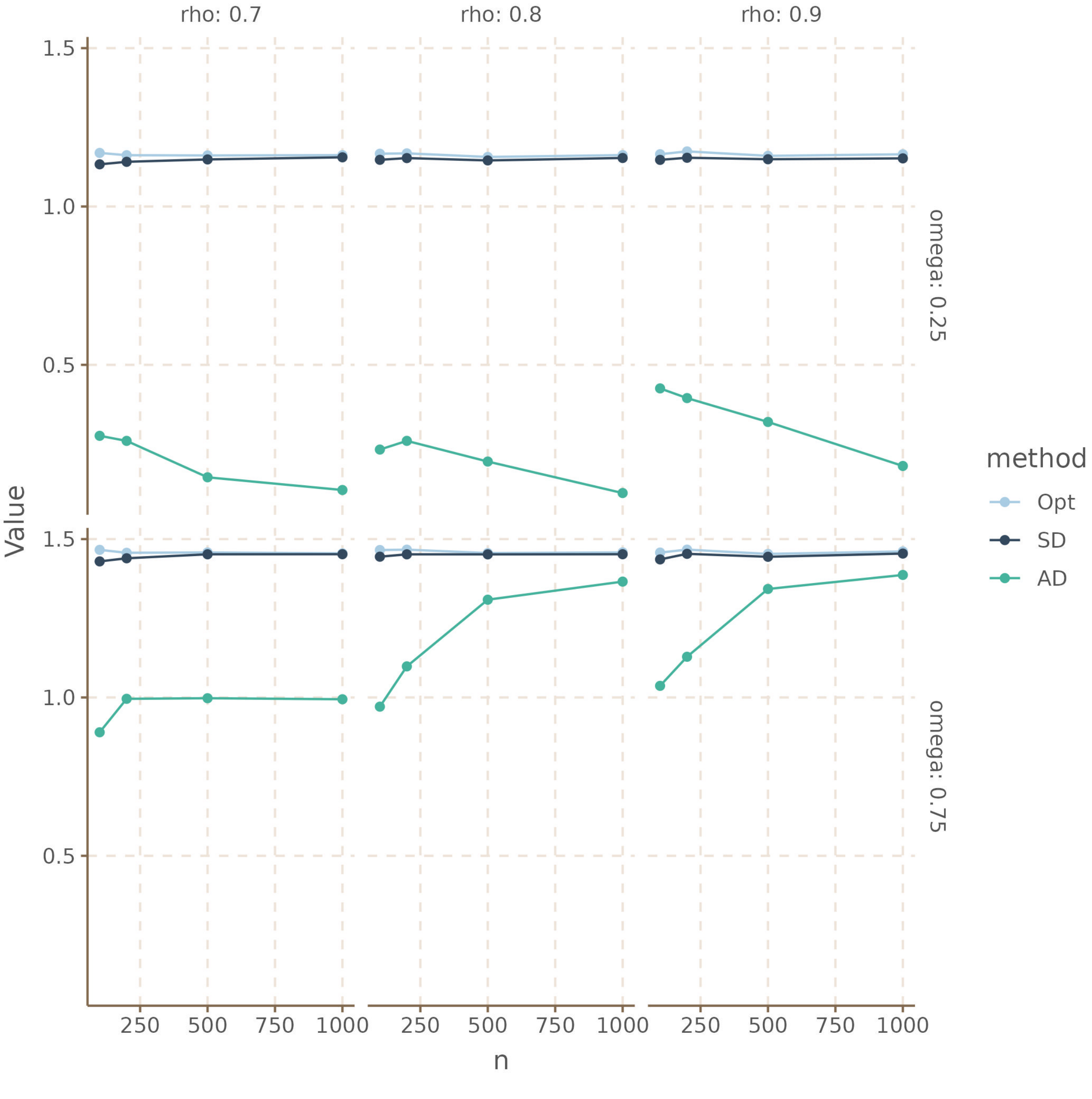

Figure 3.3: Values of each treatment rule; Opt: optimal treatment rule, SD: treatment rule estimated by SD-learning, AD: treatment rule estimated by AD-learning

3.4 Discussion

In this chapter, we suggested a methodology to obtain estimators for utilities of outcomes in cases when more than two treatments are available and presented preliminary results on simulations. The algorithms include AD-learning or SD-learning that enable the identification of complex boundaries and estimate ITRs beyond binary treatment options.

Through numerical experiments, we demonstrated that Algorithm 2 is guaranteed to provide estimators with strong performance. Moreover, the suggested algorithm provides an estimator that works well, particularly when the variances are different in patients. Therefore, we can conclude that Algorithm 2 is an effective approach for estimating a composite outcome, especially when the presence of heterogeneous variance is not certain.

However, we have empirically discovered that there is a potential issue of identifiability when $\rho < 0.7$. In Luckett et al. (2021), it was identified that the estimators of utilities are identifiable in a binary treatment setting with an inverse reinforcement learning framework when $\rho > 0.5$. Therefore, in a multi-treatment setting, there is still a need to determine the rigorous conditions for the identifiability of utilities, which will also provide conditions for determining K.

CHAPTER 4: FUTURE RESEARCH

This chapter will discuss future research directions for methods proposed in the three preceding chapters.

In Chapter 1, the suggested method, random forest informed tree-based learning, provides an interpretable treatment rule built by random forests that identified heterogeneous improvements in patients. However, there remains the possibility that the estimated final rules are suboptimal, although the estimated final rules provide superior value estimates. Therefore, as future studies, we suggest developing a method leveraged by random forests for deriving the maximum VF estimate with underlying factors that divide the individuals into disjoint subgroups.

In Chapter 2, the proposed method provides doubly robust estimators of customized utilities of patients. This indicates that more complicated formations are allowed for the outcome model. Therefore, we recommend implementing a doubly robust approach in a multi-outcome setting. Also, as future research, we recommend developing doubly robust estimators for different types of outcomes, such as survival outcomes. The development of doubly robust estimators in various outcome settings will bring a significant impact on precision medicine studies.

Moreover, we can generalize the current model for a composite outcome that is capable of embracing various models for utility functions and probability of correct recommendation.

For each $\theta \in \Theta$, let d^*_θ be the optimal ITR for preference u_θ. Assume that interventions are assigned such that

$$A = \begin{cases} d^*_{\tilde{\theta}}(\boldsymbol{X}) \text{ with probability } \zeta(\boldsymbol{x}; \beta) \\ -d^*_{\tilde{\theta}}(\boldsymbol{X}) \text{ with probability } 1 - \zeta(\boldsymbol{x}; \beta), \end{cases}$$

where $\zeta(\boldsymbol{x};\beta)$ is a parametric model indexed by $\beta \in \mathcal{B}$. Also, $\tilde{\theta} \sim N(\theta_0, \delta I_p)$, where $\delta \geq 0$, and I_p is an identity matrix of size p. Then, the likelihood is

$$\begin{aligned}\mathcal{L}_n(\beta,\theta,\delta) &= \prod_{i=1}^{n} f(\boldsymbol{X}_i)f(A_i|\boldsymbol{X}_i)f(Y_i,Z_i|\boldsymbol{X}_i,A_i)\\ &\propto \prod_{i=1}^{n}[\zeta(\boldsymbol{X}_i,\beta)\lambda(\boldsymbol{X}_i;\theta,\delta)+\{1-\zeta(\boldsymbol{X}_i,\beta)\}\{1-\lambda(\boldsymbol{X}_i,\theta,\delta)\}]^{1\{A_i=d^*_\theta(\boldsymbol{X}_i)\}}\\ &\qquad \times [\{1-\zeta(\boldsymbol{X}_i,\beta)\}\lambda(\boldsymbol{X}_i,\theta,\delta)+\lambda(\boldsymbol{X}_i,\beta)\{1-\lambda(\boldsymbol{X}_i;\theta,\delta)\}]^{1\{A_i\neq d^*_\theta(\boldsymbol{X}_i)\}},\end{aligned}$$

where $\lambda(\boldsymbol{x};\theta,\delta) = \int 1\{d^*_\theta(\boldsymbol{x}) = d^*_{\theta+\sqrt{\delta}\nu}(\boldsymbol{x})\}\tilde{\phi}(\nu)d\nu$ and $\tilde{\phi}$ is the density for a standard normal random vector. By introducing $\delta > 0$, the generalized model also allows the observations of the preference of patients to be not perfect.

In Chapter 3, we suggest a further rigorous investigation into the conditions for the identifiability of utilities when $\rho < 0.7$. By employing SD-learning, it is demonstrated numerically that the inverse reinforcement learning framework results in strong performance in a multi-armed setting. Therefore, it is reasonable to suggest theoretical proof as future research. By obtaining asymptotic consistency and weak convergence of the estimators, we can further strengthen the reliability of the suggested methodology.

APPENDIX A: TECHNICAL DETAILS FOR CHAPTER 1

This chapter contains technical details to the main text of Chapters .

A.1 Mathematical Definition

A.1.1 Value Function

$$V_0(d) = E[\frac{YI(d(\boldsymbol{X}) = A)}{P(A|\boldsymbol{X})}] \tag{A.12}$$

In (A.12), Y is outcome, $d(\boldsymbol{X})$ is an optimal treatment for a subject who has $\boldsymbol{X}$ as covariates and A is an observed treatment to a subject. $I(\cdot)$ is an indicator function. $P(A|\boldsymbol{X})$ is a propensity score, which is a probability of A given $\boldsymbol{X}$.

A.1.2 Jackknife estimator for estimating value function

$$\begin{gathered} \hat{V^{jk}}(\hat{d}_n) = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} w_i} \\ \text{where } \; u_i = y_i \frac{1\{a_i = \hat{d}_n^{(-i)}(\boldsymbol{x}_i)\}}{P(a_i|\boldsymbol{x}_i)} \; \text{ and } \; w_i = \frac{1\{a_i = \hat{d}_n^{(-i)}\}}{P(a_i|\boldsymbol{x}_i)} \end{gathered} \tag{A.13}$$

In (A.13), the jackknife approach is applied. It is computed by leaving one observation out, getting a model with the rest of the observations, then obtaining $\hat{d}$ using that model by plugging in the observation which has been left out.

A.1.3 Z-test statistic

$$T^{jk}(\hat{d}_{PMM},\hat{d}_{ZOM}) = \frac{\hat{V}^{jk}(\hat{d}_{PMM}) - \hat{V}^{jk}(\hat{d}_{ZOM})}{\sqrt{\frac{\sum_{i=1}^{n}(R^{jk}_{PMM} - R^{jk}_{ZOM})^2}{n(n-1)}}} \tag{A.14}$$

$$\text{where } R_i^{jk} = \frac{1}{\bar{W}_n}U_i - \frac{\bar{U}_n}{\bar{W}_n^2}W_i.$$

The p-value is defined as $p = 2P(|T| \leq z) = 2\int_{|T|}^{\infty} f(z)dz$ where $z \sim N(0,1)$.

A.2 Random Forest informed Tree-based Learning

A.2.1 Variable Selection

In order to obtain a set of candidate variables for split points in the algorithm, we utilized variable Importance using random forests (RF), using another leave-one-out cross-validation approach. First, we removed one patient from the data set and computed a Variable Importance plot using the remaining $(n-1)$ patients. We then obtained the top 7 variables from this Variable Importance Plot. We repeated this process for every patient in the data set, resulting in n sets of the top 7 variables. Then, we obtained the number of times each variable was selected as one of the top 7 variables out of n times. The leave-one-out cross-validation approach is used to obtain a more stable list of variables. We note that the 7 were chosen because from the many variable importance plots, 7 variables consistently appeared to be sufficient to adequately capture variable influence. Table A.13 shows a listing of the important variables for the outcomes at 12-month visit. The variables in the listing are then the candidates for which to identify cut points for each analysis.

Table A.13: Important variables for outcomes at 12-month visit

Variable Name	**Number of times selected as 1 of Top 7 variables**
WOMAC Total (WTO)	303
WOMAC Function (WF)	303
WOMAC Pain (WP)	303
WOMAC Stiffness (WS)	303
Satisfaction with physical function (Satis)	303
BMI	303
Self Efficacy Exercise (SE)	136
Age at baseline (Age)	80
Brief fear of movement (BF)	74
Education	8
Social support for exercise (family)	2
PROMIS Fatigue Score	2
phq8score	1

A.2.2 Algorithm

The algorithm starts with dividing the data set into two subgroups by utilizing all the distinct values of the variables in the important variable list as split points while avoiding subgroups having less than 15% of the sample size. Then, value function estimates are calculated after assigning an optimal treatment regime estimated by RF to the first subgroup and allocating one of PT, IBET, or WT to everyone in the other group. Out of all possible value function estimates, we retrieve the partition that produces the largest value function estimate and obtain another value function estimate after applying PT, IBET, or WT, whichever produces the largest estimate. We assessed whether this value function estimate is significantly greater than that of the ZOM and continued to the next iteration to search for a finer subgroup. In the second iteration, we only split the subgroup that received the optimal treatment estimated by Random Forest in the previous iteration. Then, we obtain the largest value function estimate as in the previous iteration and check for the significance of the value function estimate of the ZOM. We repeat this process until we have identified a set of disjoint subgroups whose Z statistic is statistically signicant to the Z statistic of the ZOM. After the third variable for the split point is chosen, we continue searching for a split point within the three variables in order to maintain feasible computations. Figure A.5 shows the flowchart for the algorithm. More detailed steps are given in the Mathematical Expression section below.

This algorithm determines a partition as a final rule that includes a sequence of subgroups accumulated through each step. In this approach, it is noteworthy that the method utilizes the advantage of Random Forests, which produces a flexible low-bias estimation, but concurrently removes its “black box” aspects.

A.2.3 Example

Figure A.5 depicts a detailed process of the algorithm when it is applied to the first data set that includes the outcome at month 4. In the first loop, the split point Age at baseline= 49.33 yields the partition that produces the largest value function estimate, $\hat{V}^{(1),jk} = 74.8662$, when IBET is given to the patients whose baseline age is less or equal to 49.33. (jk) in the superscript implies that it is calculated by the jackknife approach using the Random Forest. After assigning one of the three treatments to all patients who were initially assigned RF, three value function estimates can be calculated. However, since $\hat{V}^{(1)} = 71.1111$, which corresponds to IBET since that was the largest value function estimate among the three treatments, is the same treatment assignment as the ZOM ($\hat{V}_{IBET} = 71.1111$), we keep the split and continue to the second loop.

In the second loop, we leave the patients who have received IBET in the first loop and divide the group assigned to RF into two subgroups. Brief Fear of Movement Score (BF) =9 is the cut point with largest value function estimate $\hat{V}^{(2),jk} = 75.2701$. We assign a single treatment to patients whose BF is greater than 9. Out of three treatments, $\hat{V}^{(2)} = 71.1111$ with IBET is the largest, but this is still the same assignment to the ZOM with IBET. Thus, we continue to the third loop for the second time.

After the third loop, with the value function estimate still not being significantly large, we continue to the fourth loop within three variables, Age at baseline, BF and BMI. In other words, we restrict the number of variables to three variables for searching for a split point. The fourth split point is discovered to be BMI = 37.24, and $\hat{V}^{(4)} = 73.1939$ is not significantly greater than that of the ZOM (IBET), so we move to the next loop.

In the fifth loop, $\hat{V}^{(5)} = 75.4685$, which is significantly larger than $\hat{V}_{IBET} = 71.1111$, and the value in the sixth loop, $\hat{V}^{(6)} = 75.9797$ is not significantly larger than the value in the fifth loop. Therefore, we stop the loop and settle on the fifth assignment as the final decision rule. This final decision rule is the partition that produces the best value function estimate when the algorithm is applied.

A.2.4 Mathematical Expression

A.2.4.1 Set up

- Let the outcome y_i and the covariate $\boldsymbol{x}_i, \boldsymbol{x}_i \in \mathbb{R}^p$ for the ith patient. $i = 1, \cdots, n$.

- Let the number of the important variable in the list J. For $j = 1, \cdots, J$, x_{ij} implies the value of the jth important variable for ith patient.

- For each j, let $x_{j,1}, \cdots, x_{j,K^*}$ the distinct values of $\boldsymbol{x}_j$. *i.e.* $x_{j,1^*} < \cdots < x_{j,k^*} < \cdots < x_{j,K^*}$.

- Let $X_{j,k^*,1} = \{\boldsymbol{x}_i : x_{ij} < x_{j,k^*}\}$ and $X_{j,k^*,2} = \{\boldsymbol{x}_i : x_{ij} > x_{j,k^*}\}$. $l = 1, 2$ indicates the direction of the values of covariates. Let n_l the number of patients in $X_{j,k^*,l}$

- Let $\mathbf{y}_{j,k^*,1} = \{y_i : x_{ij} < x_{j,k^*}\}$ and $\mathbf{y}_{j,k^*,2} = \{y_i : x_{ij} > x_{j,k^*}\}$. $\mathbf{y}_{j,k^*,1} \cup \mathbf{y}_{j,k^*,2} = \mathbf{y}$

- Let $A = \{1, 2, 3\}$ a set of possible treatments.

A.2.4.2 Algorithm

1st Loop

1. Let a unique value of a variable, $x_{j,k^*} = x_{j,1}, \cdots, x_{j,K_j^*}$, be a cut point. Repeat a-d for all unique values for every variable $j = 1, \cdots, J$.

(a) Using $x^*_{j,k}$ as a cutpoint, split the data set into two subgroups, $X_{j,k,1}$ and $X \setminus X_{j,k,1}$, where the first subgroup includes patients whose value of the variable is less than a cut point and the second subgroup include the patients not in the first subgroup, whose values of the corresponding variable is greater than a cut point of that same variable.

  * Exclude any partition that generates at least one subgroup that has less than 5% of total patients.

(b) For each patient $i = 1, \cdots, n_{j,k,1}$ in the first subgroup, fit a random forest by jackknife approach, i.e. $y = \hat{f}^{(-1)}(x, a)$ where $f(\cdot, \cdot)$ is a random forest and $y \in \mathbf{y}_{j,k,1} \setminus \{y_i\}$, $x \in \boldsymbol{X}_{j,k,1} \setminus \{\boldsymbol{x}_i\}$. Get $\hat{d}(\boldsymbol{x}_i) = \arg\max_{a \in \{1,2,3\}} \hat{f}^{(-i)}(\boldsymbol{x}_i, a)$ for $\boldsymbol{x} \in \boldsymbol{X}_{j,k,1}$, which is the optimal treatment for each patient i.

(c) Assign PT to all patients not in the first subgroup $\boldsymbol{X}_{j,k,1}$ by letting $\hat{d}(\boldsymbol{x}_i) = \text{PT}$ for $\boldsymbol{x}_i \notin \boldsymbol{X}_{j,k,1}$.

(d) Obtain the value function estimate of (c), $\hat{V}^{(1),jk}_{j,k,1,PT}(\hat{d}_n) \equiv \hat{V}^{(1),jk}_{j,k,1,PT}(\hat{d}(\boldsymbol{x}))$. For $i \in I_{j,k,1}$, $\hat{d}(\boldsymbol{x}) = \hat{d}(\boldsymbol{x}_i)$ as estimated in (b), and $i \notin I_{j,k,1}$, $\hat{d}(\boldsymbol{x}) = \hat{d}(\boldsymbol{x}_i) = PT$ as in (c).

(e) Instead of PT in c, assign IBET (in the first subgroup by letting $\hat{d}(\boldsymbol{x}_i) = IBET$) and Waitlist ($\hat{d}(\boldsymbol{x}_i) = WT$) to every patients not in the first subgroup, $\boldsymbol{x}_i \notin \boldsymbol{X}_{j,k,1}$, and get two more value function estimates, $\hat{V}^{(1),jk}_{j,k,1,IBET}(\hat{d}_n)$ and $\hat{V}^{(1),jk}_{j,k,1,WT}(\hat{d}_n)$, respectively. (1) in superscript indicates that the value estimates are calculated in the first iteration.

(f) This time, obtain the estimated treatments $\hat{d}(\boldsymbol{x}_i) = \arg\max_{a \in \{1,2,3\}} \hat{f}^{(-i)}(\boldsymbol{x}_i, a)$ for the patients in second subgroup $\boldsymbol{x} \in \boldsymbol{X}_{j,k,2}$. Apply PT, IBET or Waitlist to the patients not in the second subgroup ($\boldsymbol{x} \notin \boldsymbol{X}_{j,k,2}$). Get three value function estimates, $\hat{V}^{(1),jk}_{j,k,2,PT}(\hat{d}_n)$, $\hat{V}^{(1),jk}_{j,k,2,IBET}(\hat{d}_n)$, $\hat{V}^{(1),jk}_{j,k,2,WT}(\hat{d}_n)$, respectively.

2. Obtain 6 value function estimates, $\hat{V}^{(1),jk}_{j,k,1,PT}(\hat{d}_n), \hat{V}^{(1),jk}_{j,k,1,IBET}(\hat{d}_n), \hat{V}^{(1),jk}_{j,k,1,WT}(\hat{d}_n), \hat{V}^{(1),jk}_{j,k,2,PT}$ $(\hat{d}_n), \hat{V}^{(1),jk}_{j,k,2,IBET}(\hat{d}_n), \hat{V}^{(1),jk}_{j,k,2,WT}(\hat{d}_n)$ for all $j = 1, \cdots, J$ and $k = 1, \cdots, K_j$. Out of all $6(K_1 + \cdots + K_j)$ value function estimates, obtain the largest estimate $\hat{V}^{(1),jk} = \max_{j,k,l,a} \hat{V}^{(1),jk}_{j,k,l,a}$. Also, obtain $(\hat{j}^{(1)}, \hat{k}^{(1)}, \hat{a}^{(1)}, \hat{l}^{(1)}) = \arg\max_{j,k,l,a} \hat{V}^{(1),jk}_{j,k,l,a}$, the treatment and cut point value of variables that yield the largest value function estimate, and its corresponding partition $\hat{I}^{(1)} = I_{\hat{j}^{(1)},\hat{k}^{(1)},\hat{a}^{(1)},\hat{l}^{(1)}}$, $\hat{X}^{(1)} = X_{\hat{j}^{(1)},\hat{k}^{(1)},\hat{l}^{(1)}}$.

3. Assign a single treatment a that is not equal to $\hat{a}^{(1)}$ to every patient who is in the subgroup with estimated treatment, *i.e.,* $\hat{d}^{(1)}\boldsymbol{x}_i = a, a \neq \hat{a}^{(1)}$ for $\forall i \in \hat{I}^{(1)}$. Let $\hat{d}^{(1)}(\boldsymbol{x}) = \hat{a}^{(1)}$ for $i \notin \hat{I}^{(1)}$.

4. Obtain the value function estimate $\hat{V}^{(1),a}(\hat{d}^{(1)}(\boldsymbol{x}))$ using $\hat{d}^{(1)}$ in 3 and obtain $\hat{V}^{(1)} = \max_{a \neq \hat{a}^{(1)}} \hat{V}^{(1),a}$. Test the significance of $\hat{V}^{(1)}$ compared to $\hat{V}^{ZOM}$.

5. If significant, stop the loop and finalize the rule as $\hat{d}^{*} = \hat{d}^{(1)}(\boldsymbol{x}_i)$. If not significant, move on to 6.

Step 6 applies to all mth steps in the loop. $(m \geq 2)$

6. As 1, repeat a,b for all unique values $x_{j,k}$ from all variables $j = 1, \cdots, J$. If the number of the variables used up to the end of the last loop reached 3($|\{\hat{j}^{(1)}, \cdots, \hat{j}^{(m-1)}\}| = 3$), restrict the number of candidate variables to those 3 variables from the current loop, *i.e.* $j \in \{\hat{j}^{(1)}, \cdots \hat{j}^{(m-1)}\}$ from *step m.*

   (a) For every variable j among the candidate variables, define $x^{*,(m-1)}_{j,k}$, which are distinct values of those variables remained from the last loop, cut points. Split the subgroup which is the subgroup of patients received the estimated treatments in the last loop into ($\hat{I}^{(m-1)}$)two groups by using cut points.

      * Exclude any partition that generates at least one subgroup that has less than 5% of total patients.

(b) Each partition generated by 6-(a) comprises of $(m+1)$ distinct subgroups. Assign same treatments given in the last loop for the subgroups that are not split by 6-(a). For example, assign $\hat{d}^{(1)}(\boldsymbol{x}) = \hat{a}^{(1)}$ for $\forall i \in (\hat{I}^{(1)})^c$, assign $\hat{d}^{(2)}(\boldsymbol{x}) = \hat{a}^{(2)}$ for $\forall i \in \hat{I}^{(1)} \cap (\hat{I}^{(2)})^c$,$cdots$, assign $\hat{d}^{(m-1)}(\boldsymbol{x}) = \hat{a}^{(m-1)}$ for $\forall i \in \hat{I}^{(m-2)} \cap (\hat{I}^{(m-1)})^c$.

(c) Compute a value function estimate according to the treatments given to all patients by 5-(a) and 5-(b).Obtain $\hat{V}^{(m),jk}$ and corresponding $(\hat{j}^{(m)}, \hat{k}^{(m)}, \hat{a}^{(m)}, \hat{l}^{(m)})$ and $\hat{X}^{(m)}$.

(d) Obtain all value function estimates as 1-(b), (c), and (d).

7. Out of all value function estimates in 6, obtain the largest estimate $\hat{V}^{(m)}$ and its partition.

8. If the value function estimate in *Step 7*, $\hat{V}^{(m),jk}$, is greater than the value function estimate of *Step 7* in the previous $(m-1th)$ loop, $\hat{V}^{(m-1),jk}$, move on to *Step 9*. If not, move on to *Step 9*. If not, move on to *8*-(a).

   (a) In the previous loop$(m-1th)$, replace the value function estimate from *Step 7* with the next largest value function estimate.

   (b) Go back to *Step 6* and redo the process for the current loop(mth) with the revised partition from the previous loop.

   (c) If the value function estimate in *Step 8-(b)* is not still larger than that next largest value function estimate *Step 8-(a)*, choose the next smaller value function estimate in the previous loop (*m-1 th*). Repeat the process until the value function estimate in *Step 8-(b)* is larger than the value function estimate that is calculated in the previous loop.

9. Assign a single treatment to every patient in the subgroup with estimated treatments. A treatment should be different from the treatment that has been already given to the other subgroup generated in *Step 6*. *i.e.*, $\hat{d}^{(m)}(\boldsymbol{x}_i) = a,\ a \neq \hat{a}^{(m)}$ for $\forall i \in \hat{I}^{(m)}$.

10. Obtain the value function estimate of *Step 9* and check if the difference between $\hat{V}^{(m)}$ and $\hat{V}^{ZOM}$ is significantly different.

11. If the difference in *Step 10* is significant, finalized the decision as *Step 9. i.e.*, $\hat{d}^* = \hat{d}^{(m)}(\boldsymbol{x}_i)$. If it is not significant,

    (a) If the subgroup has the number of patients more than or equal to 10% of total patients, move on to the next loop and repeat from *Step 6*.

    (b) If the subgroup has the number of patients less than 10% of total patients, replace the current partition to the partition that gives the next largest value function estimate in the previous loop($m - 1th$) and redo the current loop from *Step 6* as *Step 8(a)* and *(b)*.

A.2.5 Simulation Studies

In this simulation, we found that the estimated treatment rule by Random Forest (RF) informed Tree-based Learning resulted in greater average outcomes among individuals than the average outcomes by Zero-order models (ZOM). For the simulated data, we set the outcome to $Y = 4X_1 + 2AX_2 + \epsilon$, where the covariates are $X_1 \sim N(3,1), X_2 = 2B - 1$ such that $B \sim Bernoulli(0.5)$. Also, the treatments $A = 1$ or -1 are generated with the probability of 0.5, and the error term is generated from $N(0,1)$. 200 individuals were generated for a single data set. We repeated calculating the results 500 times. In each replication, we generated a simulated data set and estimated the treatment rule based on X_1 using RF informed Tree-based Learning. We calculated a value function estimate using equation (2), which indicates the average outcome that individuals would have obtained if they had followed the given treatment rule. Table A.14 provides averages of 500 value function estimates yielded by a treatment rule created using each of the following three methods: RF informed Tree-based Learning, ZOM with treatments equal to 1, and ZOM with treatments equal to -1. This simulation result implies that our method, RF informed Tree-based Learning is a suitable approach for determining the cut point of covariates, which are the elements of interpretable and advantageous treatment rules.

Table A.14: Average of 500 value estimates for each method

RF informed Tree-based Learning	**ZOM (1)**	**ZOM (-1)**
13.792	12.935	12.877

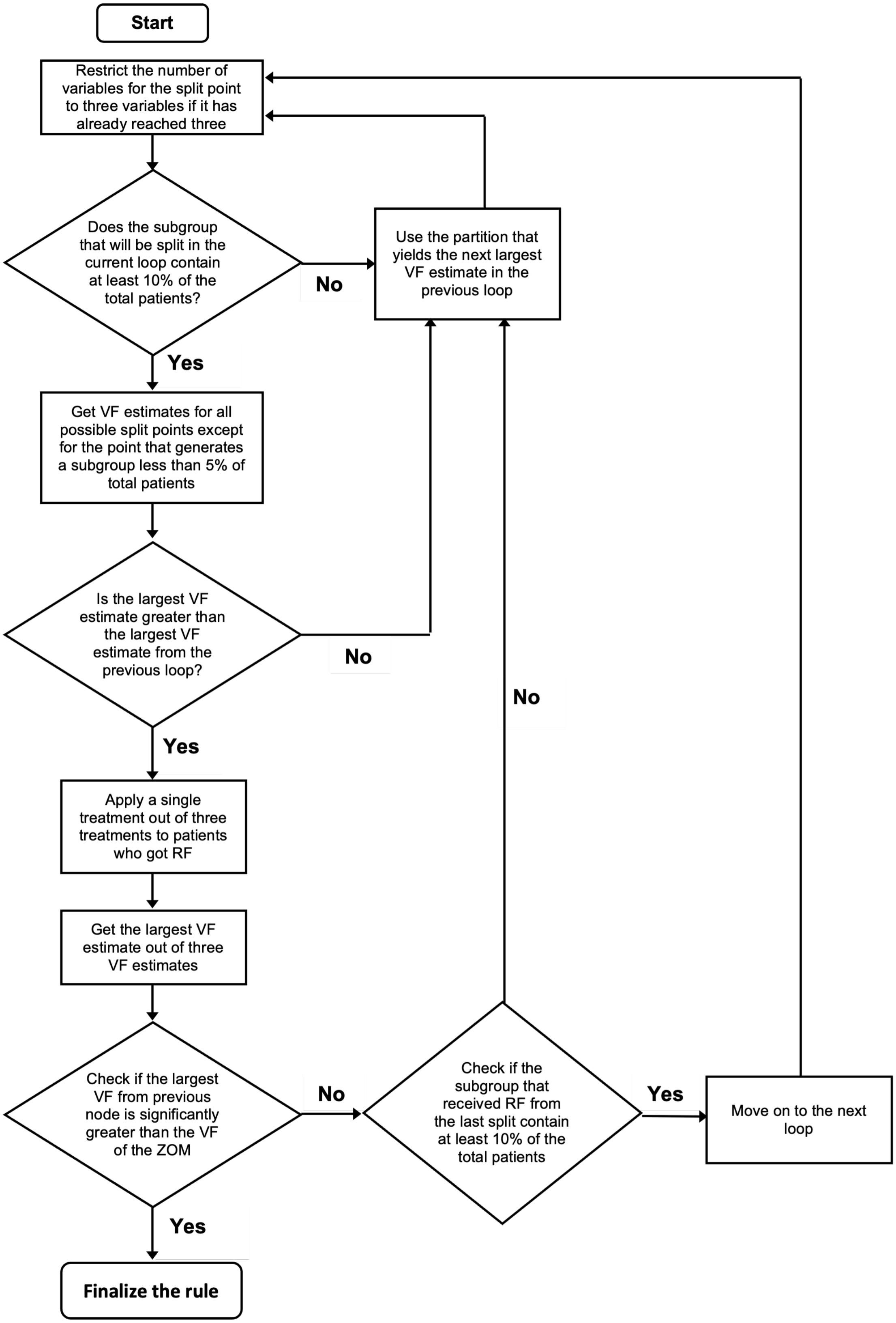
Start
Restrict the number of variables for the split point to three variables if it has already reached three
Does the subgroup that will be split in the current loop contain at least 10% of the total patients?
No
Use the partition that yields the next largest VF estimate in the previous loop
Yes
Get VF estimates for all possible split points except for the point that generates a subgroup less than 5% of total patients
Is the largest VF estimate greater than the largest VF estimate from the previous loop?
No
No
Yes
Apply a single treatment out of three treatments to patients who got RF
Get the largest VF estimate out of three VF estimates
Check if the largest VF from previous node is significantly greater than the VF of the ZOM
No
Check if the subgroup that received RF from the last split contain at least 10% of the total patients
Yes
Move on to the next loop
Yes
Finalize the rule

Figure A.4: Flowchart of the algorithm

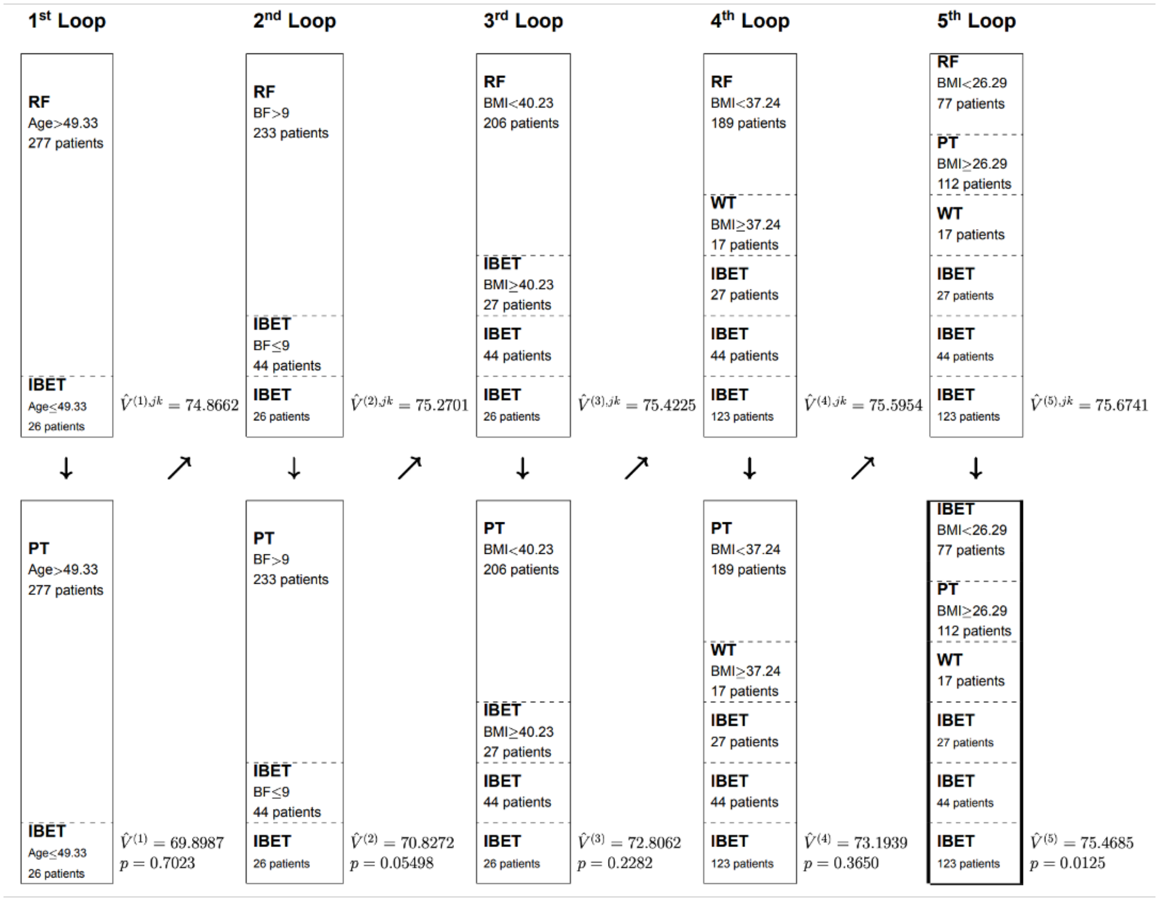

Figure A.5: Diagram for the analysis with outcome at 12-month visit

APPENDIX B: TECHNICAL DETAILS FOR CHAPTER 2

This chapter contains technical details including as assumptions, proofs, definitions, and other materials supplemental to the main text of Chapters 1.4.

B.1 Proofs

Proof of Lemma 2.1. For each $a \in \{1, -1\}$, let $W^*_{a,\theta}(U_\theta, \boldsymbol{X}, A) = \frac{U_\theta I(A=a)}{\pi(a;\boldsymbol{X})} - \frac{I(A=a)-\pi(a;\boldsymbol{X})}{\pi(a;\boldsymbol{X})} Q_\theta(\boldsymbol{X}, a) = I(A = a)\frac{U_\theta - Q_\theta(\boldsymbol{X},a)}{\pi(a;\boldsymbol{X})} + Q_\theta(\boldsymbol{X}, a)$. For fixed θ, it is straightforward to derive the Fisher consistency from the proof of Proposition 3.1 from Zhao et al. (2019), by replacing $Q(\boldsymbol{X}, a)$ to $Q_\theta(\boldsymbol{X}, a)$, and W^*_a to $W^*_{a,\theta}$, respectively. ■

Proof of Theorem 2.1. (a) Since $\mathcal{F}$ is VC-subgraph, by Lemma 9.9 in Kosorok (2008), $\{A\hat{f}_{n,\theta} : \hat{f}_{n,\theta} \in \mathcal{F}\}$ is VC, hence

$$1(A = \hat{d}_{n,\theta}(\boldsymbol{x})) = 1(A \cdot \hat{f}_{n,\theta}(\boldsymbol{x}) > 0)$$

is contained in VC class, and is thus a Glivenko-Cantelli class. Therefore,

$$\sup_{(\theta,\beta)\in\mathbb{R}^p\times\mathcal{B}} \left| (\mathbb{E}_n - \mathbb{E}) \left[\boldsymbol{X}^T\beta 1\{A = \hat{d}_{n,\theta}(\boldsymbol{X})\} - \log\{1 + \exp(\boldsymbol{X}^T\beta)\} \right] \right| \xrightarrow{P} 0 \qquad \text{(B.15)}$$

similar to Luckett et al. (2021). Let $(\tilde{\theta}_n, \tilde{\beta}_n) = \arg\max_{(\theta,\beta)\in\Theta\times\mathcal{B}} \mathbb{E}\big[\boldsymbol{X}^T\beta 1\{A = \hat{d}_{n,\theta}(\boldsymbol{X})\} - \log\{1 + \exp(\boldsymbol{X}^T\beta)\big]$. Then, by Theorem 2.12 in Kosorok (2008), $\hat{\theta}_n = \tilde{\theta}_n + o_P(1)$ and $\hat{\beta}_n = \tilde{\beta}_n + o_P(1)$.

Let

$$\begin{aligned}
W^m_{a,\theta}(U) &= \frac{U_\theta I(A = a)}{\pi^m(a; \boldsymbol{X})} - \frac{I(A = a) - \pi^m(a; \boldsymbol{X})}{\pi^m(a; \boldsymbol{X})} Q^m_\theta(\boldsymbol{X}, a) \\
\mathcal{R}^m_{\theta,\phi}(f) &= E\Big[|W^m_{1,\theta}(U)|\phi\big\{\mathrm{sgn}(W^m_{1,\theta}(U))f\big\} + |W^m_{-1,\theta}(U)|\phi\big\{-\mathrm{sgn}(W^m_{-1,\theta}(U))f\big\}\Big] \\
f^{m,*}_\theta(\boldsymbol{X}) &= \arg\min_{f\in\mathcal{F}} \mathcal{R}^m_{\theta,\phi}(f).
\end{aligned}$$

Then, $d^*_{\theta_0}(\boldsymbol{x}) = \mathrm{sgn}(f^m_{0,\theta}(\boldsymbol{x}))$ according to the Fisher consistency and $V_\theta(d)$ $= \mathbb{E}\Big[Q^m_\theta(\boldsymbol{X}, d(\boldsymbol{X})) + \frac{I\{A=d(\boldsymbol{X})\}}{\pi^m\{d(\boldsymbol{X});\boldsymbol{X}\}}\big(U_\theta - Q^m_\theta(\boldsymbol{X}; d(\boldsymbol{X}))\big)\Big]$ (from the first part of the proof in (c)). Also,

$$
\begin{aligned}
&\mathbb{E}\big[1\{A = \hat{d}_{n,\theta}(\boldsymbol{X})\}|\boldsymbol{X} = \boldsymbol{x}\big] \\
=&\mathbb{E}\big[1\{A = d^*_{\theta_0}(\boldsymbol{X})\}1\{\hat{d}_{n,\theta}(\boldsymbol{X}) = d^*_{\theta_0}(\boldsymbol{X})\} + \big(1 - 1\{A = d^*_{\theta_0}(\boldsymbol{X})\}\big) \\
&\qquad\qquad \cdot \big(1 - 1\{\hat{d}_{n,\theta}(\boldsymbol{X}) = d^*_{\theta_0}(X\boldsymbol{b})\}\big)|\boldsymbol{X} = \boldsymbol{x}\big] \\
=&\mathbb{E}\big[\big(2 \cdot 1\{A = d^*_{\theta_0}(\boldsymbol{X})\} - 1\big)1\{\hat{d}_{n,\theta}(\boldsymbol{X}) = d^*_{\theta_0}(\boldsymbol{X})\}\big] + c
\end{aligned}
$$

for some constant c. For some $\delta > 0$,

$$
\begin{aligned}
&P\big(\hat{d}_{n,\theta}(\boldsymbol{X}) = d^*_{\theta_0}(\boldsymbol{X})\big) \\
\le& P\big(d^*_\theta(\boldsymbol{X}) = d^*_{\theta_0}(\boldsymbol{x}), |f^*_{\theta_0}(\boldsymbol{X})| > \delta, |f^*_\theta(\boldsymbol{X})| > \delta\big) + P\big(|f^*_{\theta_0}(\boldsymbol{x})| \le \delta \text{ or } |f^*_\theta(\boldsymbol{x})| \le \delta\big) \\
&(P\{\hat{d}_{n,\theta}(\boldsymbol{X}) = d^*_\theta(\boldsymbol{X})|\boldsymbol{X}\} \text{ by the proof in (b)}) \\
\xrightarrow{P}& P\big(d^*_\theta(\boldsymbol{X}) = d^*_{\theta_0}(\boldsymbol{X}), |f^*_{\theta_0}(\boldsymbol{X})| > \delta, |f^*_\theta(\boldsymbol{X})| > \delta\big) + 0
\end{aligned}
$$

is continuous in θ and δ. Since δ is arbitrary, $\mathbb{E}\big[1\{A = \hat{d}_{n,\theta}(\boldsymbol{X})\}|\boldsymbol{X} = \boldsymbol{x}\big]$ is continuous in θ. Thus,

$$
\begin{aligned}
&\mathbb{E}\Big[\boldsymbol{X}^T\beta\mathbb{E}\big[1\{A = \hat{d}_{n,\theta}(\boldsymbol{X})\}|\boldsymbol{X}\big] - \log\{1 + \exp(\boldsymbol{X}^T\beta)\}\Big] && \text{(B.16)} \\
&\quad \xrightarrow{P}\mathbb{E}\Big[\boldsymbol{X}^T\beta\mathbb{E}\big[1\{A = d^*_\theta(\boldsymbol{X})\}|\boldsymbol{X}\big] - \log\{1 + \exp(\boldsymbol{X}^T\beta)\}\Big] && \text{(B.17)}
\end{aligned}
$$

uniformly in $\theta \in \Theta$ where Θ is compact, because $\sup_{\theta\in\Theta}\mathbb{E}\big[\big\|\hat{d}_{n,\theta}(\boldsymbol{X}) - d^*_\theta(\boldsymbol{X})\big\|\big] = o_P(1)$ from Theorem 2.1-(b). This leads to $\tilde{\theta}_n \xrightarrow{P} \theta_0$ and $\tilde{\beta}_n \xrightarrow{P} \beta$ by Theorem 2.12 and Lemma 14.3 of Kosorok (2008). Therefore, $\hat{\theta}_n \xrightarrow{P} \theta_0$ and $\hat{\beta}_n \xrightarrow{P} \beta_0$.

(b) Based on the influence function arguments from the beginning of the proof for the weak convergence of $\hat{f}_{n,\theta}$ in Theorem 2.4, we can conclude that proving $\mathbb{E}[\|\hat{f}_{n,\theta}(\boldsymbol{X}) - f^*_\theta(\boldsymbol{X})\|] \xrightarrow{P} 0$

is equivalent to proving $\mathbb{E}[\|\tilde{f}^{\lambda_n^\theta}_{n,\phi,\theta}(\boldsymbol{X}) - f^*_\theta(\boldsymbol{X})\|] \xrightarrow{P} 0$. We use the following lemma for the proof. Since the proof of the following Lemma is analogous to Theorem 2.12 from Kosorok (2008), we omit the proof of the lemma.

Lemma B.2. *Let (T, d_1) and $(\mathcal{H}, d_2)$ be metric spaces. Also, let $(t, h) \mapsto M_{n,t}(h)$ and $(t, h) \mapsto M_t(h)$ be stochastic processes in $l^\infty(T \times \mathcal{H})$. Assume for some $\{h^*_t : t \in T\}$, we have $\liminf_{n\to\infty} \inf_{t\in T} \big(- M_t(h_{n,t}) + M_t(h^*_t)\big) \geq 0$ implies $\sup_{t\in T} d(h_{n,t}, h^*_t) \to 0$ for any sequence $\{h_{n,t} : t \in T\}$. Then for a sequence of estimators $\{\hat{h}_{n,t} : t \in T\}$,*

(i) If $\sup_{t\in T} \big(- M_{n,t}(\hat{h}_{n,t}) - \sup_{h\in\mathcal{H}}(-M_{n,t}(h)\big) = o_P(1)$ and $\sup_{t\in T, h\in\mathcal{H}} |M_{n,t}(h) - M_t(h)| = o_P(1)$, then $\sup_{t\in T} d(\hat{h}_{n,t}, h^*_t) \xrightarrow{P} 0$.

(ii) If $\sup_{t\in T} \big(-M_{n,t}(\hat{h}_{n,t}) - \sup_{h\in\mathcal{H}}(-M_{n,t}(h)\big) = o_{as*}(1)$ and $\sup_{t\in T, h\in\mathcal{H}} |M_{n,t}(h) - M_t(h)| = o_{as*}(1)$, then $\sup_{t\in T} d(\hat{h}_{n,t}, h^*_t) \xrightarrow{as*} 0$.

Let $M_\theta(f) = \mathbb{E}\Big[|W_{1,\theta}(U)|\phi\big\{\text{sgn}(W_{1,\theta}(U))f\big\} + |W_{-1,\theta}(U)|\phi\big\{ - \text{sgn}(W_{-1,\theta}(U))f\big\}\Big]$ and $M_{n,\theta}(f) = \mathbb{E}_n\Big[|\hat{W}_{1,\theta}(U)|\phi\big\{\text{sgn}(\hat{W}_{1,\theta}(U))f\big\} + |\hat{W}_{-1,\theta}(U)|\phi\big\{ - \text{sgn}(\hat{W}_{-1,\theta}(U))f\big\} + \lambda^\theta_n\|f\|^2\Big]$. Then according to the lemma, $\sup_{\theta\in\Theta} \mathbb{E}[\|\tilde{f}^{\lambda_n^\theta}_{n,\phi,\theta}(\boldsymbol{X}) - f^*_\theta(\boldsymbol{X})\|] \leq \sup_{\theta\in\Theta} \|\tilde{\gamma}^{\lambda_n^\theta}_{n,\phi,\theta} - \gamma^*_\theta\| \cdot \mathbb{E}[\|\xi(\boldsymbol{X})\|] \xrightarrow{P} 0$, and further $\sup_{\theta\in\Theta} \mathbb{E}[\|\hat{f}_{n,\theta}(\boldsymbol{X}) - f^*_\theta(\boldsymbol{X})\|] \xrightarrow{P} 0$. Also, since $\mathbb{E}[|\hat{d}_{n,\theta}(\boldsymbol{X}) - d^*_\theta(\boldsymbol{X})|] = \mathbb{E}[1\{\hat{\gamma}^T_{n,\theta}\xi(\boldsymbol{X}) \leq 0 < \gamma^{*T}_\theta\xi(\boldsymbol{X})\} + 1\{\gamma^{*T}_\theta\xi(\boldsymbol{X}) \leq 0 < \hat{\gamma}^T_{n,\theta}\xi(\boldsymbol{X})\}] = \mathbb{E}[1\{(\hat{\gamma}_{n,\theta} - \gamma^*_\theta)^T\xi(\boldsymbol{X}) \leq -\gamma^{*T}_\theta\xi(\boldsymbol{X})\} + 1\{0 < -\gamma^{*T}_\theta\xi(\boldsymbol{X}) \leq -(\hat{\gamma}_{n,\theta} - \gamma^*_\theta)^T\xi(\boldsymbol{X})\}]$, we can conclude $\sup_{\theta\in\Theta} \mathbb{E}[|\hat{d}_{n,\theta}(\boldsymbol{X}) - d^*_\theta(\boldsymbol{X})|] \xrightarrow{P} 0$ due to $\sup_{\theta\in\Theta} \mathbb{E}[\|\hat{f}_{n,\theta}(\boldsymbol{X}) - f^*_\theta(\boldsymbol{X})\|] \xrightarrow{P} 0$.

(c)

$$
\begin{aligned}
&\left|\hat{V}_{\hat{\theta}_n}\left(\hat{d}_{n,\hat{\theta}_n}\right) - V_{\theta_0}\left(d^*_{\theta_0}\right)\right| \\
\leq &\left|\hat{V}_{\hat{\theta}_n}\left(\hat{d}_{n,\hat{\theta}_n}\right) - V_{\hat{\theta}_n}\left(\hat{d}_{n,\hat{\theta}_n}\right)\right| + \left|V_{\hat{\theta}_n}\left(\hat{d}_{n,\hat{\theta}_n}\right) - V_{\theta_0}\left(\hat{d}_{n,\hat{\theta}_n}\right)\right| + \left|V_{\theta_0}\left(\hat{d}_{n,\hat{\theta}_n}\right) - V_{\theta_0}\left(d^*_{\theta_0}\right)\right|. \quad \text{(B.18)}
\end{aligned}
$$

It is straightforward to show that the first term of (B.18) converges to 0 in probability using the proof of Lemma 2.1 in Zhao et al. (2019). Accordingly,

$$
\begin{aligned}
&\hat{V}_{\theta}^{AIPWE}(d) \\
=&\mathbb{E}_n\Big[\frac{U_\theta I\{A=d(\boldsymbol{X})\}}{\hat{\pi}\{d(\boldsymbol{X});\boldsymbol{X}\}} - \frac{I\{A=d(\boldsymbol{X})\}-\hat{\pi}\{d(\boldsymbol{X});\boldsymbol{X}\}}{\hat{\pi}\{d(\boldsymbol{X});\boldsymbol{X}\}}\hat{Q}_\theta(\boldsymbol{X},d(\boldsymbol{X}))\Big] \\
=&\mathbb{E}_n\Big[\frac{U_\theta I\{A=d(\boldsymbol{X})\}}{\pi^m\{d(\boldsymbol{X});\boldsymbol{X}\}} - \frac{I\{A=d(\boldsymbol{X})\}-\pi^m\{d(\boldsymbol{X});\boldsymbol{X}\}}{\pi^m\{d(\boldsymbol{X});\boldsymbol{X}\}}Q_\theta^m(\boldsymbol{X},d(\boldsymbol{X}))\Big] + o_P(1)
\end{aligned}
\tag{B.19}
$$

$$
\xrightarrow{P}\mathbb{E}\Big[Q_\theta^m(\boldsymbol{X},d(\boldsymbol{X})) + \frac{I\{A=d(\boldsymbol{X})\}}{\pi^m\{d(\boldsymbol{X});\boldsymbol{X}\}}\big(U_\theta - Q_\theta^m(\boldsymbol{X},d(\boldsymbol{X}))\big)\Big]. \tag{B.20}
$$

Since $\hat{\pi}(a;\boldsymbol{x}) \xrightarrow{P} \pi^m(a;\boldsymbol{x})$ and $\hat{Q}_\theta(\boldsymbol{x},a) \xrightarrow{P} Q_\theta^m(\boldsymbol{x},a)$ uniformly in $\theta \in \Theta$ for all $(\boldsymbol{x},a) \in \mathbb{R}^p \times \{1,-1\}$, (B.19) holds for $\forall\theta \in \Theta$. Assume $\pi^m(a;\boldsymbol{x}) = \pi(a;\boldsymbol{x})$. Then for each $\theta \in \Theta$, (B.20) $= \mathbb{E}\Big[Q_\theta^m(\boldsymbol{X},d(\boldsymbol{X})) + \frac{I\{A=d(\boldsymbol{X})\}}{\pi\{d(\boldsymbol{X});\boldsymbol{X}\}}\big(U_\theta - Q_\theta^m(\boldsymbol{X},d(\boldsymbol{X}))\big)\Big] = \mathbb{E}\Big[Q_\theta^m(\boldsymbol{X},d(\boldsymbol{X})) + \frac{\mathbb{E}[I\{A=d(\boldsymbol{X})\}|\boldsymbol{X}]}{\pi\{d(\boldsymbol{X});\boldsymbol{X}\}}\big(U_\theta - Q_\theta^m(\boldsymbol{X},d(\boldsymbol{X}))\big)\Big] = \mathbb{E}\Big[Q_\theta^m(\boldsymbol{X},d(\boldsymbol{X})) + \big(U_\theta - Q_\theta^m(\boldsymbol{X},d(\boldsymbol{X}))\big)\Big] = \mathbb{E}[\mathbb{E}[U_\theta|\boldsymbol{X}]] = \mathbb{E}[Q_\theta(\boldsymbol{X},d(\boldsymbol{X}))]$.

This time, let's assume $Q_\theta^m(\boldsymbol{x},a) = Q_\theta(\boldsymbol{x},a)$ for each $\theta \in \Theta$. Then, (B.20)$= \mathbb{E}[Q_\theta(\boldsymbol{X},d(\boldsymbol{X}))]$ $+\mathbb{E}\big[\frac{I(A=d(\boldsymbol{X}))}{\pi^m(d(\boldsymbol{X});\boldsymbol{X})}\big(\mathbb{E}[U_\theta|\boldsymbol{X}] - Q_\theta(\boldsymbol{X},d(\boldsymbol{X}))\big)\big] = \mathbb{E}[Q_\theta(\boldsymbol{X},d(\boldsymbol{X}))] + 0$, where $\mathbb{E}[Q_\theta(\boldsymbol{X},d(\boldsymbol{X}))]$ is equivalent to $V_\theta(d)$.

For the second term of (B.18), the Dominated Convergence Theorem could be applied. Since $\hat{\theta} \xrightarrow{P} \theta_0$, $\Big|V_{\hat{\theta}}\big(\hat{d}_{n,\hat{\theta}}\big) - V_{\theta_0}\big(\hat{d}_{n,\hat{\theta}}\big)\Big| \xrightarrow{P} 0$. For the third term of (B.18), we directly apply (b) of Proposition 3.1 from Zhao et al. (2019). For each $\theta \in \Theta$, define $\mathcal{R}_{\theta,\phi}^m(f) = E[|W_{\theta,1}^m|\phi\{\text{sgn}(W_{\theta,1}^m)f(\boldsymbol{X})\}+|W_{\theta,-1}^m|\phi\{\text{sgn}(W_{\theta,-1}^m)f(\boldsymbol{X})\}]$, and $c_{\theta,m}(\boldsymbol{x}) = E\big[\sum_a|\frac{U_\theta I(A=a)}{\pi^m(a;\boldsymbol{X})} - \frac{I(A=a)-\pi^m(a;\boldsymbol{X})}{\pi^m(a;\boldsymbol{X})}Q_\theta^m(\boldsymbol{X},a)|\big]$. Also, denote $\hat{f}_{n,\hat{\theta}} \equiv \hat{f}_{n,\hat{\theta}}(\boldsymbol{x})$, and $f_{\theta_0}^* \equiv f_{\theta_0}^*(\boldsymbol{x})$. Then,

$$
\Big|V_{\theta_0}\big(\hat{f}_{n,\hat{\theta}}\big) - V_{\theta_0}\big(f_{\theta_0}^*\big)\Big| \tag{B.21}
$$

$$
\leq \sup_{x\in\mathbb{R}^p} c_{\theta_0,m}(x)\tilde{\phi}^{-1}\Big\{\frac{|\mathcal{R}_{\theta_0,\phi}^m(\hat{f}_{n,\hat{\theta}}) - \mathcal{R}_{\theta_0,\phi}^m(f_{\theta_0}^*)|}{\inf_{x\in\mathbb{R}^p} c_{\theta_0,m}(\boldsymbol{x})}\Big\}, \tag{B.22}
$$

where $\tilde{\phi}^{-1}(t)$ is an inverse function of $\tilde{\phi}(t) = |t|$ for hinge loss, $\tilde{\phi}(t) = 1 - \sqrt{1-t^2}$ for exponential loss, $\tilde{\phi}(x) = (1+t)\log(1+t)/2 + (1-t)\log(1-t)/2$ for logistic loss, and $\tilde{\phi}(t) = t^2$ for squared hinge loss when $t \geq 0$. Since $\tilde{\phi}(\cdot)$ is continuous, (B.22) converges to 0 in probability by $\hat{f}_{n,\hat{\theta}} \xrightarrow{P} f^*_{\theta_0}$ and continuous mapping theorem. Therefore, (B.21) converges to 0 in probability, and 3 of Theorem 2.1 follows. ■

Proof of Theorem 2.2. Before giving the proof, we need the following Lemma B.3.

Lemma B.3. *Under the conditions of Theorem 2.2, $t \mapsto h_t$ is uniformly equicontinuous over T, $\hat{h}$ is separable, and $\forall$ open $G \subset \tilde{H}$ containing $\hat{h}$, we have that for every compact $K \subset H$,*

$$\inf_{\tilde{h} \in G^c \cap \tilde{K}} \inf_{t \in T} \big(M_t(\hat{h}_t) - M_t(\tilde{h}) \big) > 0. \tag{B.23}$$

Proof of Lemma B.3. Fix $t^* \in T$ and $\eta > 0$. For any $h_1 \in H$, let $B_\eta(h_1) = \{h \in H : d_2(h, h_1) < \eta\}$. Also for any $h_1 \in \tilde{H}$, let $G_\eta = \{h_2 \in \tilde{H} : h_{2,t} \in B_\eta(h_{1,t}), \forall t \in T\}$.

We know by the assumptions that there exists a compact $K \subset H$ such that $\hat{h} \in \tilde{K}$ (i.e., $\hat{h}_t \in K, \forall t \in T$). We also know by the assumed uniqueness of the maximizer $\hat{h}_t$ for each $t \in T$, that

$$M_{t_1}(\hat{h}_{t_1}) - \sup_{h \in B^c_\eta(\hat{h}_{t_1}) \cap K} M_{t_1}(h) \geq \epsilon \tag{B.24}$$

for some $\epsilon > 0$. We also know there exists a $\delta > 0$ such that

$$\sup_{t \in T : d_1(t, t_1) < \delta} \sup_{h \in K} |M_t(h) - M_{t_1}(h)| \leq \epsilon/3. \tag{B.25}$$

Thus, for any $t \in T$ such that $d_1(t, t_1) < \delta$, we have that

$$\begin{aligned} M_t(\hat{h}_t) - \sup_{h \in B^c_\eta(\hat{h}_{t_1}) \cap K} M_t(h) &\geq M_t(\hat{h}_{t_1}) - \sup_{h \in B^c_\eta(\hat{h}_{t_1}) \cap K} M_{t_1}(h) - \epsilon/3 \\ &\geq M_{t1}(\hat{h}_{t_1}) - \sup_{h \in B^c_\eta(\hat{h}_{t_1}) \cap K} M_{t_1}(h) - 2\epsilon/3 \\ &\geq \epsilon/3, \end{aligned}$$

therefore, $d_2(\hat{h}_{t_1}, \hat{h}_t) < \eta$. The first inequality follows from $\hat{h}_t$ being the maximizer of $h \mapsto M(h)$ combined with (B.25). The second inequality follows from a reapplication of (B.25), and the next inequality follows from (B.24). Since η was arbitrary, we have that $\hat{h}_t$ is continuous at $t = t_1$. Since t_1 was arbitrary and T is compact, we have the desired uniform equicontinuity of $t \mapsto \hat{h}_t$. We also conclude that $\hat{h}$ is separable. This also implies that $t \mapsto M_t(\hat{h}_t)$ is uniformly equicontinuous in t.

For the next part, for any open set $G \subset \tilde{H}$ such that $\hat{h} \in G$, there exists an $\eta > 0$ such that $\hat{h} \in G_\eta \subset G \subset H$. Recall $\tilde{K}$ from before for which $\hat{h} \in \tilde{K}$. Then,

$$\begin{aligned} &\inf_{h \in G^c \cap \tilde{K}} \inf_{t \in T} \big(M_t(\hat{h}_t) - M_t(h) \big) \\ &\qquad \geq \inf_{h \in G^c_\eta \cap \tilde{K}} \inf_{t \in T} \big(M_t(\hat{h}_t) - M_t(h) \big) \\ &\qquad = \inf_{t \in T} \big(M_t(\hat{h}_t) - \sup_{h_1 \in B^c_\eta(\tilde{h}_t) \cap K} M_t(h_1) \big) \\ &\qquad > 0, \end{aligned}$$

by uniform continuity in t of M_t, $M_t(\hat{h}_t)$ and $\hat{h}_t$.

If this were not true, the uniform continuity would imply $\exists t \in T$ such that $M_t(\hat{h}_t) - \sup_{h_1 \in B^c_\eta(\hat{h}_t) \cap K} M_t(h_1) \leq 0$, but this would violate the uniqueness of the maximum for all $t \in T$. Thus the desired results follow almost surely. ■

We continue the proof for Theorem 2.2. Let F be a closed subset of $\tilde{H}$. Fix $\epsilon > 0$, and let $K \subset H$ be compact such that $\liminf_{n\to\infty} P_*(\hat{h}_n \in \tilde{K}) \geq 1 - \epsilon$ and $P(\hat{h} \in \tilde{K}) > 1 - \epsilon$. Now,

$$\begin{aligned}
\limsup_{n\to\infty} P^*(\hat{h}_n \in F) &\leq \limsup_{n\to\infty} P^*(\hat{h}_n \in F \cap \tilde{K}) + \epsilon \\
&\leq \limsup_{n\to\infty} P^*\bigg(\sup_{h_1 \in F\cap\tilde{K}} \inf_{h_2 \in F^c\cap\tilde{K}} \inf_{t\in T} \big(M_{n,t}(h_{1,t}) - M_{n,t}(h_{2,t})\big) + o_P(1) \\
&\qquad\qquad \geq 0\bigg) + \epsilon \\
&\leq P\bigg(\sup_{h_1 \in F\cap\tilde{K}} \inf_{h_2 \in F^c\cap\tilde{K}} \inf_{t\in T} \big(M_t(h_{1,t}) - M_t(h_{2,t})\big) \geq 0\bigg) + \epsilon, \qquad \text{(B.26)}
\end{aligned}$$

where P^* is outer probability. If $\hat{h} \in F^c \cap \tilde{K}$, then, for some $\eta > 0$, $G_\eta \ni \hat{h}$ and $G_\eta \cap F = \emptyset$(the null set). Then,

$$\begin{aligned}
&- \sup_{h_1 \in F\cap\tilde{K}} \inf_{h_2 \in F^c\cap\tilde{K}} \inf_{t\in T} \big(M_t(h_{1,t}) - M_t(h_{2,t})\big) \\
&= \inf_{h_1 \in F\cap\tilde{K}} \sup_{h_2 \in F^c\cap\tilde{K}} \inf_{t\in T} \big(M_t(h_{2,t}) - M_t(h_{1,t})\big) \\
&= \inf_{h_1 \in F\cap\tilde{K}} \sup_{t\in T} \big(M_\theta(\hat{h}_t) - M_t(h_{1,t})\big) \\
&\geq \inf_{h_1 \in G_\eta^c\cap\tilde{K}} \sup_{t\in T} \big(M_t(\hat{h}_t) - M_t(h_{1,t})\big) \\
&\geq \inf_{t\in T} \big(M_t(\hat{h}_t) - \sup_{\tilde{h} \in B_\eta^c(\hat{h}_t)\cap\tilde{K}} M_t(\tilde{h})\big) > 0
\end{aligned}$$

by Lemma B.3. But this contradicts the event in (B.26) so that $\hat{h} \in (F^c \cap \tilde{K})^c = F \cup \tilde{K}^c$. Therefore, $\limsup_{n\to\infty} P^*(\hat{h}_n \in F) \leq P(\hat{h} \in F) + 2\epsilon$, and the results follow since ϵ is arbitrary. The other conclusions follow from Lemma B.3. ■

Proof of Theorem 2.3. Assume that $\inf_{t\in T}\big(M_{n,t}(\hat{h}_{n,t})-\sup_{h\in N_t}M_{n,t}(h)\big)\geq -K_1r_n^{-2}$ for some $K_1>0$, and the quadratic condition (2.6) holds for all $t\in T$ and $h\in\mathcal{H}$. Then,

$$\begin{aligned}
&P^*(r_n\sup_{t\in T}d(\hat{h}_{n,t},h_t^*)>2^M)\\
&=\sum_{j\geq M}P^*(2^{j-1}<r_n\sup_{t\in T}d(\hat{h}_{n,t},h_t^*)\leq 2^j).
\end{aligned}$$

Let the "peels", $S_{j,n}=\{h:2^{j-1}<r_n\sup_{t\in T}d(h,h_t^*)\leq 2^j\}$. Then,

$$\begin{aligned}
&P(2^{j-1}<r_n\sup_{t\in T}d(\hat{h}_{n,t},h_t^*)\leq 2^j)\\
&\quad\leq P\big(\sup_{t\in T}\sup_{h\in S_{j,n}}M_{n,t}(h)-M_{n,t}(h_t^*)+K_1r_n^{-2}\geq 0\big).
\end{aligned}$$

Let $M_{n,t}^{'}=M_{n,t}-M_t$. Then,

$$\begin{aligned}
&P\big(\sup_{t\in T,h\in S_{j,n}}\big[M_{n,t}(h)-M_{n,t}(h_t^*)+K_1r_n^{-2}\big]\geq 0\big)\\
&\quad=P\big(\sup_{t\in T,h\in S_{j,n}}\big[M_{n,t}^{'}(h)-M_n^{'}(h_t^*)+M_t(h)-M_t(h_t^*)+K_1r_n^{-2}\big]\geq 0\big)\\
&\quad\leq P\big(\sup_{t\in T,h\in S_{j,n}}\big[M_{n,t}^{'}(h)-M_n^{'}(h_t^*)+K_1r_n^{-2}\big]\geq -\sup_{t\in T,f\in S_{j,n}}\big[M_t(h)-M_t(h_t^*)\big]\big).
\end{aligned}$$

By (2.6), (2.7), and (2.8),

$$\begin{aligned}
&P^*\big(\sup_{t\in T,h\in S_{j,n}}\big[M_{n,t}(h)-M_{n,t}(h_t^*)+K_1r_n^{-2}\big]\geq 0\big)\\
&\leq P^*\big(\sup_{t\in T,h\in S_{j,n}}\big[M_{n,t}^{'}(h)-M_{n,t}^{'}(h_t^*)+K_1r_n^{-2}\big]\geq -\sup_{t\in T,h\in S_{j,n}}\big[M_t(h)-M_t(h_t^*)\big]\big)\\
&\leq P^*\big(\sup_{t\in T,h\in S_{j,n}}\big[M_{n,t}^{'}(h)-M_{n,t}^{'}(h_t^*)+K_1r_n^{-2}\big]\geq \sup_{t\in T}c_1d^2(h,h_t^*)\big)\\
&\leq P^*\big(\sup_{t:t\in T,h\in S_{j,n}}\big[M_{n,t}^{'}(h)-M_{n,t}^{'}(h_t^*)\big]\geq \frac{c_12^{2j-2}-K_1}{r_n^2}\big)\\
&\leq P^*\big(\sup_{t\in T,h:d(h,h_t^*)<\frac{2^j}{r_n}}\big[M_{n,t}^{'}(h)-M_{n,t}^{'}(h_t^*)\big]\geq \frac{c_12^{2j-2}-K_1}{r_n^2}\big)\\
&\leq \mathbb{E}^*\big(\sup_{t\in T,h:d(h,h_t^*)<\frac{2^j}{r_n}}\big[M_{n,t}^{'}(h)-M_{n,t}^{'}(h_t^*)\big]\big)/\frac{c_12^{2j-2}-K_1}{r_n^2}\big)\\
&\leq \frac{c_2\phi_n(2^j/r_n)r_n^2}{\sqrt{n}(c_12^{2j-2}-K_1)}\\
&\leq \frac{c_2c_32^{j\alpha}}{c_12^{2j-2}-K_1}.
\end{aligned}$$

Therefore,

$$P(r_n\sup_{t\in T}d(\hat{h}_{n,t},h_t^*)>2^M)\leq\sum_{j\geq M}\frac{c_2c_32^{j\alpha}}{c_12^{2j-2}-K_1}.$$

Thus there exists a constant M such that $\limsup_{n\to\infty}P(r_nd(\hat{h}_{n,t},h_t^*)>2^M)\leq 2\epsilon$ since the right term goes to 0 as $M\to\infty$. Since ϵ is arbitrary, $r_nd(\hat{h}_{n,t},h_t^*)=O_P(1)$. ■

Proof of Theorem 2.4. We split $\hat{f}_{n,\hat{\theta}_n}(\boldsymbol{X})-f_{\theta_0}^*(\boldsymbol{X})$ as below.

$$\hat{f}_{n,\hat{\theta}_n}(\boldsymbol{X})-f_{\theta_0}^*(\boldsymbol{X}) \tag{B.27}$$

$$=\big(\hat{f}_{n,\hat{\theta}_n}(\boldsymbol{X})-f_{\hat{\theta}_n}^*(\boldsymbol{X})\big)+\big(f_{\hat{\theta}_n}^*(\boldsymbol{X})-f_{\theta_0}^*(\boldsymbol{X})\big). \tag{B.28}$$

$$=\big\{\big(\hat{\gamma}_{n,\hat{\theta}_n}-\gamma_{\hat{\theta}_n}^*\big)+\big(\gamma_{\hat{\theta}_n}^*-\gamma_{\theta_0}^*\big)\big\}^T\xi(\boldsymbol{X}). \tag{B.29}$$

Using Lemma B.4 below, we obtain the rate of convergence of the first term of (B.28) is equal to $\sqrt{n}$. We start with identifying the asymptotic distribution of $\sqrt{n}(\hat{\gamma}_{\hat{\theta}_n} - \gamma^*_{\hat{\theta}_n})$. Recall $U = (X, A, Y, Z)$. Define $m^{\theta}_{\gamma_h}(U) \equiv -\sum_a |W_{a,\theta}(U)|\phi(a \cdot \text{sgn}(W_{a,\theta}(U))\xi(X)^T(\gamma^*_\theta + \frac{\gamma_h}{\sqrt{n}}))$ such that $M_\theta(\gamma_h) = \mathbb{E}[m^{\theta}_{\gamma_h}(U)]$, where $h(\boldsymbol{X}) = \gamma_h^T\xi(\boldsymbol{X}) \in \mathcal{H}$. Similarly, let $\hat{m}^{\theta}_{\gamma_h}(U) \equiv -\sum_a |\hat{W}_{a,\theta}(U)|\phi(a \cdot \text{sgn}(\hat{W}_{a,\theta}(U))\xi(X)^T(\gamma^*_\theta + \frac{\gamma_h}{\sqrt{n}}))$ such that $\hat{M}_\theta(\gamma_h) = \mathbb{E}_n[\hat{m}^{\theta}_{\gamma_h}(U)]$. We also let $\hat{m}^{(j),\theta}_{\gamma^*_\theta}(U) = (\frac{\partial}{\partial\gamma})^j \hat{m}^{\theta}_{\gamma}(U)\big|_{\gamma=\gamma^*_\theta}$ for $j = 1, 2$, and $m^{(j),\theta}_{\gamma^*_\theta}(U)$ be similarly defined for $m^{\theta}_{\gamma^*_\theta}(U)$. We first show that the limiting distribution of $\sqrt{n}(\hat{\gamma}_{n,\theta} - \gamma^*_\theta)$ is equivalent to the limiting distribution of $\sqrt{n}(\tilde{\gamma}^{\lambda_n}_{n,\phi,\theta} - \gamma^*_\theta)$ such that $\tilde{f}^*_{n,\phi,\theta}(\boldsymbol{X}) = \tilde{\gamma}^{\lambda_n}_{n,\phi,\theta}\xi(\boldsymbol{X})$ where $\tilde{f}^{\lambda_n}_{n,\phi,\theta}$ is estimated by (2.3) without sample splitting technique. In order to prove this, we borrow the influence function vectors of $\sqrt{n}(\tilde{\gamma}^{\lambda_n}_{n,\phi,\theta} - \gamma^*_\theta)$, which is derived later in this proof. Then, we have

$$
\begin{aligned}
\sqrt{n}(\tilde{\gamma}^{\lambda_n}_{n,\phi,\theta} - \gamma^*_\theta) &= -V_\theta^{-1}(\mathbb{G}_n m^{(1),\theta}_{\gamma^*_\theta}(U) + N_\theta^T\sqrt{n}(\hat{V} - V_0)) + o_P(1) \\
&= \mathbb{G}_n(-V_\theta^{-1} m^{(1),\theta}_{\gamma^*_\theta}(U)) + (-V_\theta^{-1}N_\theta^T\sqrt{n}(\hat{V} - V_0)) + o_P(1) \\
&= n^{-1/2}\sum_{i=1}^n(-V_\theta^{-1} m^{(1),\theta}_{\gamma^*_\theta}(U_i) + n^{-1/2}\sum_{i=1}^n(-V_\theta^{-1}N_\theta^T\psi_V(U_i)) + o_P(1) \\
&= n^{-1/2}\sum_{i=1}^n I_1^\theta(U_i) + n^{-1/2}\sum_{i=1}^n I_2^{\theta^T} I_3(U_i) + o_P(1),
\end{aligned}
$$

where $V_\theta = \mathbb{E}\Big[\sum_a -|W_{a,\theta}(U)|\ddot{\phi}\big(a \cdot \text{sgn}(W_{a,\theta}(U))\gamma^{*T}_\theta\xi(\boldsymbol{X})\big)\xi(\boldsymbol{X})\xi(\boldsymbol{X})^T\Big]$ and $I^\theta(U_i) = I_1^\theta(U_i) + I_2^{\theta^T} I_3(U_i)$ is a $q \times 1$ independent and identically distributed influence function vector such that $I_1^\theta(U_i) = -V_\theta^{-1} m^{(1),\theta}_{\gamma^*_\theta}(U_i)$, $I_2^\theta = -N_\theta V_\theta^{-1}$, and $I_3(U_i) = \psi_V(U_i)$. Denote $\mathbb{G}_n^{(j)} = \sqrt{n_j}(\mathbb{E}_n^{(j)} - \mathbb{E})$ and $\mathbb{E}_n^{(j)} = \frac{n}{n_j}\mathbb{E}_n 1_{\{i \in I_j\}}$ where I_j denotes the jth fold. Similarly, $\mathbb{G}_n^{(-j)} = \sqrt{n - n_k}(\mathbb{E}_n^{(-j)} - \mathbb{E})$ and $\mathbb{E}_n^{(-j)} = \frac{n}{n-n_j}\mathbb{E}_n 1_{\{i \notin I_j\}}$. Let $\hat{\gamma}^{(j)}_\theta$ be an estimator for γ^*_θ by samples not included in $\{i : i \in I_j\}$ such that $\hat{f}^{(j)}_{n,\theta} = (\hat{\gamma}^{(j)}_{n,\theta})^T\xi$ where $\hat{f}^{(j)}_\theta \equiv \hat{f}^{\lambda_n,(j)}_{n,\phi,\theta}$. Then considering the sample splitting technique,

$$\begin{aligned}\sqrt{n-n_j}(\hat{\gamma}_{n,\theta}^{(j)}-\gamma_\theta^*) &= \mathbb{G}_n^{(-j)} I_1^\theta(U) + (-V_\theta^{-1} N_\theta^T \sqrt{n-n_j}(\hat{V}_n^{(j)} - V_0)) \\ &= \mathbb{G}_n^{(-j)} I_1(U) + I_2^{\theta T} \frac{\sqrt{n-n_j}}{\sqrt{n_j}} \mathbb{G}_n^{(j)} I_3(U)\end{aligned}$$

where $\hat{V}_n^{(k)}$ is estimated using samples in the jth fold. Thus, for $\theta \in \Theta$,

$$\begin{aligned}\sqrt{n}(\hat{\gamma}_{n,\theta}-\gamma_\theta^*) &= \sqrt{n}(\frac{1}{J}\sum_{j=1}^{J}(\hat{\gamma}_\theta^{(j)}-\gamma_\theta^*)) \\ &= \frac{\sqrt{n}}{J}\sum_{j=1}^{J}\frac{1}{\sqrt{n-n_j}}\big[\mathbb{G}_n^{(-j)} I_1^\theta(U) + \frac{\sqrt{n-n_j}}{\sqrt{n_j}} I_2^{\theta T} \mathbb{G}_n^{(j)} I_3(U)\big] \\ &= \frac{\sqrt{n}}{J}\sum_{j=1}^{J}\frac{n}{n-n_j}(\mathbb{E}_n-\mathbb{E})1(i\notin I_j)I_1^\theta(U_i) \qquad \text{(B.30)} \\ &+ \frac{\sqrt{n}}{J}\sum_{j=1}^{J}\frac{n}{n_j}(\mathbb{E}_n-\mathbb{E})1(i\in I_j)I_2^{\theta T} I_3(U_i). \qquad \text{(B.31)}\end{aligned}$$

Since $\{I_1^\theta(U;\theta) : \theta \in \Theta\}$ and $\{I_2^{\theta T} I_3(U) : \theta \in \Theta\}$ are Glivenko-Cantelli class with integrable envelope, $T_{1,n} \equiv \sqrt{n}\max_{1\le j\le J}\left|\frac{(J-1)n}{J(n-n_j)}-1\right| \to 0$, and $T_{2,n} \equiv \sqrt{n}\max_{1\le j\le J}\left|\frac{n}{Jn_j}-1\right| \to 0$ as $n\to\infty$, for (B.30),

$$\begin{aligned}&\left|\frac{\sqrt{n}}{J}\sum_{j=1}^{J}\frac{n}{n-n_j}(\mathbb{E}_n-\mathbb{E})1(i\notin I_j)I_1^\theta(U_i) - \mathbb{G}_n I_1^\theta(U;\theta)\right| \\ \le& T_{1,n}\|\mathbb{E}_n I_1^\theta(U;\theta)\|_\Theta \xrightarrow{P} 0,\end{aligned}$$

uniformly over Θ. Similarly for (B.31),

$$\begin{aligned}&\left|\frac{\sqrt{n}}{J}\sum_{j=1}^{J}\frac{n}{n_j}(\mathbb{E}_n-\mathbb{E})1(i\in I_j)I_2^{\theta T} I_3(U_i) - \mathbb{G}_n I_2^{\theta T} I_3(U)\right| \\ \le& T_{2,n}\|\mathbb{E}_n I_2^{\theta T} I_3(U)\|_\Theta \xrightarrow{P} 0.\end{aligned}$$

Therefore, we can substitute $\sqrt{n}(\hat{\gamma}_{n,\theta} - \gamma^*_\theta)$ with $\sqrt{n}(\tilde{\gamma}^{\lambda_n}_{n,\phi,\theta} - \gamma^*_\theta)$, which is estimated without the sample splitting technique. For the remainder of the proof, we will show the weak convergence of $\sqrt{n}(\tilde{\gamma}^{\lambda_n}_{n,\phi,\theta} - \gamma^*_\theta)$. Fix a compact $\Gamma \subset \mathbb{R}^q$. We have

$$\begin{aligned}
\tilde{M}_{n,\theta}(\gamma_h) &\equiv n\big(\hat{M}_{n,\theta}(\gamma^*_\theta + \frac{\gamma_h}{\sqrt{n}}) - \hat{M}_{n,\theta}(\gamma^*_\theta)\big) \\
&= n\mathbb{E}_n\big[-\sum_a |\hat{W}_{a,\theta}(U)|\big\{\phi\big(a \cdot \text{sgn}(\hat{W}_{a,\theta}(U))\xi(\boldsymbol{X})^T(\gamma^*_\theta + \frac{\gamma_h}{\sqrt{n}})\big) \\
&\quad - \phi\big(a \cdot \text{sgn}(\hat{W}_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma^*_\theta\big)\big\} + \lambda_n(\|(\gamma^*_\theta + \frac{\gamma_h}{\sqrt{n}})^T\xi(\boldsymbol{X})\|^2 - \|\gamma^{*T}_\theta\xi(\boldsymbol{X})\|^2)\big]
\end{aligned}$$

$$= \mathbb{G}_n\big[\sqrt{n}(\hat{m}^\theta_{\gamma^*_\theta+\gamma_h/\sqrt{n}}(U) - \hat{m}^\theta_{\gamma^*_\theta}(U)) - \gamma_h^T\hat{m}^{(1),\theta}_{\gamma^*_\theta}(U)\big] \tag{B.32}$$

$$+ \gamma_h^T\mathbb{G}_n\hat{m}^{(1),\theta}_{\gamma^*_\theta}(U) \tag{B.33}$$

$$+ n\mathbb{E}\big[\hat{m}^\theta_{\gamma^*_\theta+\gamma_h/\sqrt{n}}(U) - \hat{m}^\theta_{\gamma^*_\theta}(U)\big], \tag{B.34}$$

where $\hat{m}^{(1),\theta}_{\gamma^*_\theta}(U) = -\sum_a |\hat{W}_{a,\theta}(U)|\dot{\phi}\{a \cdot \text{sgn}(\hat{W}_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma^*_\theta\}a \cdot \text{sgn}(\hat{W}_{a,\theta}(U))\xi(\boldsymbol{X}) + 2\lambda_n(\gamma^{*T}_\theta\xi(\boldsymbol{X}))\gamma^*_\theta$.

Using more empirical process arguments, we discover the limiting distributions of (B.32), (B.33), and (B.34). Firstly for (B.32), we show that the conditions in Theorem 11.20 in Kosorok (2008) are satisfied. Let γ_1 and γ_2 be some arbitrary elements in Γ. We have

$$\begin{aligned}
&|\hat{m}^\theta_{\gamma_1}(u) - \hat{m}^\theta_{\gamma_2}(u)| \\
=&\big||\hat{W}_{1,\theta}(u)|\phi\{\text{sgn}(\hat{W}_{1,\theta}(u))\gamma_1^T\xi(\boldsymbol{x})\} + |\hat{W}_{-1,\theta}(u)|\phi\{-\text{sgn}(\hat{W}_{-1,\theta}(u))\gamma_1^T\xi(\boldsymbol{x})\} + \lambda_n\|\gamma_1^T\xi(\boldsymbol{x})\|^2 \\
-&|\hat{W}_{1,\theta}(u)|\phi\{\text{sgn}(\hat{W}_{1,\theta}(u))\gamma_2^T\xi(\boldsymbol{x})\} - |\hat{W}_{-1,\theta}(u)|\phi\{-\text{sgn}(\hat{W}_{-1,\theta}(u))\gamma_2^T\xi(\boldsymbol{x})\} - \lambda_n\|\gamma_2^T\xi(\boldsymbol{x})\|^2\big| \\
\leq&\big||\hat{W}_{1,\theta}(u)|\big(\phi\{\text{sgn}(\hat{W}_{1,\theta}(u))\gamma_1^T\xi(\boldsymbol{x})\} - \phi\{\text{sgn}(\hat{W}_{1,\theta}(u))\gamma_2^T\xi(\boldsymbol{x})\}\big)\big| \\
&+ \big||\hat{W}_{-1,\theta}(u)|\big(\phi\{-\text{sgn}(\hat{W}_{-1,\theta}(u))\gamma_1^T\xi(\boldsymbol{x})\} - \phi\{-\text{sgn}(\hat{W}_{-1,\theta}(u))\gamma_2^T\xi(\boldsymbol{x})\}\big)\big| \\
&+ |\lambda_n(\|\gamma_1^T\xi(\boldsymbol{x})\|^2 - \|\gamma_2^T\xi(\boldsymbol{x})\|^2)| \\
\lesssim&\big(|\hat{W}_{1,\theta}(u)| + |\hat{W}_{-1,\theta}(u)| + |(\gamma_1 + \gamma_2)^T\xi(\boldsymbol{x})|\big)\|\xi(\boldsymbol{x})\| \cdot \|\gamma_1 - \gamma_2\|.
\end{aligned}$$

Let $\dot{m}_\theta(\cdot) = c\big(|\hat{W}_{1,\theta}(\cdot)| + |\hat{W}_{-1,\theta}(\cdot)| + |(\gamma_1+\gamma_2)^T\xi(\boldsymbol{x})|\big)$ for some constant c. Also, define $k^n_{\gamma_{h_\theta}}(u) \equiv \sqrt{n}(\hat{m}^\theta_{\gamma^*_\theta+\frac{\gamma_{h_\theta}}{\sqrt{n}}}(u) - \hat{m}^\theta_{\gamma^*_\theta}(u)) - \gamma^T_{h_\theta}\hat{m}^{(1),\theta}_{\gamma^*_\theta}(u)$, where the subscript θ of h emphasizes that it is in the neighborhood of the maximum of $m^\theta_{\gamma^*_\theta}$, which is γ^*_θ. Let $h_{1,\theta}$ and $h_{2,\theta}$ be arbitrary $h_\theta \in \mathcal{H}$. For each $\theta \in \Theta$,

$$
\begin{aligned}
&|\sup_{\theta\in\Theta} k^n_{\gamma_{h_{1,\theta}}}(u) - \sup_{\theta\in\Theta} k^n_{\gamma_{h_{2,\theta}}}(u)| \\
\le& \sup_{\theta\in\Theta}|k^n_{\gamma_{h_{1,\theta}}}(u) - k^n_{\gamma_{h_{2,\theta}}}(u)| \\
=& \sup_{\theta\in\Theta}\big|\sqrt{n}(\hat{m}^\theta_{\gamma^*_\theta+\frac{\gamma_{h_{1,\theta}}}{\sqrt{n}}}(x) - \hat{m}^\theta_{\gamma^*_\theta}(u)) - \gamma^T_{h_{1,\theta}}\hat{m}^{(1),\theta}_{\gamma^*_\theta}(u) - \sqrt{n}(\hat{m}_{\gamma^*_\theta+\frac{\gamma_{h_{2,\theta}}}{\sqrt{n}}}(u) - \hat{m}_{\gamma^*_\theta}(u)) \\
&\qquad + \gamma^T_{h_{2,\theta}}\hat{m}^{(1),\theta}_{\gamma^*_\theta}(u)\big| \\
=& \sup_{\theta\in\Theta}\big|\sqrt{n}\Big(\hat{m}_{\gamma^*_\theta+\frac{\gamma_{h_{1,\theta}}}{\sqrt{n}}}(u) - \hat{m}_{\gamma^*_\theta+\frac{\gamma_{h_{2,\theta}}}{\sqrt{n}}}(u)\Big) - (\gamma_{h_{1,\theta}} - \gamma_{h_{2,\theta}})^T\hat{m}^{(1),\theta}_{\gamma^*_\theta}(u)\big| \\
\le& \sup_{\theta\in\Theta}\big\{\dot{m}_\theta(u)\|\xi(\boldsymbol{x})\| + \|\hat{m}^{(1),\theta}_{\gamma^*_\theta}(u)\|\big\}\|\gamma_{h_{1,\theta}} - \gamma_{h_{2,\theta}}\|.
\end{aligned}
$$

Thus for $\mathcal{F}_n \equiv \{\sup_{\theta\in\Theta} u^n_{\gamma_{h_\theta}} : \gamma_{h_\theta} \in \Gamma\}$,

$$
\begin{aligned}
&\sup_Q N(\epsilon\|F_n\|_{\mathcal{Q},2}, \mathcal{F}_n, L_2(\mathcal{Q})) \\
\le& \sup_Q N_{[]}(\epsilon\|F_n\|_{\mathcal{Q},2}, \mathcal{F}_n, L_2(\mathcal{Q})) \\
\le& N(\frac{\epsilon}{2}, \Gamma, d(\gamma_{h_{1,\theta}}, \gamma_{h_{2,\theta}})) \\
\lesssim& (\frac{1}{\epsilon})^p,
\end{aligned}
$$

where the envelope $F_n \equiv \sup_{\theta\in\Theta}\{|\dot{m}_\theta|\|\xi\|_{\mathcal{X}} + \|\hat{m}^{(1),\theta}_{\gamma_{0,\theta}}\|\}\|\gamma_{h_\theta}\|_\Gamma$ and $E|F^2_n(u)| < \infty$. Theorem 9.18 from Kosorok (2008) is applied to the first inequality, and the second inequality is derived by utilizing Theorem 9.23 from Kosorok (2008) with $\|\cdot\|_{\mathcal{Q},2}$ for any probability measure $\mathcal{Q}$ on $\mathcal{X}$. Therefore, $\limsup_{n\to\infty}\sup_Q \int_0^1 \sqrt{\log N(\epsilon\|F_n\|_{\mathcal{Q},2}, \mathcal{F}_n, L_2(\mathcal{Q}))}d\epsilon < \infty$.

Now let $H(s,t) = \lim_{n\to\infty} \mathbb{E}(\sup_{\theta\in\Theta} k^n_{\gamma_{h_{s,\theta}}} k^n_{\gamma_{h_{t,\theta}}}) - \mathbb{E}(\sup_{\theta\in\Theta} k^n_{\gamma_{h_{s,\theta}}})\mathbb{E}(\sup_{\theta\in\Theta} k^n_{\gamma_{h_{t,\theta}}})$. Then,

$$
\begin{aligned}
&\|\mathbb{E}(\sup_{\theta\in\Theta} k^n_{\gamma_{h_{s,\theta}}}(u) k^n_{\gamma_{h_{t,\theta}}}(u))\|_\Gamma \\
=&\|\mathbb{E}(\sup_{\theta\in\Theta} n(\hat{m}_{\gamma^*_\theta + \frac{\gamma_{h_{s,\theta}}}{\sqrt{n}}}(u) - \hat{m}_{\gamma^*_\theta}(u) - \frac{\gamma^T_{h_{s,\theta}}}{\sqrt{n}} \hat{m}^{(1),\theta}_{\gamma^*_\theta}(u))^2)\|_\Gamma \\
=&n\mathbb{E}\Big(\sup_{\theta\in\Theta}\{\frac{\gamma^T_{h_{s,\theta}}}{\sqrt{n}}\dot{m}_{\gamma_{0,\theta}}(u) + o\big((\frac{\|\gamma_{h_{s,\theta}}\|}{\sqrt{n}})^2\big) - \frac{\gamma^T_{h_{s,\theta}}}{\sqrt{n}}\dot{m}_{\gamma_{0,\theta}}(u)\}\Big)^2 \\
=&o(\sup_{\theta\in\Theta}\frac{\|\gamma_{h_{s,\theta}}\|^4}{n}).
\end{aligned}
$$

The second equality follows from Taylor's theorem for fixed u. Therefore, $H(s,t) = 0$. Moreover, since F_n does not depend on n and $F_n < \infty$, $\limsup_{n\to\infty} \mathbb{E}F^2_n < \infty$ and $\lim_{n\to\infty} \mathbb{E}F^2_n 1\{F_n > \epsilon\sqrt{n}\} = 0$, for each $\epsilon > 0$. Also, we have,

$$
\begin{aligned}
&\sup_{\sup_{\theta\in\Theta}\|\gamma_{h_{1,\theta}},\gamma_{h_{2,\theta}}\|<\delta_n} (\sup_{\theta\in\Theta} k^n_{\gamma_{h_{1,\theta}}} - \sup_{\theta\in\Theta} k^n_{\gamma_{h_{2,\theta}}})^2 \\
=&\sup_{\sup_{\theta\in\Theta}\|\gamma_{h_{1,\theta}},\gamma_{2_{t,\theta}}\|<\delta_n} \big(\sqrt{n}(\hat{m}_{\gamma^*_\theta + \frac{\gamma_{h_{1,\theta}}}{\sqrt{n}}} - \hat{m}_{\gamma^*_\theta + \frac{\gamma_{h_{2,\theta}}}{\sqrt{n}}}) - (\gamma_{h_{1,\theta}} - \gamma_{h_{2,\theta}})^T \hat{m}^{(1),\theta}_{\gamma^*_\theta}\big)^2 \\
\leq&\sup_{\theta\in\Theta}\big(|\dot{m}_\theta|\|\xi\| + \|\dot{m}^{(1),\theta}_{\gamma^*_\theta}\|\big)^2 \sup_{\sup_{\theta\in\Theta}\|\gamma_{h_{1,\theta}},\gamma_{h_{2,\theta}}\|<\delta_n} \|\gamma_{h_{1,\theta}} - \gamma_{h_{2,\theta}}\|^2 \\
\leq&\sup_{\theta\in\Theta}\big(|\dot{m}_\theta|\|\xi\| + \|\hat{m}^{(1),\theta}_{\gamma_0}\|\big)^2 \delta^2_n \\
\to& 0,\ as\ \delta_n \downarrow 0.
\end{aligned}
$$

Therefore, $\sup_{\theta\in\Theta} \mathbb{G}_n(k^n_{\gamma_{h_\theta}}) \rightsquigarrow 0$ in $l^\infty(\Gamma)$. Next, for (B.34), we have,

$$
\begin{aligned}
&n\mathbb{E}\big[\hat{m}^\theta_{\gamma^*_\theta+\gamma_h/\sqrt{n}}(U) - \hat{m}^\theta_{\gamma^*_\theta}(U)\big] \\
&= n\mathbb{E}\big[m_{\gamma^*_\theta+\gamma_h/\sqrt{n}}(U) - m_{\gamma^*_\theta}(U)\big] && \text{(B.35)} \\
&+ n\mathbb{E}\big[(\hat{m}_{\gamma^*_\theta+\gamma_h/\sqrt{n}}(U) - \hat{m}_{\gamma^*_\theta}(U) - m_{\gamma^*_\theta+\gamma_h/\sqrt{n}}(U) + m_{\gamma^*_\theta}\big]. && \text{(B.36)}
\end{aligned}
$$

It is easily shown that (B.35)= $\frac{1}{2}\gamma_h^T V_\theta \gamma_h + o(1)$, where $V_\theta = \mathbb{E}\Big[\sum_a -|W_{a,\theta}(U)|\ddot{\phi}\big(a \cdot \text{sgn}(W_{a,\theta}(U))\gamma_\theta^{*T}\xi(\boldsymbol{X})\big)\xi(\boldsymbol{X})\xi(\boldsymbol{X})^T\Big]$. For (B.36),

$$
\begin{aligned}
(B.36) &= n\mathbb{E}\big[\hat{m}_{\gamma_\theta^*+\gamma_h/\sqrt{n}}(U) - m_{\gamma_\theta^*+\gamma_h/\sqrt{n}}(U) - (\hat{m}_{\gamma_\theta^*}(U) - m_{\gamma^{\theta*}}(U))\big] \\
&= -\sum_a \mathbb{E}\Big[\sqrt{n}\big(|\hat{W}_{a,\theta}(U)| - |W_{a,\theta}(U)|\big)\dot{\phi}\big(a \cdot \text{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma_\theta^*\big)a \\
&\qquad \cdot \text{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma_h\Big] + o_P(1) \\
&= -\sum_a \mathbb{E}\Big[\sqrt{n}\big(|W_{a,\theta}(U) + \sqrt{n}(\hat{W}_{a,\theta}(U) - W_{a,\theta}(U))/\sqrt{n}| - |W_{a,\theta}(U)|\big)\cdot \\
&\qquad \dot{\phi}\big(a \cdot \text{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma_\theta^*\big)a \cdot \text{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma_h\Big] + o_P(1) \\
&= -\sum_a \mathbb{E}\Big[\Big\{1(W_{a,\theta}(U) \pm \frac{K_\epsilon}{\sqrt{n}} \in N_0) \cdot \sqrt{n}\big(|W_{a,\theta}(U) + \sqrt{n}(\hat{W}_{a,\theta}(U) - W_{a,\theta}(U)) \\
&\qquad /\sqrt{n}| - |W_{a,\theta}(U)|\big) \\
&\qquad + 1(W_{a,\theta}(U) \pm \frac{K_\epsilon}{\sqrt{n}} \notin N_0) \cdot \sqrt{n}\big(|W_{a,\theta}(U) + \sqrt{n}(\hat{W}_{a,\theta}(U) - W_{a,\theta}(U)) \\
&\qquad /\sqrt{n}| - |W_{a,\theta}(U)|\big)\Big\} \cdot \dot{\phi}\big(a \cdot \text{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma_\theta^*\big)a \cdot \text{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})^T \\
&\qquad \gamma_h\Big] + o_P(1) \\
&= o_P(1) - \sum_a \mathbb{E}\Big[\sqrt{n}(\hat{W}_{a,\theta}(U) - W_{a,\theta}(U))\dot{\phi}\big(a\xi(\boldsymbol{X})^T\gamma_\theta^*\big)a \cdot \text{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma_h\Big] \\
&= \sqrt{n}(\hat{V}_n - V_0)^T\mathbb{E}\Big[-\sum_a D_a^\theta(U)\dot{\phi}\big(a \cdot \text{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma_\theta^*\big)a\xi(\boldsymbol{X})^T\Big]\gamma_h + o_P(1),
\end{aligned}
$$

where N_0 is a neighborhood of 0 and K_ϵ is a compact interval depending on arbitrary $\epsilon > 0$. Let's denote $N_\theta \equiv \mathbb{E}\Big[-\sum_a D_a^\theta(U)\dot{\phi}\big(a \cdot \text{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})^T\gamma_\theta^*\big)a\xi(\boldsymbol{X})^T\Big]$. Then, $\tilde{M}_{n,\theta}(\gamma_h) = \frac{1}{2}\gamma_h^T V_\theta \gamma_h + \gamma_h^T\big(\mathbb{G}_n m_{\gamma_\theta^*}^{(1),\theta}(U) + N_\theta^T\sqrt{n}(\hat{V}_n - V_0)\big) + \mathbb{E}_n^1(\theta)$, where $\sup_{\theta\in\Theta}\|\mathbb{E}_n^1(\theta)\| = o_P(1)$. Therefore, $\sqrt{n}(\hat{\gamma}_{n,\theta} - \gamma_\theta^*) = -V_\theta^{-1}\big(Z_1(\theta) + N_\theta^T Z_2\big) + o_P(1)$ where Z_1 and Z_2 are defined as the following limiting distributions:

$$
\begin{pmatrix} \mathbb{G}_n m_{\gamma_\theta^*}^{(1),\theta}(U) \\ \sqrt{n}(\hat{V}_n - V_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} Z_1(\theta) \\ Z_2 \end{pmatrix},
$$

uniformly over Θ, where $Z_1(\theta)$ is a tight, mean zero p-dimensional Gaussian process vector with covariance $\Sigma_U(\theta_1, \theta_2) = \mathbb{E}[Z_1(\theta_1)Z_1(\theta_2)]$, $Z_2 \sim N(0, \Sigma_{22})$ of dimensional d_p, and the covariance of $Z_1(\theta)$ and Z_2 is $\mathbb{E}[Z_1(\theta)Z_2] = \mathbb{E}[m^{(1),\theta}_{\gamma^*_\theta}(U)\psi_V(U)]]$. Then, the limiting distribution of $\sqrt{n}(\hat{\gamma}_{n,\theta} - \gamma^*_\theta)$ is distribution of $-V_\theta^{-1}\tilde{Z}(\theta)$, where $\tilde{Z}(\theta)$ is a vector Gaussian process indexed by θ with covariance $\mathbb{E}\big[m^{(1),\theta}_{\gamma^*_\theta}(U)\big(m^{(1),\theta}_{\gamma^*_\theta}(U)\big)^T\big] + N_\theta^T\mathbb{E}\big[\psi_V(U)\psi_V^T(U)\big]N_\theta + \mathbb{E}\big[m^{(1),\theta}_{\gamma^*_\theta}(U)\psi_V^T(U)\big]N_\theta + N_\theta^T\mathbb{E}\big[\psi_V(U)\big(m^{(1),\theta}_{\gamma^*_\theta}(U)\big)^T\big]$. Let $g_\theta = -V_\theta^{-1}(\mathbb{G}_n m^{(1),\theta}_{\gamma^*_\theta} + N_\theta^T\sqrt{n}(\hat{V}_n - V_0))$. Also, let $A_\theta \equiv \mathbb{E}\big[m^{(1),\theta}_{\gamma^*_\theta}(U)\big(m^{(1),\theta}_{\gamma^*_\theta}(U)\big)^T\big] + N_\theta^T\mathbb{E}\big[\psi_V(U)\psi_V^T(U)\big]N_\theta + \mathbb{E}\big[m^{(1),\theta}_{\gamma^*_\theta}(U)\psi_V^T (U)\big]N_\theta + N_\theta^T\mathbb{E}\big[\psi_V(U)\big(m^{(1),\theta}_{\gamma^*_\theta}(U)\big)^T\big]$, and $A_{\theta,\theta'} \equiv \mathbb{E}\big[m^{(1),\theta}_{\gamma^*_\theta}(U)\big(m^{(1),\theta}_{\gamma^*_{\theta'}}(U)\big)^T\big] + N_\theta^T\mathbb{E}\big[\psi_V(U) \psi_V^T(U)\big]N_{\theta'} + \mathbb{E}\big[m^{(1),\theta}_{\gamma^*_\theta}(U)\psi_V^T(U)\big]N_{\theta'} + N_\theta^T\mathbb{E}\big[\psi_V(U)\big(m^{(1),\theta}_{\gamma^*_{\theta'}}(U)\big)^T\big]$. It is straightforward now to verify that, as $\theta \to \theta_0$, $W_{a,\theta} \to W_{a,\theta_0}$, $m^{(1),\theta}_{\gamma^*_\theta} \to m^{(1),\theta}_{\gamma^*_{\theta_0}}$, and $N_\theta \to N_{\theta_0}$. Also, $V_\theta \to V_{\theta_0}$ and $V_\theta^{-1} \to V_{\theta_0}^{-1}$. Then, $\mathbb{E}\big[\|g_\theta - g_{\theta_0}\|^2\big] = V_\theta^{-1}A_\theta V_\theta^{-1} + V_{\theta_0}^{-1}A_{\theta_0}V_{\theta_0}^{-1} - 2V_\theta^{-1}A_{\theta,\theta_0}V_{\theta_0}^{-1} \to 0$. Therefore, by Lemma 13.3 in Kosorok (2008), since $\hat{\theta} \to \theta_0$,

$$\sqrt{n}(\hat{\gamma}_{\hat{\theta}} - \gamma^*_{\hat{\theta}}) \rightsquigarrow -V_{\theta_0}^{-1}G_{\theta_0}, \tag{B.37}$$

where G_{θ_0} is mean zero Gaussian process with covariance $A_{\theta_0} = \mathbb{E}\big[m^{(1),\theta}_{\gamma^*_{\theta_0}}(U)(m^{(1),\theta}_{\gamma^*_{\theta_0}}(U)^T\big] + N_{\theta_0})^T\mathbb{E}\big[\psi_V(U)\psi_V^T(U)\big]N_{\theta_0} + \mathbb{E}\big[m^{(1),\theta}_{\gamma^*_{\theta_0}}(U)\psi_V^T(U)\big]N_{\theta_0} + N_{\theta_0}^T\mathbb{E}\big[\psi_V(U)\big(m^{(1),\theta}_{\gamma^*_{\theta_0}}(U)\big)^T\big]$. For the latter term on (B.28), we obtain the limiting distribution of $\sqrt{n}(\gamma^*_{\hat{\theta}} - \gamma^*_{\theta_0})$. Let $\dot{M}_\theta(\gamma) \equiv \frac{\partial}{\partial\gamma}M_\theta(\gamma) = \mathbb{E}\big[-\sum_a |W_{a,\theta}(U)|\dot{\phi}\{a \cdot \operatorname{sgn}(W_{a,\theta}(U))\gamma^T\xi(\boldsymbol{X})\}a \cdot \operatorname{sgn}(W_{a,\theta}(U))\xi(\boldsymbol{X})\big]$, and note that $\dot{M}_\theta(\gamma) = 0$. Then, from Taylor's Theorem,

$$\begin{aligned} &\sqrt{n}(\gamma_{\hat{\theta}^*} - \gamma^*_{\theta_0}) && \text{(B.38)} \\ =&\sqrt{n}B_0(\hat{\theta} - \theta_0) + o_P(\sqrt{n}\|\hat{\theta} - \theta_0\|)), && \text{(B.39)} \end{aligned}$$

where $B_0 = \left(\frac{\partial\gamma}{\partial\theta}\right)_{\theta=\theta_0}$ is a $q \times p$ matrix. We can obtain this since $\left(\frac{\partial\gamma}{\partial\theta}\right)_{\theta=\theta_0} = -\big(A_1^{-1} A_2\big)_{\theta=\theta_0}$ where

$$
\begin{aligned}
A_1 &= \frac{\partial \dot{M}_\theta(\gamma)}{\partial \gamma} \\
&= -\mathbb{E}\Big[\big\{\sum_a |W_{a,\theta}(U)|\ddot{\phi}(a \cdot \mathrm{sgn}(W_{a,\theta}(U))a\gamma_\theta^{*T}\xi(\boldsymbol{X}))\big\}\xi(\boldsymbol{X})\xi(\boldsymbol{X})^T\Big], \text{ and} \\
A_2 &= \frac{\partial \dot{M}_\theta(\gamma)}{\partial \theta} \\
&= -\mathbb{E}\Big[\sum_a a \cdot \Big(\frac{I(A=a)}{\pi(a;\boldsymbol{X})}(Y-Z) - \frac{I(A=a)-\pi(a;\boldsymbol{X})}{\pi(a;\boldsymbol{X})}(Q_Y(\boldsymbol{X}) - Q_Z(\boldsymbol{X}))\Big) \\
&\cdot \omega_\theta(\boldsymbol{X})(1-\omega_\theta(\boldsymbol{X}))\dot{\phi}\{a \cdot \mathrm{sgn}(W_{a,\theta}(U))\gamma_\theta^{*T}\xi(\boldsymbol{X})\}\xi(\boldsymbol{X})X^T\Big] \\
&- 2\mathbb{E}\Big[\sum_a W_{a,\theta}(U)\Big(\frac{I(A=a)}{\pi(a;\boldsymbol{X})}(Y-Z) - \frac{I(A=a)-\pi(a;\boldsymbol{X})}{\pi(a;\boldsymbol{X})}(Q_Y(\boldsymbol{X}) - Q_Z(\boldsymbol{X}))\Big) \\
&\cdot \omega_\theta(\boldsymbol{X})(1-\omega_\theta(\boldsymbol{X}))\ddot{\phi}\big\{a \cdot \mathrm{sgn}(W_{a,\theta}(U))\gamma_\theta^{*T}\xi(\boldsymbol{X})\big\}\gamma_\theta^{*T}\xi(\boldsymbol{X})\xi(\boldsymbol{X})\boldsymbol{X}^T\Big| W_{a,\theta}(U) = 0\Big].
\end{aligned}
$$

Then the desired conclusions follow.

■

Lemma B.4. *Assume that* $E\|\hat{\pi}(a;\boldsymbol{X}) - \pi(a;\boldsymbol{X})\|_{P,2}^2 = O(n^{-1})$ *and* $E\|\hat{Q}_\theta(\boldsymbol{X}, a) - Q_\theta(\boldsymbol{X}, a)\|_{P,2}^2 = O(n^{-1})$*. Let* $\hat{f}_{n,\theta} = \arg\max_{f\in\mathcal{F}} \hat{M}_{n,\theta}(f)$*, where* $\hat{M}_{n,\theta}(f) = \mathbb{E}_n[-\sum_a |\hat{W}_{a,\theta}|\phi(a \cdot \mathrm{sgn}(\hat{W}_{a,\theta})f) + \lambda_n\|f\|^2]$*, and* $f_\theta^* = \arg\max_{f\in\mathcal{F}} M_\theta(f)$*, where* $M_\theta(f) = \mathbb{E}[-\sum_a |W_{a,\theta}|\phi(a \cdot \mathrm{sgn}(W_{a,\theta})f)]$*. Then, the rate of convergence of* $\hat{f}_{n,\theta}$ *to* f_θ^* *is* $r_n = \sqrt{n}$ *uniformly over* $\theta \in \Theta_0^\epsilon$*.*

Proof of Lemma B.4. We prove this by verifying the conditions in Theorem 2.3. For f in the neighborhood N_θ,

$$M_\theta(f) - M_\theta(f_\theta^*) \lesssim -c_1 d(f, f_\theta^*)$$

for $c_1 > 0$ by the definition of f^*_θ, and non-singularity and continuity of the second derivative matrix of $\ddot{M}_\theta(f)$ uniformly over $\theta \in \Theta_0^\epsilon$. Next, for all $\theta \in \Theta_0^\epsilon$

$$\begin{aligned}
&\sqrt{n}(\hat{M}_{n,\theta} - M_\theta)f - \sqrt{n}(\hat{M}_{n,\theta} - M_\theta)f^*_\theta \\
&= \mathbb{G}_n\Bigg(- |\hat{W}_{1,\theta}|\phi(\operatorname{sgn}(\hat{W}_{1,\theta})f) - |\hat{W}_{-1,\theta}|\phi(-\operatorname{sgn}(\hat{W}_{-1,\theta})f) - \lambda_n\|f\|^2 \\
&\qquad\qquad + |\hat{W}_{1,\theta}|\phi(\operatorname{sgn}(\hat{W}_{1,\theta})f^*_\theta) + |\hat{W}_{-1,\theta}|\phi(-\operatorname{sgn}(\hat{W}_{-1,\theta})f^*_\theta) + \lambda_n\|f^*_\theta\|^2\Bigg) \qquad \text{(B.40)} \\
&- \sqrt{n}\mathbb{E}\Bigg(\sum_a \Big(|\hat{W}_{a,\theta}|\big\{\phi(a \cdot \operatorname{sgn}(\hat{W}_{a,\theta})f) - \phi(a \cdot \operatorname{sgn}(\hat{W}_{a,\theta})f^*_\theta\big\} \\
&\qquad\qquad - |W_{a,\theta}|\big\{\phi(a \cdot \operatorname{sgn}(W_{a,\theta}) - \phi(a \cdot \operatorname{sgn}(W_{a,\theta}))\big\}\Big) + \lambda_n^\theta(\|f\|^2 - \|f^*_\theta\|^2)\Bigg). \qquad \text{(B.41)}
\end{aligned}$$

For fixed $\theta \in \Theta_0^\epsilon$, let $\mathcal{G}_\theta = \{g_\theta(f) : g_\theta(f) = -|\hat{W}_{1,\theta}|\phi(\operatorname{sgn}(\hat{W}_{1,\theta})f) - |\hat{W}_{-1,\theta}|\phi(-\operatorname{sgn}(\hat{W}_{-1,\theta})f) + |\hat{W}_{1,\theta}|\phi(\operatorname{sgn}(\hat{W}_{1,\theta})f^*_\theta) + |\hat{W}_{-1,\theta}|\phi(-\operatorname{sgn}(\hat{W}_{-1,\theta})f^*_\theta) - \lambda_n^\theta(\|f\|^2 - \|f^*_\theta\|^2), f \in N_\theta\}$, and let G_θ be an envelope function for $\mathcal{G}_\theta$. Then for (B.40), by Theorem 11.1 in Kosorok (2008), we have

$$\begin{aligned}
&\mathbb{E} \sup_{d(f,f_0)<\delta} \big|\mathbb{G}_n(g_\theta(f))\big| \\
&\leq \mathbb{E} \sup_{g_\theta \in \mathcal{G}} \big|\mathbb{G}_n(g_\theta(f))\big| \\
&\leq c_2 J^*(1, \mathcal{G}_\theta)\|G_\theta\|_{P,1}
\end{aligned}$$

for some $c_2 < \infty$. $J^*(1, \mathcal{G})$ is computed as

$$\begin{aligned}
J^*(1, \mathcal{G}_\theta) &= \sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon\|G_\theta\|_{Q,2}, \mathcal{G}_\theta, L_2(Q))}d\epsilon \\
&\leq \int_0^1 \sqrt{1 + K_{VC_\theta}(\frac{1}{\epsilon})^{2-2/VC_\theta}}d\epsilon \\
&\lesssim 1.
\end{aligned}$$

The first inequality is from Corollary 9.5 in Kosorok (2008). VC_θ is the VC-index of $\mathcal{G}_\theta$, and K_{VC_θ} is some constant greater than 0. Let $f_{1,\theta} = \arg\max_{d(f,f_\theta^*)\leq\delta} \big\{ \sum_{a\in\{-1,1\}} |\hat{W}_{a,\theta}|\{\phi(a\cdot \operatorname{sgn}(\hat{W}_{a,\theta})f) - \phi(a\cdot\operatorname{sgn}(\hat{W}_{a,\theta})f_\theta^*)\} - \lambda_n^\theta(\|f\|^2 - \|f_\theta^*\|^2)\big\}$. Then by Assumption 2.8 and Hölder's inequality,

$$\begin{aligned}
&\sup_{\theta\in\Theta_0^\epsilon} \|G_\theta\|_{P,1} \\
&= \sup_{\theta\in\Theta_0^\epsilon} \Bigg\| \sum_{a\in\{-1,1\}} \Big| (\frac{I(A=a)}{\hat{\pi}(a;\boldsymbol{X})} - \frac{I(A=a)}{\pi(a;\boldsymbol{X})})(U_\theta - Q_\theta(\boldsymbol{X},a)) + \frac{I(A=a)}{\pi(a;\boldsymbol{X})}(U_\theta - Q_\theta(\boldsymbol{X},a)) \\
&+ (\frac{I(A=a)}{\hat{\pi}(a;\boldsymbol{X})} - \frac{I(A=a)}{\pi(a;\boldsymbol{X})})(Q_\theta(\boldsymbol{X},a) - \hat{Q}_\theta(\boldsymbol{X},a)) + (\frac{I(A=a)}{\pi(a;\boldsymbol{X})} - 1)(Q_\theta(\boldsymbol{X},a) \\
&\qquad - \hat{Q}_\theta(\boldsymbol{X},a)) \\
&+ Q_\theta(X,a)\Big| (\phi(a\cdot\operatorname{sgn}(\hat{W}_{a,\theta})f_{1,\theta}) - \phi(a\cdot\operatorname{sgn}(\hat{W}_{a,\theta})f_\theta^*)) - \lambda_n^\theta(\|f_{1,\theta}\|^2 - \|f_\theta^*\|^2) \Bigg\|_{P,2} \\
&\leq \sup_{\theta\in\Theta_0^\epsilon} \sum_{a\in\{-1,1\}} \Bigg\| \frac{I(A=a)}{\hat{\pi}(a;\boldsymbol{X})} - \frac{I(A=a)}{\pi(a;\boldsymbol{X})})(U_\theta - Q_\theta(\boldsymbol{X},a)) + \frac{I(A=a)}{\pi(a;\boldsymbol{X})}(U_\theta - Q_\theta(\boldsymbol{X},a)) \\
&+ (\frac{I(A=a)}{\pi(a;\boldsymbol{X})} - 1)(Q_\theta(\boldsymbol{X},a) - \hat{Q}_\theta(\boldsymbol{X},a)) + Q_\theta(\boldsymbol{X},a)\Bigg\|_{P,2} \cdot \Bigg\| \phi(a\cdot\operatorname{sgn}(\hat{W}_{a,\theta})f_{1,\theta}) \\
&\qquad - \phi(a\cdot\operatorname{sgn}(\hat{W}_{a,\theta})f_\theta^*)\Bigg\|_{P,2} \\
&+ \sup_{\theta\in\Theta_0^\epsilon} \sum_{a\in\{-1,1\}} \Bigg\| (\frac{I(A=a)}{\hat{\pi}(a;\boldsymbol{X})} - \frac{I(A=a)}{\pi(a;\boldsymbol{X})})(Q_\theta(\boldsymbol{X},a) - \hat{Q}_\theta(\boldsymbol{X},a))\Bigg\|_{P,2} \cdot \Bigg\| \phi(a\cdot\operatorname{sgn}(\hat{W}_{a,\theta}) \\
&\qquad \cdot f_{1,\theta}) - \phi(a\cdot\operatorname{sgn}(\hat{W}_{a,\theta})f_\theta^*)\Bigg\|_{P,2} + \Bigg\| \lambda_n^\theta(\|f_{1,\theta}\|^2 - \|f_\theta^*\|^2)\Bigg\|_{P,2} \\
&\leq O(n^{-\frac{1}{2}}\delta)
\end{aligned}$$

For (B.41),

$$
\begin{aligned}
&\sqrt{n}\sup_{\theta\in\Theta_0^\epsilon}\mathbb{E}\Bigg(\sum_a\Big(|\hat{W}_{a,\theta}|\big\{\phi(a\cdot\mathrm{sgn}(\hat{W}_{a,\theta})f)-\phi(a\cdot\mathrm{sgn}(\hat{W}_{a,\theta})f_\theta^*\big\}\\
&\qquad\qquad-|W_{a,\theta}|\big\{\phi(a\cdot\mathrm{sgn}(W_{a,\theta})-\phi(a\cdot\mathrm{sgn}(W_{a,\theta}))\big\}\Big)+\lambda_n^\theta(\|f\|^2-\|f_\theta^*\|^2)\Bigg)\\
&\lesssim\sqrt{n}O(n^{-\frac{1}{2}})\delta\\
&=O(1)\delta.
\end{aligned}
$$

Let $\phi_n(\delta)=\delta$. Also, let $\gamma=\frac{3}{2}$. Then $\frac{\phi_n(\delta)}{\delta^\gamma}=\delta^{-\frac{1}{2}}$. Then, (2.7) is satisfied. Also, $r_n^2\phi(r_n^{-1})=r_n^2\cdot\frac{1}{r_n}=r_n\leq c_3\sqrt{n}$. Choose $r_n=n^{\frac{1}{2}}$. Then, all conditions in Theorem 2.3 are satisfied. ■

Proof of Theorem 2.5. Let $\hat{l}_n(\theta,\beta)$ be a log of pseudo-likelihood $\hat{L}_n(\theta,\beta)$. Also, let $\hat{u}_n(\theta)=n^{-1/2}\sum_{i=1}^n\boldsymbol{X}_i\big[1\big\{A_i=\hat{d}_{n,\theta}(\boldsymbol{X}_i)\big\}-P_{\beta_0}(\boldsymbol{X}_i)\big]$. We use similar arguments as in the proof for the Theorem 13 in Luckett et al. (2021). We have

$$
\begin{aligned}
&n^{-1/2}\{\hat{l}_n(\hat{\theta}_n,\hat{\beta}_n)\}\\
=&n^{-1/2}\sum_{i=1}^n\boldsymbol{X}_i^T\beta_0\big[1\big\{A_i=\hat{d}_{n,\hat{\theta}_n}(\boldsymbol{X})\big\}-\big\{A_i=\hat{d}_{n,\theta_0}(\boldsymbol{X})\big\}\big]+\frac{1}{2}v_n(\hat{\theta}_n,\beta_*)\\
=&n^{1/2}\mathbb{E}\Big(\boldsymbol{X}^T\beta_0\big[1\big\{A=\hat{d}_{n,\hat{\theta}_n}(\boldsymbol{X})\big\}-1\big\{A=d_{\theta_0}^*(\boldsymbol{X})\big\}\big]\Big)\\
&-n^{1/2}\mathbb{E}\Big(\boldsymbol{X}^T\beta_0\big[1\big\{A=\hat{d}_{n,\theta_0}(\boldsymbol{X})\big\}-1\big\{A=d_{\theta_0}^*(\boldsymbol{X})\big\}\big]\Big)+o_P(1+\sqrt{n}\|\hat{\theta}_n-\theta_0\|)\\
=&-\mathbb{E}\Big(X^T\beta_0\{2P_{\beta_0}(\boldsymbol{X})-1\}\big|\sqrt{n}\{\hat{f}_{n,\hat{\theta}}(\boldsymbol{X})-f_{\theta_0}^*(\boldsymbol{X})\}\big|\Big|f_{\theta_0}^*(\boldsymbol{X})=0\Big)\{1+o_P(1)\}g_0\\
&\quad+O_P(1)+o_P(1+\sqrt{n}\|\hat{\theta}_n-\theta_0\|)\\
\leq&-\mathbb{E}\Big(\boldsymbol{X}^T\beta_0\{2P_{\beta_0}(\boldsymbol{X})-1\}\big|(-V_{\theta_0}^{-1}\tilde{Z}+B_0\sqrt{n}(\hat{\theta}-\theta_0))^T\xi(\boldsymbol{X})\big|\Big|f_{\theta_0}^*(\boldsymbol{X})=0\Big)\\
&\quad\cdot\{1+o_P(1)\}g_0+O_P(1)+o_P(1+\sqrt{n}\|\hat{\theta}_n-\theta_0\|)\\
\leq&-\delta_2\delta_1^2\left(\frac{\exp(\delta_1)-1}{\exp(\delta_1)+1}\right)\sqrt{n}\|\hat{\theta}-\theta_0\|\{1+o_P(1)\}+O_P(1)+o_P(1+\sqrt{n}\|\hat{\theta}-\theta_0\|),
\end{aligned}
$$

where $v_n(\hat{\theta}_n, \beta_*) = n^{-1/2}\hat{u}_n(\hat{\theta}_n)^T I_n^{-1}(\beta_*)\hat{u}_n(\hat{\theta}_n)$, and β_* is a point between $\hat{\beta}_n$ and β_0. Therefore, $\sqrt{n}\|\hat{\theta} - \theta_0\| = O_P(1)$.

Since the second term of the last equality does not depend on $\hat{\theta}$, we can let $\hat{\theta} = \arg\max_{\tilde{u}\in\mathbb{R}^p}$ $M_n(\theta_0 + \tilde{u}/\sqrt{n})$ where

$$M_n(\theta_0 + \tilde{u}/\sqrt{n}) = n^{-1/2}\sum_{i=1}^{n} \boldsymbol{X}_i^T\beta_0\Big[1\big\{A_i = \hat{d}_{n,\theta_0+\tilde{u}/\sqrt{n}}(\boldsymbol{X}_i)\big\} - 1\big\{A_i = d^*_{\theta_0}(\boldsymbol{X}_i)\big\}\Big] \\ + v_n(\theta_0 + \tilde{u}/\sqrt{n}, \beta_*)/2.$$

Let $h_n(\tilde{u}) = \theta_0 + \tilde{u}/\sqrt{n}$. We use similar arguments as in Luckett et al. (2021). Since we have $o_P(1 + \sqrt{n}\|\hat{\theta} - \theta_0\|) = o_P(1)$, for any compact set $K \subset \mathbb{R}^p$,

$$\begin{aligned}
&\arg\min_{\tilde{u}\in K} M_n\{h_n(\tilde{u})\} \\
=&\arg\min_{\tilde{u}\in K} n^{1/2}\mathbb{E}\bigg(\boldsymbol{X}^T\beta_0\big[1\big\{A_i = \hat{d}_{h_n(\tilde{u})}(\boldsymbol{X}_i)\big\} - 1\big\{A_i = d^*_{\theta_0}(\boldsymbol{X}_i)\big\}\big]\bigg) + o_P(1) \\
=&\arg\min_{\tilde{u}\in K} \mathbb{E}\Big(\boldsymbol{X}^T\beta_0\{2P_{\beta_0}(\boldsymbol{X}) - 1\}\big|\sqrt{n}\{\hat{f}_{h_n(\tilde{u})}(\boldsymbol{X}) - f^*_{\theta_0}(\boldsymbol{X})\}\big|\Big|f_{\theta_0}(\boldsymbol{X}) = 0\Big)g_0 + o_P(1) \\
\rightsquigarrow&\arg\min_{\tilde{u}\in K} \mathbb{E}\Big(\boldsymbol{X}^T\beta_0\{2P_{\beta_0}(\boldsymbol{X}) - 1\}|(-V_{\theta_0}^{-1}\tilde{Z} + B_0\tilde{u})^T\xi(\boldsymbol{X})|\Big|f_{\theta_0}(\boldsymbol{X}) = 0\Big)g_0,
\end{aligned}$$

where Z is a tight mean zero Gaussian process with covariance A_{θ_0}. Let $M(\tilde{u}) = \mathbb{E}\Big(\boldsymbol{X}^T\beta_0\{2 \cdot P_{\beta_0}(\boldsymbol{X}) - 1\}|(-V_{\theta_0}^{-1}\tilde{Z} + B_0\tilde{u})^T\xi(\boldsymbol{X})\Big|f_{\theta_0}(\boldsymbol{X}) = 0\Big)$. By the *argmax* theorem in chapter 14 of Kosorok (2008), since $M_n(h_n(\tilde{u})) \rightsquigarrow M(\tilde{u})$ in $l^\infty(K)$, we have $\tilde{U}_n \rightsquigarrow \tilde{U}$ where $\tilde{U}_n = \arg\min_{u\in\mathbb{R}^p} M_n(h_n(u))$ and $\tilde{U} = \arg\min_{u\in\mathbb{R}^p} M(u)$. Also, since $\sqrt{n}(\hat{\beta}_n - \beta_0) = I_n(\beta_*)^{-1}\hat{u}_n(\hat{\theta}_n)$ and $\hat{u}(\hat{\theta}_n) = Z_{A,n} + \sqrt{n}\mathbb{E}_n[\boldsymbol{X}(1\{A = \hat{d}_{n.\hat{\theta}_n}(\boldsymbol{X})\} - 1\{A = d^*_{\theta_0}(\boldsymbol{X})\}]$, $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges weakly to $I_0^{-1}(Z_A - k(\tilde{Z}, \tilde{U}))$.

■

BIBLIOGRAPHY

Abdolell, M., LeBlanc, M., Stephens, D., and Harrison, R. (2002). Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. *Statistics in medicine*, 21(22):3395–3409.

Allen, K. D., Arbeeva, L., Callahan, L. F., Golightly, Y. M., Goode, A. P., Heiderscheit, B., Huffman, K., Severson, H., and Schwartz, T. (2018). Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. *Osteoarthritis and cartilage*, 26(3):383–396.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178.

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973.

Bannuru, R. R., Osani, M., Vaysbrot, E., Arden, N., Bennell, K., Bierma-Zeinstra, S., Kraus, V., Lohmander, L. S., Abbott, J., Bhandari, M., et al. (2019). Oarsi guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. *Osteoarthritis and cartilage*, 27(11):1578–1589.

Bartlett, P. L., Jordan, M. I., and McAuliffe, J. D. (2006). Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473):138–156.

Bellamy, N. (2002). Womac: a 20-year experiential review of a patient-centered self-reported health status questionnaire. *The Journal of rheumatology*, 29(12):2473–2476.

Bénard, C., Biau, G., Da Veiga, S., and Scornet, E. (2021). Interpretable random forests via rule extraction. In *International Conference on Artificial Intelligence and Statistics*, pages 937–945. PMLR.

Blatt, D., Murphy, S. A., and Zhu, J. (2004). A-learning for approximate planning. *Ann Arbor*, 1001:48109–2122.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and regression trees.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Cho, H., Holloway, S. T., Couper, D. J., and Kosorok, M. R. (2020). Multi-stage optimal dynamic treatment regimes for survival outcomes with dependent censoring. *arXiv preprint arXiv:2012.03294*.

Cho, H., Jewell, N. P., and Kosorok, M. R. (2021). Interval censored recursive forests. *Journal of Computational and Graphical Statistics*, pages 1–13.

Ciampi, A. (1991). Generalized regression trees. *Computational Statistics & Data Analysis*, 12(1):57–78.

Coffman, C., Arbeeva, L., Schwartz, T., Callahan, L., Golightly, Y., Goode, A., Huffman, K., and Allen, K. (2021). Application of heterogeneity of treatment effects methods: exploratory analyses of a trial of exercise-based interventions for knee oa. *Arthritis Care & Research*.

Cui, Y., Kosorok, M. R., Sverdrup, E., Wager, S., and Zhu, R. (2020). Estimating heterogeneous treatment effects with right-censored data via causal survival forests. *arXiv preprint arXiv:2001.09887*.

Cui, Y., Zhu, R., and Kosorok, M. (2017). Tree based weighted learning for estimating individualized treatment rules with censored data. *Electronic journal of statistics*, 11(2):3927.

Davis, R. B. and Anderson, J. R. (1989). Exponential survival trees. *Statistics in medicine*, 8(8):947–961.

Dusseldorp, E. and Van Mechelen, I. (2014). Qualitative interaction trees: a tool to identify qualitative treatment–subgroup interactions. *Statistics in medicine*, 33(2):219–237.

Fan, A., Lu, W., and Song, R. (2016). Sequential advantage selection for optimal treatment regime. *The annals of applied statistics*, 10(1):32.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139.

Freund, Y. and Schapire, R. E. (1998). Large margin classification using the perceptron algorithm. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 209–217.

Friedman, J., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2):337–407.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Goldberg, Y. and Kosorok, M. R. (2012). Q-learning with censored data. *Annals of statistics*, 40(1):529.

Gordon, L. and Olshen, R. A. (1985). Tree-structured survival analysis. *Cancer treatment reports*, 69(10):1065–1069.

Hernán, M. A. and Robins, J. M. (2010). Causal inference.

Jiang, X., Nelson, A. E., Cleveland, R. J., Beavers, D. P., Schwartz, T. A., Arbeeva, L., Alvarez, C., Callahan, L. F., Messier, S., Loeser, R., et al. (2020). A precision medicine approach to develop and internally validate optimal exercise and weight loss treatments for overweight and obese adults with knee osteoarthritis. *Arthritis Care & Research*.

Kallus, N. (2017). Recursive partitioning for personalization using observational data. In *International conference on machine learning*, pages 1789–1798. PMLR.

Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science*, 22(4):523–539.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2):119–127.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.

Kim, H. and Loh, W.-Y. (2001). Classification trees with unbiased multiway splits. *Journal of the American Statistical Association*, 96(454):589–604.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics*, pages 191–219.

Kolasinski, S. L., Neogi, T., Hochberg, M. C., Oatis, C., Guyatt, G., Block, J., Callahan, L., Copenhaver, C., Dodge, C., Felson, D., et al. (2020). 2019 american college of rheumatology/arthritis foundation guideline for the management of osteoarthritis of the hand, hip, and knee. *Arthritis & Rheumatology*, 72(2):220–233.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer.

Kosorok, M. R. and Laber, E. B. (2019). Precision medicine. *Annual review of statistics and its application*, 6:263–286.

Kosorok, M. R. and Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics*, 35(3):957–989.

Laber, E. B., Linn, K. A., and Stefanski, L. A. (2014). Interactive model building for q-learning. *Biometrika*, 101(4):831–847.

Lange, K. and Tong Wu, T. (2008). An mm algorithm for multicategory vertex discriminant analysis. *Journal of Computational and Graphical Statistics*, 17(3):527–544.

LeBlanc, M. and Crowley, J. (1992). Relative risk trees for censored survival data. *Biometrics*, pages 411–425.

Liu, Y., Wang, Y., Kosorok, M. R., Zhao, Y., and Zeng, D. (2018). Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Statistics in medicine*, 37(26):3776–3788.

Liu, Y. and Yuan, M. (2011). Reinforced multicategory support vector machines. *Journal of Computational and Graphical Statistics*, 20(4):901–919.

Loh, W.-Y. (2009). Improving the precision of classification trees. *The Annals of Applied Statistics*, pages 1710–1737.

Loh, W.-Y. and Vanichsetakul, N. (1988). Tree-structured classification via generalized discriminant analysis. *Journal of the American Statistical Association*, 83(403):715–725.

Luckett, D. J., Laber, E. B., Kahkoska, A. R., Maahs, D. M., Mayer-Davis, E., and Kosorok, M. R. (2019). Estimating dynamic treatment regimes in mobile health using v-learning. *Journal of the American Statistical Association*.

Luckett, D. J., Laber, E. B., Kim, S., and Kosorok, M. R. (2021). Estimation and optimization of composite outcomes. *J. Mach. Learn. Res.*, 22:167–1.

Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine*, 23(19):2937–2960.

Ma, S. and Kosorok, M. R. (2005). Penalized log-likelihood estimation for partly linear transformation models with current status data. *The Annals of Statistics*, 33(5):2256–2290.

Meinshausen, N. (2010). Node harvest. *The Annals of Applied Statistics*, pages 2049–2072.

Meng, H. and Qiao, X. (2020). Doubly robust direct learning for estimating conditional average treatment effect. *arXiv preprint arXiv:2004.10108*.

Moodie, E. E., Richardson, T. S., and Stephens, D. A. (2007). Demystifying optimal dynamic treatment regimes. *Biometrics*, 63(2):447–455.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355.

Murphy, S. A. (2005). A generalization error for q-learning. *Journal of machine learning research: JMLR*, 6:1073.

Murray, T. A., Thall, P. F., and Yuan, Y. (2016). Utility-based designs for randomized comparative trials with categorical outcomes. *Statistics in medicine*, 35(24):4285–4305.

Neugebauer, R. and van der Laan, M. (2005). Why prefer double robust estimators in causal inference? *Journal of statistical planning and inference*, 129(1-2):405–426.

Ng, A. Y., Russell, S., et al. (2000). Algorithms for inverse reinforcement learning. In *Icml*, volume 1, page 2.

Oprescu, M., Syrgkanis, V., and Wu, Z. S. (2019). Orthogonal random forest for causal inference. In *International Conference on Machine Learning*, pages 4932–4941. PMLR.

Qi, Z., Liu, D., Fu, H., and Liu, Y. (2020). Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *Journal of the American Statistical Association*, 115(530):678–691.

Qi, Z. and Liu, Y. (2018). D-learning to estimate optimal individual treatment rules. *Electronic Journal of Statistics*, 12(2):3601–3638.

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

Quinlan, J. R. (1993). *C4. 5: Programs for Machine Learning*. Morgan Kaufmann.

Robins, J. M. (1989). The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS*, pages 113–159.

Robins, J. M. (1997). Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*, pages 69–117. Springer.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the second seattle Symposium in Biostatistics*, pages 189–326. Springer.

Robins, J. M. and Rotnitzky, A. (2001). Comment on the bickel and kwon article,"inference for semiparametric models: Some questions and an answer". *Statistica Sinica*, 11(4):920–936.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866.

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

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593.

Sachs, G. S., Nierenberg, A. A., Calabrese, J. R., Marangell, L. B., Wisniewski, S. R., Gyulai, L., Friedman, E. S., Bowden, C. L., Fossey, M. D., Ostacher, M. J., et al. (2007). Effectiveness of adjunctive antidepressant treatment for bipolar depression. *New England Journal of Medicine*, 356(17):1711–1722.

Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2):197–227.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120.

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.

Segal, M. R. (1992). Tree-structured methods for longitudinal data. *Journal of the American Statistical Association*, 87(418):407–418.

Seijo, E. and Sen, B. (2011). A continuous mapping theorem for the smallest argmax functional. *Electronic Journal of Statistics*, 5:421–439.

Shah, K. S., Fu, H., and Kosorok, M. R. (2022). Stabilized direct learning for efficient estimation of individualized treatment rules. *Biometrics*.

Shi, C., Fan, A., Song, R., and Lu, W. (2018). High-dimensional a-learning for optimal dynamic treatment regimes. *Annals of statistics*, 46(3):925.

Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. (1990). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, pages 465–472.

Stekhoven, D. J. and Bühlmann, P. (2012). Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160.

Tian, L., Alizadeh, A. A., Gentles, A. J., and Tibshirani, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association*, 109(508):1517–1532.

United States Bone and Joint Initiative (2020). United States Bone and Joint Initiative: The Burden of Musculoskeletal Diseases in the United States (BMUS), Fourth Edition. Rosemont, IL. Available at http://www.boneandjointburden.org. Accessed on Jan 12, 2021.

Van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, 6(1).

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Wang, L. and Shen, X. (2007). On l 1-norm multiclass support vector machines: methodology and theory. *Journal of the American Statistical Association*, 102(478):583–594.

Williams, Q. I., Gunn, A. H., Beaulieu, J. E., Benas, B. C., Buley, B., Callahan, L. F., Cantrell, J., Genova, A. P., Golightly, Y. M., Goode, A. P., et al. (2015). Physical therapy vs. internet-based exercise training (path-in) for patients with knee osteoarthritis: study protocol of a randomized controlled trial. *BMC musculoskeletal disorders*, 16(1):1–12.

Yadlowsky, S., Pellegrini, F., Lionetto, F., Braune, S., and Tian, L. (2021). Estimation and validation of ratio-based conditional average treatment effects using observational data. *Journal of the American Statistical Association*, 116(533):335–352.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, 100(3):681–694.

Zhang, C. and Liu, Y. (2013). Multicategory large-margin unified machines. *Journal of Machine Learning Research*, 14(5).

Zhang, C. and Liu, Y. (2014). Multicategory angle-based large-margin classification. *Biometrika*, 101(3):625–640.

Zhang, H. (1998). Classification trees for multiple binary responses. *Journal of the American Statistical Association*, 93(441):180–193.

Zhang, Y., Laber, E. B., Davidian, M., and Tsiatis, A. A. (2018). Interpretable dynamic treatment regimes. *Journal of the American Statistical Association*, 113(524):1541–1549.

Zhang, Y., Laber, E. B., Tsiatis, A., and Davidian, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, 71(4):895–904.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118.

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.

Zhao, Y.-Q., Laber, E. B., Ning, Y., Saha, S., and Sands, B. E. (2019). Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *The Journal of Machine Learning Research*, 20(1):1821–1843.

Zhao, Y.-Q., Zeng, D., Laber, E. B., and Kosorok, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598.

Zhou, X., Mayer-Hamblett, N., Khan, U., and Kosorok, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, 112(517):169–187.

Zhu, J. and Hastie, T. (2001). Kernel logistic regression and the import vector machine. *Advances in neural information processing systems*, 14.

Zhu, R. and Kosorok, M. R. (2012). Recursively imputed survival trees. *Journal of the American Statistical Association*, 107(497):331–340.