Statistical Learning Methods for Personalized Medicine

Xin Qiu

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

Statistical Learning Methods for Personalized Medicine

Xin Qiu

The theme of this dissertation is to develop simple and interpretable individualized treatment rules (ITRs) using statistical learning methods to assist personalized decision making in clinical practice. Considerable heterogeneity in treatment response is observed among individuals with mental disorders. Administering an individualized treatment rule according to patient-specific characteristics offers an opportunity to tailor treatment strategies to improve response. Black-box machine learning methods for estimating ITRs may produce treatment rules that have optimal benefit but lack transparency and interpretability. Barriers to implementing personalized treatments in clinical psychiatry include a lack of evidence-based, clinically interpretable, individualized treatment rules, a lack of diagnostic measure to evaluate candidate ITRs, a lack of power to detect treatment modifiers from a single study, and a lack of reproducibility of treatment rules estimated from single studies. This dissertation contains three parts to tackle these barriers: (1) methods to estimate the best linear ITR with guaranteed performance among the class of linear rules; (2) a tree-based method to improve the performance of a linear ITR fitted from the overall sample and identify subgroups with a large benefit; and (3) an integrative learning combining information across trials to provide an integrative ITR with improved efficiency and reproducibility.

In the first part of the dissertation, we propose a machine learning method to estimate optimal linear individualized treatment rules for data collected from single stage randomized controlled trials (RCTs). In clinical practice, an informative and practically useful treatment rule should be simple and transparent. However, because simple rules are likely to be far from optimal, effective methods to construct such rules must guarantee performance, in terms of yielding the best clinical outcome (highest reward) among the class of simple rules under consideration. Furthermore, it is important to evaluate the benefit of the derived rules on the whole sample and in pre-specified

subgroups (e.g., vulnerable patients). To achieve both goals, we propose a robust machine learning algorithm replacing zero-one loss with an authentic approximation loss (ramp loss) for value maximization, referred to as the asymptotically best linear O-learning (ABLO), which estimates a linear treatment rule that is guaranteed to achieve optimal reward among the class of all linear rules. We then develop a diagnostic measure and inference procedure to evaluate the benefit of the obtained rule and compare it with the rules estimated by other methods. We provide theoretical justification for the proposed method and its inference procedure, and we demonstrate via simulations its superior performance when compared to existing methods. Lastly, we apply the proposed method to the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial on major depressive disorder (MDD) and show that the estimated optimal linear rule provides a large benefit for mildly depressed and severely depressed patients but manifests a lack-of-fit for moderately depressed patients.

The second part of the dissertation is motivated by the results of real data analysis in the first part, where the global linear rule estimated by ABLO from the overall sample performs inadequately on the subgroup of moderately depressed patients. Therefore, we aim to derive a simple and interpretable piece-wise linear ITR to maintain certain optimality that leads to improved benefit in subgroups of patients, as well as the overall sample. In this work, we propose a tree-based robust learning method to estimate optimal piece-wise linear ITRs and identify subgroups of patients with a large benefit. We achieve these goals by simultaneously identifying qualitative and quantitative interactions through a tree model, referred to as the composite interaction tree (CITree). We show that it has improved performance compared to existing methods on both overall sample and subgroups via extensive simulation studies. Lastly, we fit CITree to Research Evaluating the Value of Augmenting Medication with Psychotherapy (REVAMP) trial for treating major depressive disorders, where we identified both qualitative and quantitative interactions and subgroups of patients with a large benefit.

The third part deals with the difficulties in the low power of identifying ITRs and replicating ITRs due to small sample sizes of single randomized controlled trials. In this work, a novel integrative learning method is developed to synthesize evidence across trials and provide an integrative ITR that improves efficiency and reproducibility. Our method does not require all studies to collect a common set of variables and thus allows information to be combined from ITRs identified from

randomized controlled trials with heterogeneous sets of baseline covariates collected from different domains with different resolution. Based on the research goal, the integrative learning can be used to enhance a high-resolution ITR by borrowing information from coarsened ITRs or improve the coarsened ITR from a high-resolution ITR. With a simple modification, the proposed integrative learning can also be applied to improve the estimation of ITRs for studies with blockwise missing feature variables. We conduct extensive simulation studies to show that our method has improved performance compared to existing methods where only single-trial ITRs are used to learn personalized treatment rules. Lastly, we apply the proposed method to RCTs of major depressive disorder and other comorbid mental disorders. We found that by combining information from two studies, the integrated ITR has a greater benefit and improved efficiency compared to single-trial rules or universal non-personalized treatment rule.

Key Words: Personalized medicine; Machine learning; Treatment response heterogeneity; Individualized treatment rules; Qualitative interaction; Quantitative interaction; Robust loss function; Tree-based method; Integrative learning; Blockwise missing data

Table of Contents

List of Figures

List of Tables

Acknowledgments

Foremost, I would like to express my deepest gratitude to my advisor, Professor Yuanjia Wang, for her invaluable guidance and constant support during the past four years. I could not have come this far without her motivation, encouragement, and trust in my abilities. During our weekly meeting, Professor Wang always gave me inspiring advice and insightful feedback to facilitate my research. I am very lucky and grateful to work with her for the past few years as a PhD student and research assistant. A special thanks goes to Professor Donglin Zeng, for providing invaluable suggestions and insightful comments to improve my work from the beginning of thesis proposal to the end of defense.

I would also like to express my sincere gratitude to my dissertation committee members, Professor Bruce Levin, Professor Bin Cheng, Professor Gen Li, and Professor Eva Petkova, who offered helpful feedback and inspiring questions to perfect this dissertation. And I would like to thank Professor Christine Mauro and Professor Katherine Shear for supervising my research assistantship.

I would like to thank the Department of Biostatistics for providing me fellowship. I wish to acknowledge all the faculty and staff in the department. In particular, I wish to thank all my classmates for sharing this journey together.

Last but not the least, I would like to thank my family and friends for their unconditional love and support.

To my parents and husband

Chapter 1

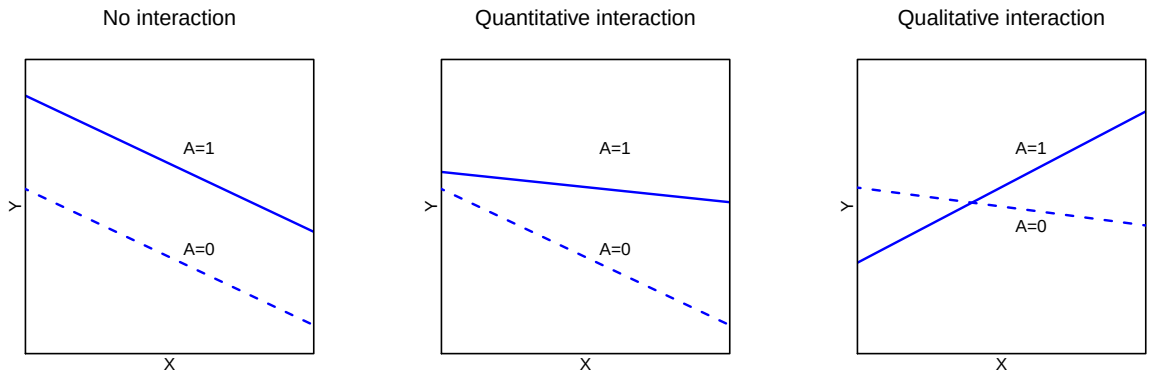
Introduction

1.1 Background and Overview

Heterogeneity in patient response to treatment is a long-recognized challenge in the clinical community. For example, in adults affected by major depression, only around 30% of patients achieve remission with a single acute phase of treatment (Trivedi et al., 2006; Rush et al., 2004); the remaining 70% of patients require augmentation of the current treatment or a switch to a new treatment (Trivedi and Daly, 2008). Heterogeneity in treatment response also has been observed among children with attention deficit and hyperactivity disorder (Pelham and Fabiano, 2008), and autism spectrum disorders (Jones et al., 2010). Thus, a universal strategy that treats all patients with the same treatment is inadequate, and individualized treatment strategies are required to improve response in individual patients. In this regard, rapid advances in technologies for collecting patient-level data have made it possible to tailor treatments to individual patients based on their characteristics, thereby enabling the new paradigm of personalized medicine.

Personalized medicine aims to provide each patient with the right medicine, the right dose, and

Figure 1.1: Three types of covariates (prognostic, predictive and prescriptive)



at the right time, in order to improve patient care and reduce potential side effects and health care cost (Carini et al., 2014). As depicted in Figure 1.1, three types of patient's covariates may be considered to achieve this goal. The first class of covariates includes prognostic variables which inform selecting subgroups of subjects at high risk for disease, irrespective of the treatment they receive (Carini et al., 2014). The other two classes of covariates correspond to variables with either quantitative or qualitative interaction with treatments, which are referred as predictive or prescriptive variables, respectively. In particular, qualitative interaction refers to that the treatment effect changes direction based on some function of covariates, which indicates that a treatment could be superior in one subgroup but inferior in another subgroup. This type of interaction provides important information on estimating a personalized treatment rule. Quantitative interaction refers to that the treatment effects are in the same direction in the covariate space but differ in magnitude in some subgroup. Predictive variables manifest quantitative interaction and define subgroups of subjects who are likely to experience large treatment benefit, while prescriptive variables manifest qualitative interaction and define the optimal treatment for a given individual. Both classes of

variables can be used to guide the development of personalized medicine (Carini et al., 2014).

In the following of this dissertation, we develop several new statistical learning methods for estimating simple and interpretable individualized treatment rules for single-stage randomized controlled trials. The dissertation consists of three projects. We start by introducing background and motivation of each project. In Chapter 2, we propose a machine learning method to estimate the optimal linear individualized treatment rule and a diagnostic measure to assess the optimality of candidate rules. In Chapter 3, we improve the performance of a linear ITR fitted from the overall sample using a tree-based model to identify both qualitative and quantitative interactions and subgroups of subjects with a large benefit. In Chapter 4, we use integrative learning to synthesize evidence across multiple trials to improve efficiency and reproducibility of the estimation of ITRs.

1.2 Introduction to Estimation and Evaluation of Linear Individualized Treatment Rules

Statistical methods have been proposed to estimate optimal individualized treatment rules (ITRs) (Lavori and Dawson, 2004) using predictive and prescriptive clinical variables that manifest quantitative and qualitative treatment interactions, respectively (Carini et al., 2014; Gunter et al., 2011). Q-learning (Watkins, 1989; Qian and Murphy, 2011) and A-learning (Murphy, 2003; Blatt et al., 2004) are proposed to identify the optimal ITR. Q-learning is a regression-based method, which estimates an ITR by directly modeling the Q-function ("Q" stands for "quality of action"). A-learning only requires posited models for contrast functions and uses a doubly robust estimating equation to estimate the contrast functions. This makes A-learning more robust to model misspecification than Q-learning and provides a consistent estimation of an ITR (Schulte et al., 2014). Other pro-

posed approaches include semiparametric methods and machine learning methods (Moodie et al., 2007; Zhang et al., 2012; Foster et al., 2011; Zhao et al., 2012; Chakraborty and Moodie, 2013). For example, the virtual twins approach (Foster et al., 2011) uses tree-based estimators to identify subgroups of patients who show larger than expected treatment effects. Zhang et al. (2012, 2013) estimated the optimal ITR by directly maximizing the value function over a specified parametric class of treatment rules through augmented inverse probability weighting. In contrast, Zhao et al. (2012) proposed outcome weighted learning (O-learning), which utilizes weighted support vector machine to directly maximize the value function (expected clinical outcome by following the ITR). More recently, Huang and Fong (2014) proposed a robust machine learning method to select the ITR that minimizes a total burden score due to disease and treatment for a binary clinical outcome. Interactive Q-learning (Laber et al., 2014) models two ordinary mean-variance functions instead of modeling the predicted future optimal outcomes. Fan et al. (2017) proposed a concordance function for prescribing treatment, where a patient is more likely to be assigned to a treatment than another patient if s/he has a greater benefit than the other patient.

In clinical practice, simple treatment rules such as linear rules, are preferred due to their transparency and convenience for interpretation. However, when only linear rules are in consideration, many existing methods including semiparametric models and some machine learning methods may not yield a rule with optimal performance, because they focus on optimization of a surrogate objective function of treatment benefit. Using surrogate objective functions may only guarantee the optimality when there is no restriction on the functional form of the treatment rules. For example, with O-learning, the objective function is a weighted hinge-loss, which yields the optimal rule among nonparametric rules, but may not be optimal when the candidate rules are restricted to the linear form. Therefore, learning algorithms are desired to derive a treatment rule with guaranteed

performance when constraints are placed on the class of candidate rules.

An additional consideration is the need to evaluate, through diagnostics, any approach for rule estimation. However, less emphasis has been placed on the evaluation of the estimated ITR in the context of personalized medicine. Residual plots were used to evaluate model fit for G-estimation (Rich et al., 2010) and Q-learning (Ertefaie et al., 2016). In the recent work by Wallace et al. (2016), a dynamic treatment regime (DTR) is estimated by G-estimation and double robustness is exploited for model diagnosis. How to evaluate the optimality of an ITR in general remains an open research question.

The purpose of this work is two-fold: we first develop a general approach to identify a linear ITR with guaranteed performance; we then propose a diagnostic method to evaluate the performance of any derived ITR including the proposed one. Our two-stage approach separates the estimation of the ITR from its evaluation and the sample used in each stage. Specifically, in the first stage, we propose ramp-loss-based (McAllester and Keshet, 2011; Huang and Fong, 2014) learning for the estimation and we show that this approach guarantees the derived linear ITR to be asymptotically optimal within the class of all linear rules. We refer our method as Asymptotically Best Linear O-learning, ABLO. For the second stage, in practice, it is infeasible to expect that an ITR that benefits each individual can be identified due to the unknown treatment mechanism and the likely omission of some prescriptive variables. Thus, we propose a practical solution to calibrate the average ITR effect in the population given the observed variables, or in pre-specified important subgroups (e.g., patients in the most severe state). Specifically, to obtain an ITR evaluation criterion, we define the benefit of a candidate ITR as the average difference in the value function between those who follow the ITR and those who do not. We then use the ITR benefit as a diagnostic measure to evaluate its optimality. Our method exploits the fact that if an ITR is truly optimal for all individuals, then

for any given patient subgroup, the average outcome for patients who are treated according to the ITR should be greater than for those who are not treated according to the ITR. On the contrary, if the average outcome of the ITR is worse for some patients who follow the ITR than for those who do not, then the ITR is not optimal on this subgroup.

Compared to the existing literature, two main contributions of this work are to propose a benefit function to calibrate an ITR, and a diagnostic procedure to evaluate optimality of a derived ITR, while most of the existing work focuses on the estimation of ITR/DTR. A third contribution is to prove asymptotic properties of ITR estimated under the ramp loss (Huang and Fong, 2014). Asymptotic results in the existing literature (e.g., Zhao et al. (2012)) are obtained for the hinge loss. Due to these theoretical results, we can provide valid statistical inference procedure for testing optimality of an ITR using asymptotic normality.

In Chapter 2, we show that ABLO consistently estimates the ITR benefit for a class of candidate rules regardless of two potential pitfalls: 1) the consistency of benefit estimator is maintained even though the functional form of the rule is misspecified; 2) the rule does not include all prescriptive/tailoring variables and thus the true global optimal rule is not in the specified class. We further derive the asymptotic distribution for the proposed diagnostic measure. We conduct simulation studies to demonstrate finite sample performance and show advantages over existing machine learning methods. Lastly, we apply the method to the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial on major depressive disorder (MDD), where substantial treatment response heterogeneity has been documented (Trivedi et al., 2006; Huynh and McIntyre, 2008). Our analyses estimate an optimal linear ITR, and we demonstrate a large benefit in mildly depressed and severely depressed patients but a lack-of-fit among moderately depressed patients.

1.3 Introduction to Composite Interaction Tree for Learning Optimal Individualized Treatment Rules and Subgroups

Due to their simplicity and interpretability, decision trees are often constructed to assist personalized medical decision making. Two types of interactions are useful for personalizing treatments: qualitative interactions inform the selection of optimal treatment from several competing choices; quantitative interactions inform the identification of subgroups with a substantially greater or smaller response than the overall sample. Methodologies are proposed to estimate an optimal ITR by hunting for qualitative interactions between covariates and treatments on the outcome (Carini et al., 2014; Gunter et al., 2011; Dusseldorp and Van Mechelen, 2014; Laber and Zhao, 2015). Specifically, Qualitative interaction tree (QUINT, Dusseldorp and Van Mechelen, 2014) partitions patients into terminal nodes where the average effect of one treatment for patients in the nodes is superior or two treatments are similar. Unlike the usual classification trees or regression trees where class labels are known, in tree-based methods for estimating optimal ITRs the optimal treatment (class label) for an individual is unknown. In this regard, Minimum Impurity Decision Assignments decision trees (MIDAs, Laber and Zhao, 2015) focused on the estimation of a tree-structured ITR that maximizes a value function and splits the parent nodes where the value function will be most dramatically improved. Fu et al. (2016) proposed to maximize the value function by exhaustive search; their approach was shown to be more stable due to using residuals of a regression model to remove main effects of covariates on the outcome when learning a tree (Liu et al., 2014). In a recent work by Zhu et al. (2017), outcome weighted learning (Zhao et al., 2012) and reinforcement learning tree (Zhu et al., 2015) were combined to construct a tree-based model to perform treatment selection. The algorithm uses greedy search where at each step, a weighted classification score is

used to evaluate the potential contribution of each variable.

Another group of methods is proposed to identify treatment response heterogeneity through estimating quantitative interactions. In this case, the optimal treatment may be the same for a subgroup of patients but they manifest a greater benefit than the overall sample. Subgroup Identification based on Differential Effect Search (SIDES, Lipkovich et al., 2011) searches within certain regions of the covariate spaces and identifies multiple subgroups with enhanced treatment effects. At each step, SIDES partitions a parent node into two child nodes for each covariate in a pre-specified set and retains the child node with a larger treatment effect. Virtual twins (Foster et al., 2011) finds a subgroup of patients who will have larger treatment effect using tree-based estimators. Interaction trees (ITs, Su et al., 2009) can identify both qualitative and quantitative interaction, but cannot distinguish strong quantitative interaction from qualitative interaction.

Some of the above existing tree-based approaches (e.g., QUINT, IT) examine qualitative interaction by exploring each candidate feature variable in turn. However, due to biological and clinical heterogeneity among patients, a single variable is unlikely to successfully guide treatment choice in individual patients; information carried by a single variable is limited, and a large number of variables each with small effects may play a role. Thus, tree-based approaches that partitions by individual variables may have reduced performance. For instance, in a randomized trial treating major depressive disorder (STAR*D, Rush et al., 2004), QUINT did not return any individual variable that can distinguish the best treatment to reduce depressive symptoms. In contrast, using the same study data, machine learning and regression-based approaches (Chakraborty and Moodie, 2013; Qiu et al., 2017) identified ITRs as linear combinations of feature variables that manifest a qualitative interaction to differentiate optimal treatments for individual patients. However, a linear ITR fitted from the overall sample using these methods may not be optimal on some patient

subgroups (due to the misspecification of linear rules), and thus do not identify subgroups with large ITR benefits.

Recognizing that estimating an ITR and identifying subgroups with large ITR benefits are two important goals, we propose a novel composite interaction tree, referred to as CITree, to simultaneously estimate qualitative and quantitative interactions. CITree contains two types of splits: a qualitative-interaction split by fitting asymptotically best linear rule (ABLO) at each stage; and a quantitative-interaction split based on a heterogeneity of ITR benefit (HTB) test. Specifically, in the qualitative split CITree estimates an interpretable, simple decision tree that guarantees enhanced performance on subgroups of patients by improving the value function fitted from the overall sample. Patients are partitioned into homogeneous subgroups of similar optimal ITR (reduce optimal treatment rule heterogeneity). In the quantitative split, CITree partitions patients into homogeneous subgroups of ITR benefit such that the within-group optimal ITR effect is similar while the between-group difference is large (reduce ITR benefit heterogeneity).

CITree leverages a diagnostic measure of the goodness-of-fit of a decision rule and a measure of the ITR benefit (i.e., the difference in the mean response for patients who follow the ITR and those who do not). It has been shown in Chapter 2 that if an optimal ITR genuinely maximizes the clinical response in each individual patient, then the ITR will have a positive effect within any arbitrarily defined subgroup of patients. Thus, by an HTB test, it is feasible to determine for which patients a linear (and potentially misspecified) ITR leads to a significantly lower than average (and potentially negative) benefit; thus the treatment is more likely to be non-optimal for these patients. To remedy this lack-of-fit of linear rules, as patients travel down the CITree, non-optimal ITRs with a poor benefit at top nodes are rectified in subgroups and patients organize into nodes of a high or low ITR benefit. By recursively detecting subgroups with a low benefit (quantitative splits) and

re-fit ITR within the subgroups (qualitative split), CITree will improve the overall value function and increase the subgroup benefit.

In Chapter 3 of this dissertation, we further introduce the rationale and algorithm of CITree. We show that CITree can successfully reduce the benefit heterogeneity and rule heterogeneity. We then perform extensive simulation studies to show improvement as compared to existing machine learning methods (e.g., IT, QUINT). Lastly, we fit a CITree using the Research Evaluating the Value of Augmenting Medication with Psychotherapy (REVAMP) trial data for treating major depressive disorder (MDD), where substantial treatment response heterogeneity was documented in the literature (Shankman et al., 2013).

1.4 Introduction to Integrative Learning to Synthesize Individualized Treatment Rules Across Multiple Trials

Several challenges hamper the success of developing and implementing personalized treatment decisions in clinics. First, recent machine learning methods (Zhao et al., 2012, 2015) for discovering individualized treatment rules (ITRs) lack interpretability, and thus, they are difficult to understand by clinicians and translate into clinical practice. Second, most randomized controlled trials (RCTs) are powered to detect average treatment effects instead of subgroup effects, let alone optimal individualized treatment decisions. Thus, subgroup or ITR findings are difficult to replicate due to small sample sizes. Third, the target population for the application of ITRs can be different from the patient sample used in estimating the ITR due to time, geographic, or other differences (Justice et al., 1999). When learning an ITR based on a single study, the research aim is to estimate an ITR that performs best on the current study, which may not be able to transport or generalize

to a future sample due to sample difference between studies and the study-specific noise variables.

A cost-effective method to remedy the small sample size problem and improve the reproducibility and transportability is to pool and analyze data from multiple RCTs, which includes meta-analysis (Haidich, 2010) and integrative analysis approaches (Ma et al., 2011). Meta-analysis uses a weighted average to compute a pooled estimate of an average overall treatment effect from individual studies (Cipriani et al., 2018; Jakobsen et al., 2017). Subgroup analyses are exploratory in nature, and well-designed studies testing the same subgroup effect is scarce in the literature. Therefore, conducting subgroup meta-analysis across multiple trials is difficult. For example, we performed a systematic review of subgroup analysis of RCTs of major depressive disorder (MDD). There were 211 studies that met our inclusion criteria, but with only one consensus predictive variable (i.e., baseline severity of depression), suggesting that the literature is incomplete and inconclusive given the substantial observed heterogeneity in treatment effects (HTE). On the other hand, we didn't find any research performing meta-analysis for ITRs because there is no straightforward method to average ITRs from individual studies given that an ITR is usually the sign of a decision function. Integrative analysis approaches pool raw data from multiple studies and analyze the pooled data as if it is from a single trial. Integrative analysis can be more effective than meta-analysis (Ma et al., 2011), and has been used in detecting genetic risk factors from multiple cancer studies or cancer subtypes. However, it requires that multiple studies share the same biological ground and the same candidate feature variables should be collected from all the studies.

In practice, due to different study designs and hypotheses, RCTs often collect different sets of baseline covariates or feature variables. When new evidence from neuroscience research emerges, new hypotheses are proposed regarding various biomarkers as predictive/prescriptive variables for pharmacotherapy of MDD (Trivedi et al., 2016). Thus, biomarkers are one of the foci of data collec-

tion, in addition to clinical and neuropsychiatric measures (EMBARC, Trivedi et al. (2016)), while in prior studies (e.g, STAR*D (Rush et al., 2004), CO-MED (Rush et al., 2011), HEAL (Shear et al., 2016), REVAMP (Kocsis et al., 2009), Nefazodone-CBASP (Keller et al., 2000), Bulimia (Sysko et al., 2010)), only comprehensive clinical variables are available. In particular, EMBARC trial collects comprehensive baseline measures including clinical measures, depression and anxiety measures, behavioral phenotyping (BP), functional magnetic resonance imaging (fMRI), electroencephalogram (EEG), and diffusion tensor imaging (DTI). In addition to clinical measures, depression and anxiety measures, STAR*D and Co-Med also include information for measuring quality of life, work and social adjustment, while HEAL consists of additional variables to measure grief symptoms and treatment expectancy.

Non-uniform, heterogeneous feature collection poses challenges for the integration of ITRs or meta-analysis across RCTs. Combining data directly and performing a single analysis is often inefficient or inappropriate. For example, one may consider to use all subjects with common feature variables across trials and treat it as a single study, which may lose information of important feature variables. On the other hand, to include as many feature variables as possible, one may end up with a small number of subjects in the analysis, which may lead to a biased sample to represent the entire population.

Because available feature variables in each trial differ, treatment rules estimated from each trial will include different "resolutions" of the patient-specific characteristics (i.e., ITRs from EMBARC may depend on both clinical variables and brain imaging biomarkers, whereas ITRs from STAR*D depend only on clinical variables). The fitted ITRs may yield opposite treatment recommendations for the same patient based on different feature variables included. A conundrum is that although ITRs learned with a rich set of feature variables and finer resolution may prescribe more specific

treatment for an individual patient, these ITRs can be less reliable due to an increased number of tailoring variables. Conversely, coarsened ITRs learned using fewer variables are more robust and practical in resource-limited clinical settings (e.g., when collecting data on expensive biomarkers is prohibitive). However, these ITRs may not lead to a truly optimal personalized treatment. Therefore, integration and reconciliation of the ITRs learned on heterogeneous scales are important.

We will propose novel analytic solutions to the above challenges. Our proposed method is related to Multi-Task Learning (MTL) and Multi-Objective Reinforcement Learning (MORL). One type of MTL (Ruder, 2017) is to optimize one loss function with related auxiliary tasks to improve the main task. MORL (Liu et al., 2015) aims to optimize multiple objectives by summarizing them into one single objective. When multiple studies collecting features with different resolutions are available, depending on the main study of interest and information available in the future target population of applying the estimated ITR, other studies can be included as auxiliary data sets to improve the efficiency and reproducibility of the ITR comparing to using the main study data set alone. When learning a high-resolution ITR is of interest, the auxiliary data sets often collect a subset of feature variables which can provide low-resolution ITRs. If a simple and easy to interpret low-resolution ITR is of interest, auxiliary data sets with high-resolution ITRs can still assist improving the coarsened ITR based on the low-resolution information.

To improve efficiency and reproducibility of ITRs from both directions, we propose a novel integrative learning to synthesize evidence across trials and provide an integrative ITR. Our method does not require all studies to collect common sets of variables. Thus, the integrative learning allows evidence to be combined from ITRs identified in recent RCTs that collected emerging biomarkers (e.g., neuroimaging measures) with earlier RCTs that focused on clinical and psychosocial markers. The proposed method will summarize information across different trials by a regularized value func-

tion and use a data-driven method to determine how much evidence each study contributes to the integrative ITR. Optimization of the regularized value function can be easily solved using existing outcome-weighted learning methods (Zhao et al., 2012; Qiu et al., 2017) with an augmentation term related to the clinical outcome.

In Chapter 4 of this dissertation, we first introduce the rationale and algorithm of the proposed integrative machine learning method for learning both high-resolution ITRs and coarsened ITRs. We then extend the proposed method to studies collecting blockwise feature domains. We derive the underlying Bayesian rules for the proposed method. We show that the proposed method can successfully improve the efficiency and reproducibility of the estimated ITRs compared to existing machine learning methods for single studies (e.g., ABLO) via extensive simulation studies. Lastly, we fit ITRs using the proposed integrative learning method on Establishing Moderators and Biosignatures of Antidepressant Response in clinical Care (EMBARC) trial for treating major depressive disorder.

Chapter 2

Estimation and Evaluation of Linear Individualized Treatment Rules to Guarantee Performance

2.1 Overview

In this chapter, we propose a machine learning method, asymptotically best linear O-learning (ABLO) to estimate the optimal linear ITR. We also propose a diagnostic measure to evaluate candidate ITRs. In Section 2.2, we propose the statistical method of ABLO and several tests for goodness-of-fit. In Section 2.3, we show the asymptotic properties. In Section 2.4, we conduct simulation studies to investigate the performance of the proposed method. In Section 2.5, we apply the method to a study of patients with major depressive disorder, the STAR*D data. Finally, we summarize our findings and discuss possible extensions in Section 2.6.

2.2 Methodologies

We start by introducing some notations for single-stage randomized clinical trials. Let R denote a continuous variable measuring clinical response after treatment (e.g., reduction of depressive symptoms). Without loss of generality, assume a large value of R is desirable. Let $\boldsymbol{X}$ denote a vector of subject-specific baseline feature variables, and let $A = 1$ or $A = -1$ denote two alternative treatments being compared. Assume that we observe $(A_i, \boldsymbol{X}_i, R_i)$ for the ith subject in a two-arm randomized trial with randomization probability $P(A_i = a|\boldsymbol{X}_i = \boldsymbol{x}) = \pi(a|\boldsymbol{x})$, for $i = 1, ..., n$.

An ITR, denoted as $\mathcal{D}(\boldsymbol{X})$, is a binary decision function that maps $\boldsymbol{X}$ into the treatment domain $A = \{-1, 1\}$. Let $P^{\mathcal{D}}$ denote the distribution of $(A, \boldsymbol{X}, R)$ in which $\mathcal{D}$ is used to assign treatments. The value function of $\mathcal{D}$ satisfies

$$V(\mathcal{D}) = E^{\mathcal{D}}(R) = \int R \, dP^{\mathcal{D}} = \int R \, \frac{dP^{\mathcal{D}}}{dP} dP = E\left\{ \frac{RI(A = \mathcal{D}(\boldsymbol{X}))}{\pi(A|\boldsymbol{X})} \right\}, \tag{2.1}$$

where P is the distribution of $(A, \boldsymbol{X}, R)$ and $P^{\mathcal{D}}$ is the distribution under $A = \mathcal{D}(\boldsymbol{X})$. In most applications, $\mathcal{D}(\boldsymbol{X})$ is determined by the sign of a function, $f(\boldsymbol{X})$, which is referred to as the ITR decision function. That is, $\mathcal{D}(\boldsymbol{X}) = \text{sign}(f(\boldsymbol{X}))$. In general settings, $f \in \mathcal{F}$ can take any form, either a parametric function or a non-parametric function. To quantify the benefit of an ITR, a measure related to the value function is a natural choice. The mean difference is widely used to compare the average effect of two treatments. Analogously, we define the benefit function corresponding to an ITR as the difference in the value function between two complementary strategies: one that assigns treatments according to $\mathcal{D}(\boldsymbol{X})$ and the other assigns according to the complementary rule $-\mathcal{D}(\boldsymbol{X})$ for any given feature variables $\boldsymbol{X}$. That is, the benefit function for $\mathcal{D}(\boldsymbol{X}) = \text{sign}(f(\boldsymbol{X}))$ is

$$\delta(f(\boldsymbol{X})) = E\Big\{ R|A = \text{sign}(f(\boldsymbol{X})), \boldsymbol{X} \Big\} - E\Big\{ R|A = -\text{sign}(f(\boldsymbol{X})), \boldsymbol{X} \Big\}. \tag{2.2}$$

2.2.1 Estimating Optimal Linear Treatment Rule

To obtain a practically useful and transparent ITR, we consider a class of linear ITR decision functions, denoted by $\mathcal{L}$, and estimate the optimal linear function $f_L^* \in \mathcal{L}$, that maximizes the value function (2.1) among this class. To this end, following the original idea of Liu et al. (2014), we note that maximizing $V(\mathcal{D})$ is equivalent to minimizing a residual-weighted misclassification error given as

$$E\left[|R - r(\boldsymbol{X})| \frac{I\left\{A \operatorname{sign}(R - r(\boldsymbol{X})) \neq \mathcal{D}(\boldsymbol{X})\right\}}{\pi(A|\boldsymbol{X})}\right],$$

where $r(\boldsymbol{X})$ is any function of $\boldsymbol{X}$, taken as an approximation to the conditional mean of R given $\boldsymbol{X}$. Thus, we aim to minimize the empirical version of the above quantity, given as

$$\frac{1}{n}\sum_i \frac{|W_i| I(A_i Z_i \neq \mathcal{D}(\boldsymbol{X}_i))}{\pi(A_i|\boldsymbol{X}_i)} = \frac{1}{n}\sum_i \frac{|W_i| I(A_i Z_i f(\boldsymbol{X}_i) < 0)}{\pi(A_i|\boldsymbol{X}_i)}$$

for $f \in \mathcal{L}$, where $W_i = R_i - \widehat{r}(\boldsymbol{X}_i)$, $Z_i = \operatorname{sign}(W_i)$, and $\widehat{r}(\boldsymbol{X})$ is obtained from a working model by regressing R_i on $\boldsymbol{X}_i$ (Liu et al., 2014).

The above optimization with zero-one loss is a non-deterministic polynomial-time hard (NP-hard) problem (Natarajan, 1995). To avoid this computational challenge, the zero-one loss was replaced by some convex surrogate loss in existing methods, for instance, the squared loss or hinge loss. Let f^* denote the global optimal decision function corresponding to the optimal treatment rule among any decision functions. That is, $f^*(\boldsymbol{X}) = E(R|A = 1, \boldsymbol{X}) - E(R|A = -1, \boldsymbol{X})$. When $\mathcal{L}$ consists of linear decision functions that are far from the global optimal rule such that $f^* \notin \mathcal{L}$, estimating optimal linear rule by minimizing the surrogate loss (e.g., hinge loss or squared loss) no longer guarantees that the induced value or benefit is maximized among the linear class.

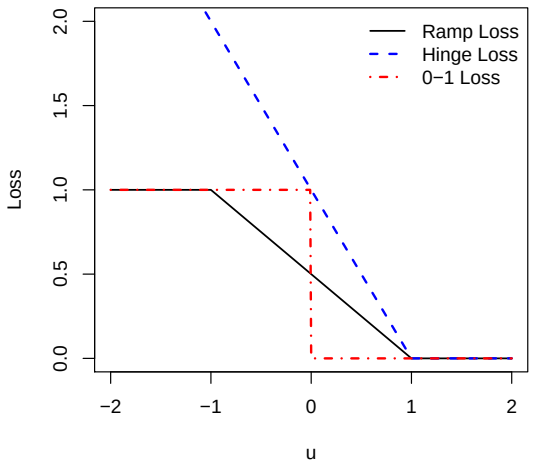
In order to obtain the best linear ITR with guaranteed performance, we propose to use an authentic approximation loss that will converge to zero-one loss, referred to as the ramp loss

(McAllester and Keshet, 2011; Huang and Fong, 2014), for value maximization. The ramp loss, as plotted in Figure 2.1, has been used in the machine learning literature to provide a tight bound on the misclassification rate (McAllester and Keshet, 2011; Collobert et al., 2006). Mathematically, this function can be expressed as

$$h_s(u) = I(u \leq -\frac{s}{2}) - \frac{1}{s}(u - \frac{s}{2})I(-\frac{s}{2} < u < \frac{s}{2}), \tag{2.3}$$

where s is a tuning parameter to be chosen in a data-adaptive fashion. Clearly, when s converges to zero, the ramp loss function converges to the zero-one loss; thus, we expect that the estimated rule from this loss function should approximately maximize the value function among class $\mathcal{L}$.

Figure 2.1: Different approximation functions of the zero-one loss



Specifically, with the ramp loss (2.3), we propose to estimate the optimal linear ITR decision function, $f_L^*(\boldsymbol{X})$, by minimizing the penalized weighted sum of ramp loss of a linear decision function $f(\boldsymbol{X}) = \beta_0 + \boldsymbol{X}^T\boldsymbol{\beta}$,

$$L(f) = C\sum_{i=1}^{n} \frac{|W_i| h_s(Z_i A_i f(\boldsymbol{X}_i))}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2, \tag{2.4}$$

where C is the cost parameter, which is a tuning parameter that determines the penalty placed on misclassifying a subject's optimal treatment. Because the ramp loss is not convex, we solve the optimization by the difference of convex functions algorithm (DCA) (An et al., 1996). First, we express $h_s(u)$ as the difference of two convex functions. That is,

$$h_s(u) = h_{1,s}(u) - h_{2,s}(u) = (\frac{1}{2} - \frac{u}{s})_+ - (-\frac{1}{2} - \frac{u}{s})_+,$$

where function $(x)_+$ denotes the positive part of x. Let η_i denote $Z_i A_i f(\boldsymbol{X}_i)$. Then the penalized weighted sum of ramp loss can be simplified as $L = \sum_{i=1}^n C\frac{|W_i|h_s(\eta_i)}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2$, and the minimization in (2.4) can be carried out in three steps:

- Step 1: Start with an initial value of $\boldsymbol{\beta}$, i.e. $\boldsymbol{\beta}_0$, which can be derived from the linear rule estimated by the O-learning with hinge loss. Then, the initial value of $\boldsymbol{\eta}$ can be calculated and we denote it as $\boldsymbol{\eta}_0$.

- Step 2: Solve

$$\widehat{\boldsymbol{\beta}} = \arg\min \sum_{i=1}^n C\frac{|W_i|\{h_{1,s}(\eta_i) - \widehat{h}_{2,s}(\eta_i, \eta_i^0)\}}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2, \tag{2.5}$$

  where $\widehat{h}_{2,s}(\eta_i, \eta_i^0) = h_{2,s}(\eta_i^0) + h'_{2,s}(\eta_i^0)\eta_i$ and $h'_{2,s}(u) = \dfrac{-I(u/s < -1/2)}{s}$.

- Step 3: Compute $\boldsymbol{\eta}^0$ and update it in step 2 until the change in L is less than a pre-specified threshold.

In order to solve the optimization problem in Step 2, we introduce slack variables ξ_i to replace $h_{1,s}(\eta_i)$. Therefore, (2.5) is equivalent to minimize

$$\sum_{i=1}^n C\frac{|W_i|\{\xi_i - h'_{2,s}(\eta_i^0)\eta_i\}}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2, \ s.t. \ \ \xi_i \geq \frac{1}{2} - \frac{\eta_i}{s}, \text{ and } \ \xi_i \geq 0.$$

By adding two non-negative Lagrange multipliers $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$, we obtain

$$L = \sum_{i=1}^{n} C\frac{|W_i|\{\xi_i - h'_{2,s}(\eta_i^0)\eta_i\}}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2 - \sum_{i=1}^{n}\alpha_i(\xi_i + \frac{\eta_i}{s} - \frac{1}{2}) - \sum_{i=1}^{n}\tau_i\xi_i.$$

Let $\boldsymbol{\gamma}$ be a vector with i-element $\gamma_i = \frac{|W_i|h'_{2,s}(\eta_i^0)}{\pi(A_i|\boldsymbol{X}_i)}$. Notice that $\eta_i = A_iZ_i(\beta_0 + \boldsymbol{X}_i^T\boldsymbol{\beta})$, and take derivative with regard to β_0, $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, we obtain the following equations

$$0 = \sum_{i=1}^{n} A_iZ_i\left(C\gamma_i + \frac{\alpha_i}{s}\right), \tag{2.6}$$

$$\boldsymbol{\beta} = \sum_{i=1}^{n} C\frac{|W_i|h'_{2,s}(\eta_i^0)A_iZ_i\boldsymbol{X}_i}{\pi(A_i|\boldsymbol{X}_i)} + \sum_{i=1}^{n}\alpha_iA_iZ_i\boldsymbol{X}_i/s = \sum_{i=1}^{n} A_iZ_i\left(C\gamma_i + \frac{\alpha_i}{s}\right)\boldsymbol{X}_i, \tag{2.7}$$

$$0 = \frac{|W_i|}{\pi(A_i|\boldsymbol{X}_i)}C - \alpha_i - \tau_i. \tag{2.8}$$

By (2.8), $\xi_i's$ cancel out and the penalized weighted sum of ramp loss becomes

$$\begin{aligned}
L &= -\sum_{i=1}^{n} C\gamma_i\left\{A_iZ_i(\beta_0 + \boldsymbol{X}_i^T\boldsymbol{\beta})\right\} + \frac{1}{2}||\boldsymbol{\beta}||^2 - \sum_{i=1}^{n}\frac{\alpha_i}{s}\left\{A_iZ_i(\beta_0 + \boldsymbol{X}_i^T\boldsymbol{\beta})\right\} + \frac{1}{2}\sum_{i=1}^{n}\alpha_i \\
&= -\sum_{i=1}^{n} A_iZ_i\left(C\gamma_i + \frac{\alpha_i}{s}\right)\boldsymbol{X}_i^T\boldsymbol{\beta} + \frac{1}{2}\sum_{i=1}^{n}\alpha_i + \frac{1}{2}||\boldsymbol{\beta}||^2 \qquad \text{by (2.6)} \\
&= -\frac{1}{2}||\boldsymbol{\beta}||^2 + \frac{1}{2}\sum_{i=1}^{n}\alpha_i \qquad \text{by (2.7)} \\
&\propto \frac{1}{2}\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\left(\sum A_iZ_i\frac{\alpha_i}{s}\boldsymbol{X}_i^T\sum A_iZ_i\frac{\alpha_i}{s}\boldsymbol{X}_i + 2\sum A_iZ_iC\gamma_i\boldsymbol{X}_i^T\sum A_iZ_i\frac{\alpha_i}{s}\boldsymbol{X}_i\right) \\
&= -\frac{1}{2s^2}\boldsymbol{\alpha}^T\boldsymbol{Q}\boldsymbol{\alpha} + \frac{1}{2}\left(\mathbf{1} - 2C\boldsymbol{Q}\boldsymbol{\gamma}/s\right)\boldsymbol{\alpha},
\end{aligned}$$

where $\mathbf{Q}$ is a square matrix where $Q_{i,j} =< A_iZ_i\boldsymbol{X}_i, A_jZ_j\boldsymbol{X}_j >$.

Hence, the dual problem is

$$\min \quad \frac{1}{2s^2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} - \frac{1}{2}\left(\mathbf{1} - 2C\mathbf{Q}\boldsymbol{\gamma}/s\right)^T\boldsymbol{\alpha}, \tag{2.9}$$

subject to $0 \le \alpha_i \le C|W_i|/\pi(A_i|\boldsymbol{X}_i)$ and $\sum CA_iZ_i\gamma_i + \sum A_iZ_i\alpha_i/s = 0$. Thus, the optimization problem can be solved via quadratic programming. After obtaining α_i, the original coefficient can

be derived by $\widehat{\boldsymbol{\beta}} = \sum A_i Z_i \left(C\boldsymbol{\gamma}_i + \frac{\alpha_i}{s}\right) \boldsymbol{X}_i$. Based on the KarushKuhnTucker (KKT) condition $\xi_i(C|W_i|/\pi(A_i|\boldsymbol{X}_i) - \alpha_i) = 0$, when $0 < \alpha_i < C|W_i|/\pi(A_i|\boldsymbol{X}_i)$, we have $\xi_i = 0$ and $A_i Z_i(\widehat{\beta}_0 + \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}) - \frac{1}{2}s = 0$. The intercept term $\widehat{\beta}_0$ can be calculated by taking the average of $\dfrac{s}{2A_iZ_i} - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}$.

Therefore, we obtain the optimal linear ITR as

$$\widehat{f}_L^*(\boldsymbol{X}) = \widehat{\beta}_0 + \boldsymbol{X}^T\widehat{\boldsymbol{\beta}},$$

and denote the optimal ITR as $\text{sign}(\widehat{f}_L^*(\boldsymbol{X}))$. In Section 2.3, we show that $\widehat{f}_L^*$ converges to the true best linear rule, f_L^*, asymptotically, at a slower rate than the usual root-n rate. We refer the proposed estimation procedure as Asymptotically Best Linear O-learning, ABLO. We also prove the asymptotic normality of $\widehat{\boldsymbol{\beta}}$ and the estimated benefit function, which provides justification of the inference procedures proposed in Section 2.2.2 and 2.2.3.

2.2.2 Performance Diagnostics for the Estimated ITR

ABLO guarantees that the optimal value among the class $\mathcal{L}$ is achieved asymptotically. Nevertheless, the optimal linear rule $f_L^*(\boldsymbol{X})$ may still be far from the global optimal, f^*, such that for some important subgroups, $f_L^*(\boldsymbol{X})$ may be non-optimal or even worse than the complementary treatment rule. Therefore, an empirical measure must be constructed to evaluate the performance of an estimated ITR.

To develop a practically feasible diagnostic method for any estimated ITR, given by $\text{sign}(\widehat{f}(\boldsymbol{X}))$, we note that if $\widehat{f}(\boldsymbol{X})$ is truly optimal among any decision functions in $\mathcal{F}$, i.e., $\widehat{f}(\boldsymbol{X})$ has the same sign as $f^*(\boldsymbol{X})$, then for any subgroup defined by $\boldsymbol{X} \in \mathcal{C}$ for a given set $\mathcal{C}$ (e.g., $\mathcal{C}$ can represent the subset of mildly depressed patients with QIDs score less than 11) in the domain of $\boldsymbol{X}$, the value function for those subjects whose treatments are the same as $\text{sign}(\widehat{f}(\boldsymbol{X}))$ should always be larger

than or equal to the value function for those subjects with the same $\boldsymbol{X} \in \mathcal{C}$, but whose treatments are opposite to $\text{sign}(\widehat{f}(\boldsymbol{X}))$. This is because

$$\begin{aligned} & E\left[\frac{RI\left\{A=\text{sign}(\widehat{f}(\boldsymbol{X}))\right\}}{\pi(A|\boldsymbol{X})}\Big|\boldsymbol{X}\right] - E\left[\frac{RI\left\{A=-\text{sign}(\widehat{f}(\boldsymbol{X}))\right\}}{\pi(A|\boldsymbol{X})}\Big|\boldsymbol{X}\right] \\ = \;& I(f^*(\boldsymbol{X})>0)E(R|A=1,\boldsymbol{X}) + I(f^*(\boldsymbol{X})\le 0)E(R|A=-1,\boldsymbol{X}) \\ & -I(f^*(\boldsymbol{X})>0)E(R|A=-1,\boldsymbol{X}) - I(f^*(\boldsymbol{X})\le 0)E(R|A=1,\boldsymbol{X}) \\ = \;& |f^*(\boldsymbol{X})| \\ \ge \;& 0. \end{aligned}$$

It then follows that the group-average benefit for $\widehat{f}$, defined as

$$\delta_{\mathcal{C}}(\widehat{f}) \equiv E\left[\frac{RI\left\{A=\text{sign}(\widehat{f}(\boldsymbol{X}))\right\}}{\pi(A|\boldsymbol{X})}\Big|\boldsymbol{X}\in\mathcal{C}\right] - E\left[\frac{RI\left\{A=-\text{sign}(\widehat{f}(\boldsymbol{X}))\right\}}{\pi(A|\boldsymbol{X})}\Big|\boldsymbol{X}\in\mathcal{C}\right],$$

should be non-negative. On the other hand, if $\delta_{\mathcal{C}}(\widehat{f}) \ge 0$ holds for any subset $\mathcal{C}$, then the above derivation also indicates that $\widehat{f}(\boldsymbol{X})$ must have the same sign as $f^*(\boldsymbol{X})$, i.e., $\widehat{f}(\boldsymbol{X})$ is the optimal treatment rule for subjects in $\mathcal{C}$.

These observations suggest a diagnostic measure $\delta_{\mathcal{C}}(\widehat{f})$ for any subgroup $\mathcal{C}$. Specifically, we propose an empirical ITR diagnostic measure as

$$\begin{aligned} \widehat{\delta}_{\mathcal{C}}(\widehat{f}) \;=\; & \frac{\sum_{i=1}^n I\left\{\boldsymbol{X}_i\in\mathcal{C}, A_i=\text{sign}(\widehat{f}(\boldsymbol{X}_i))\right\} R_i/\pi(A_i,\boldsymbol{X}_i)}{\sum_{i=1}^n I(\boldsymbol{X}_i\in\mathcal{C})} \\ & -\frac{\sum_{i=1}^n I\left\{\boldsymbol{X}_i\in\mathcal{C}, A_i\ne\text{sign}(\widehat{f}(\boldsymbol{X}_i))\right\} R_i/\pi(A_i,\boldsymbol{X}_i)}{\sum_{i=1}^n I(\boldsymbol{X}_i\in\mathcal{C})}. \end{aligned} \tag{2.10}$$

Because $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$ approximates $\delta_{\mathcal{C}}(\widehat{f})$, the measure $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$ is expected to be positive with a high probability if $\widehat{f}(\boldsymbol{X})$ is close to the global true optimal. Furthermore, the evidence that $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$ is positive for a rich class of subsets $\mathcal{C}$ will support the approximate optimality of $\widehat{f}$ in the class. However, because it is infeasible to exhaust all subgroups, we suggest a class of pre-specified subgroups

$\mathcal{C}_1, ..., \mathcal{C}_m$ and calculate the corresponding $\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}), ..., \widehat{\delta}_{\mathcal{C}_m}(\widehat{f})$. An aggregated diagnostic measure is $\widehat{\Delta}(\widehat{f}) = \min\left\{\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}), ..., \widehat{\delta}_{\mathcal{C}_m}(\widehat{f})\right\}$. A positive value of $\widehat{\Delta}(\widehat{f})$ implies approximate optimality of $\widehat{f}$ when m is large enough. In practice, we consider $\mathcal{C}_k$ to be pre-specified groups or the sets determined by the tertiles of each component of $\boldsymbol{X}$, for example, the jth component of $\boldsymbol{X}$ below its first tertile, between the first and the second tertiles, or above the second tertile. Moreover, using the proposed diagnostic measure, by examining the subsets $\mathcal{C}$ (or tertiles defined by variables) with negative or close to zero values of $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$, we can identify subgroups or components of $\boldsymbol{X}$ for which the estimated rule $\widehat{f}$ may not be sufficiently optimal. Thus, we can further improve the rule estimation in this subgroup to obtain an improved ITR.

If the same data are used for estimating the optimal ITR and performing diagnostics, the latter may not be an honest measure of performance (Athey and Imbens, 2016). Thus, we suggest the following sample-splitting scheme. Divide the data into K folds, and denote $\widehat{f}^{(-k)}$ as the optimal ITR obtained using data without the kth-fold. Next, benefit of each $\widehat{f}^{(-k)}$ is calibrated on the kth-fold data using the diagnostic measure and then averaged. Let n_k denote the sample size of the kth-fold, and let I_k index subjects in this fold. The honest diagnostic measure for subgroup $\mathcal{C}$ is estimated by $\widehat{\delta}_{\mathcal{C}}(\widehat{f}) = \frac{1}{K}\sum_{k=1}^{K}\widehat{\delta}_{\mathcal{C}}^{(k)}$, where

$$\widehat{\delta}_{\mathcal{C}}^{(k)} = \frac{1}{n_k}\sum_{\{i:i\in I_k\}}\left[I\left\{A_i = \text{sign}(\widehat{f}^{(-k)}(\boldsymbol{X}_i))\right\} - I\left\{A_i = -\text{sign}(\widehat{f}^{(-k)}(\boldsymbol{X}_i))\right\}\right] R_i/\pi(A_i|\boldsymbol{X}_i).$$

We will implement this scheme in subsequent analysis.

2.2.3 Inference Using the Diagnostic Measure

The proposed diagnostic measure, $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$, can be used to compare different ITRs and non-personalized rules, make comparisons within certain subgroups, and assess heterogeneity of ITR benefit (HTB)

across subgroups. Hypotheses of interest may include:

- Test significance of the optimal linear rule compared to the non-personalized rule in the overall sample, i.e.,

$$H_0 : \delta(f_L^*) - \delta_0 = 0 \text{ v.s. } H_1 : \delta(f_L^*) - \delta_0 > 0,$$

  where δ_0 is the average treatment effect of a non-personalized rule (difference in the mean response between treatment groups). For this purpose, we can construct the test statistic based on $\widehat{\delta}_{\mathcal{C}}(\widehat{f}) - \delta_0$, where $\widehat{f}$ is obtained from any method, and $\mathcal{C}$ is the whole population. We reject the null hypothesis at a significance level of α if the $(1-\alpha)$-confidence interval with ∞ as the upper bound for $\widehat{\delta}_{\mathcal{C}}(\widehat{f}) - \delta_0$ does not contain 0.

- Test significance of the optimal linear rule compared to the non-personalized rule in a subgroup k, i.e.,

$$H_0 : \delta_{\mathcal{C}_k}(f_L^*) - \delta_{0k} = 0 \text{ v.s. } H_1 : \delta_{\mathcal{C}_k}(f_L^*) - \delta_{0k} > 0,$$

  where δ_{0k} is the average treatment effect in the subgroup. The same test statistic as the previous one can be used but with $\mathcal{C} = \mathcal{C}_k$.

- Test the HTB across subgroups $\{\mathcal{C}_1, \cdots, \mathcal{C}_K\}$, i.e.,

$$H_0 : \delta_{\mathcal{C}_k}(f_L^*) - \delta_{\mathcal{C}_K}(f_L^*) = 0, k = 1, \cdots, K-1.$$

  We propose the HTB test statistic

$$T = \widehat{\mathbf{\Delta}}_{\mathcal{C}}^T \{\text{cov}(\widehat{\mathbf{\Delta}}_{\mathcal{C}})\}^{-1} \widehat{\mathbf{\Delta}}_{\mathcal{C}},$$

  where $\widehat{\mathbf{\Delta}}_{\mathcal{C}}^T = (\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) - \widehat{\delta}_{\mathcal{C}_K}(\widehat{f}), \cdots, \widehat{\delta}_{\mathcal{C}_{K-1}}(\widehat{f}) - \widehat{\delta}_{\mathcal{C}_K}(\widehat{f}))$. It can be shown that T asymptotically follows χ^2_{K-1} under H_0, so we reject H_0 when T is larger than the $(1-\alpha)$-quantile of χ^2_{K-1}.

- Test the non-optimality of the best linear rule f_L^* in a subgroup $\mathcal{C}$ by evaluating

$$H_0 : \delta_{\mathcal{C}}(f_L^*) \geq 0 \text{ v.s. } H_1 : \delta_{\mathcal{C}}(f_L^*) < 0.$$

  For this purpose, we can directly use $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$ and reject the null hypothesis if the confidence interval with lower bound of $-\infty$ does not contain zero.

The asymptotic properties of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$ are required to perform inference above. Based on the theoretical properties (asymptotic normality) given in Section 2.3, we propose a bootstrap method to compute confidence interval for the diagnostic measure. We denote the bth bootstrap sample as $(\tilde{A}_i^{(b)}, \tilde{X}_i^{(b)}, \tilde{R}_i^{(b)})$, where $i = 1, 2, \cdots, n$, and re-estimate residuals as $\tilde{W}_i^{(b)}$ in (2.5). Next, we re-fit treatment rule $\tilde{f}^{(b)}$ and obtain $\tilde{\delta}_{\mathcal{C}}^{(b)}(\tilde{f}^{(b)})$. The 95% confidence interval for $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$ is constructed from the empirical quantiles of $\tilde{\delta}_{\mathcal{C}}^{(b)}(\tilde{f}^{(b)})$, $b = 1, 2, \cdots, B$.

2.3 Asymptotic Properties

Let $\boldsymbol{X}$ denote a vector with one as the first component and the remaining components as feature variables. To emphasize that the tuning parameter s of ramp loss may depend on the sample size to establish asymptotic properties, we denote it by s_n in this section. We assume

(a) The true optimal linear function, $f_L^*(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}^*$, is the unique minimizer of $E\{RI(Af(\boldsymbol{X}) < 0)\}$ for $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}$ where $\|\boldsymbol{\beta}\| = 1$. Furthermore, there exists a positive constant δ_0 such that $P(|\boldsymbol{X}^T\boldsymbol{\beta}^*| > \delta_0) = 1$.

(b) The joint densities of $(R, \boldsymbol{X})$ given $A = 1$ and -1 are twice-continuously differentiable.

(c) There exits a function $r(\boldsymbol{x})$ such that $\{\widehat{r}(\boldsymbol{x}) - r(\boldsymbol{x})\} = o((ns_n)^{-1/2})$ uniformly in $\boldsymbol{x}$, where $\widehat{r}(\boldsymbol{x})$ is estimated from a working regression model of R on $\boldsymbol{X}$.

(d) $(nC_n)^{-1} \to 0$, $ns_n \to \infty$, $ns_n^3 \to 0$, and $(ns_n)^{1/2}(nC_n)^{-1} \to 0$.

(e) There exits a unique minimizer, denoted by $\boldsymbol{\beta}_n$, that minimizes

$$E\left[|R - r(\boldsymbol{X})|h_s\left\{A \text{ sign}(R - r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}\right\}/\pi(A|\boldsymbol{X})\right].$$

Assume that $\boldsymbol{\beta}_n$ belongs to a bounded set. Furthermore, let

$$IF_n(R, \boldsymbol{X}, A) = \left[\frac{\partial}{\partial\beta}E\left\{A(R - r(\boldsymbol{X}))\boldsymbol{X}/\pi(A|\boldsymbol{X})|Z(\boldsymbol{\beta}) = 0\right\}f_{Z(\boldsymbol{\beta})}(0)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_n}\right]^{-1}$$

$$\times\left[|R - r(\boldsymbol{X})|A \text{ sign}(R - r(\boldsymbol{X}))\boldsymbol{X}(2s_n)^{-1}I(A \text{ sign}(R - r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}_n \in [-s_n/2, s_n/2])/\pi(A|\boldsymbol{X})\right],$$

we assume that $s_n^{1/2}IF_n(R, \boldsymbol{X}, A)$ has a bounded third moment and converges to a random variable in $L_2(P)$ norm.

Condition (a) requires a separable boundary condition, but this condition can be further relaxed to allow $\boldsymbol{X}^T\boldsymbol{\beta}^*$ to have positive probability around the boundary and the density vanishes faster than a linear rate when close to the boundary. Condition (c) usually holds if we estimate $r(\boldsymbol{x})$ through some parametric models. In condition (d), s_n and C_n are the tuning parameters to be chosen depending on n, for example, $C_n = 1$ and $s_n = n^{-1/2}$. Condition (e) assumes the convergence of the minimizer associated with the ramp loss. Under these assumptions, we first show the consistency of ABLO, $\widehat{f}_L^*(\boldsymbol{x}) = \boldsymbol{x}^T\widehat{\boldsymbol{\beta}}$. The proof follows the standard M-estimation theory by Van der Vaart (2000). Let $\mathbf{P}_n$ denote the empirical measure, then $\widehat{f}_L^*$ minimizes

$$\mathbf{P}_n\left[|R - \widehat{r}(\boldsymbol{X})|h_s\left\{A \text{ sign}(R - \widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}\right\}/\pi(A|\boldsymbol{X})\right] + (2nC_n)^{-1}\|\boldsymbol{\beta}\|^2.$$

It is clear that from assumptions (a), (b) and (c),

$$\sup_{\boldsymbol{\beta}}\Big|\mathbf{P}_n\left[|R - \widehat{r}(\boldsymbol{X})|h_s\left\{A \text{ sign}(R - \widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}\right\}/\pi(A|\boldsymbol{X})\right]$$

$$+(2nC_n)^{-1}\|\beta\|^2 - E\left[|R - r(\boldsymbol{X})|h_s\left\{A \text{ sign}(R - r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}\right\}/\pi(A|\boldsymbol{X})\right]\Big| \to 0$$

almost surely. By condition (b) and (d), $E\left[|R-r(\boldsymbol{X})|h_s\left\{A\ \text{sign}(R-r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}\right\}/\pi(A|\boldsymbol{X})\right]$ converges uniformly to $E\left[|R-r(\boldsymbol{X})|I\left\{A\ \text{sign}(R-r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}<0\right\}/\pi(A|\boldsymbol{X})\right]$, which is equivalent to

$$E\left[RI(A\boldsymbol{X}^T\boldsymbol{\beta}<0)/\pi(A|\boldsymbol{X})\right]-E[(R-r(\boldsymbol{X}))^-]-r(\boldsymbol{X}).$$

This gives

$$\mathbf{P}_n\left[|R-\widehat{r}(\boldsymbol{X})|h_s\left\{A\ \text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}\right\}/\pi(A|\boldsymbol{X})\right]+(2nC_n)^{-1}\|\boldsymbol{\beta}\|^2$$

$$\to E\left[RI(A\boldsymbol{X}^T\boldsymbol{\beta}<0)/\pi(A|\boldsymbol{X})\right]-E[(R-r(\boldsymbol{X}))^-]-r(\boldsymbol{X})$$

uniformly in $\boldsymbol{\beta}$. Since (a) implies f_L^* is also the unique minimizer of the latter limit for $\|\boldsymbol{\beta}\|=1$, it yields that any convergent subsequence of $\widehat{\boldsymbol{\beta}}$ should converge to a limit proportional to $\boldsymbol{\beta}^*$. Therefore, we conclude that $\widehat{\boldsymbol{\beta}}/\|\widehat{\boldsymbol{\beta}}\|$ converges to $\boldsymbol{\beta}^*$ almost surely. Furthermore, by noting

$$\sup_{\boldsymbol{\beta}}\Big|\mathbf{P}\left[|R-\widehat{r}(\boldsymbol{X})|h_s\left\{A\ \text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}\right\}/\pi(A|\boldsymbol{X})\right]$$

$$-E\left[|R-r(\boldsymbol{X})|h_s\left\{A\ \text{sign}(R-r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}\right\}/\pi(A|\boldsymbol{X})\right]\Big|\to 0,$$

we can easily show that $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n\|$ converges to zero almost surely.

To obtain the asymptotic normality for $\widehat{\boldsymbol{\beta}}$, we follow Koo et al. (2008) by noting $\widehat{\boldsymbol{\beta}}$ solves

$$\mathbf{P}_n\left[|R-\widehat{r}(\boldsymbol{X})|A\ \text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}h_s'\left\{A\ \text{sign}(R-\widehat{r}(\boldsymbol{X})\boldsymbol{X}^T\widehat{\boldsymbol{\beta}}\right\}/\pi(A|\boldsymbol{X})\right]+(nC_n)^{-1}\widehat{\boldsymbol{\beta}}=0.$$

This gives

$$\begin{aligned}
&\sqrt{ns_n}(\mathbf{P}_n-\mathbf{P})\left[|R-\widehat{r}(\boldsymbol{X})|A\ \text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}h_s'\left\{A\ \text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\widehat{\boldsymbol{\beta}}\right\}/\pi(A|\boldsymbol{X})\right]\\
=\ &-(ns_n)^{1/2}(nC_n)^{-1}\widehat{\boldsymbol{\beta}}\\
&-(ns_n)^{1/2}\mathbf{P}\left[|R-\widehat{r}(\boldsymbol{X})|A\ \text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}h_s'\left\{A\ \text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\widehat{\boldsymbol{\beta}}\right\}/\pi(A|\boldsymbol{X})\right]\\
=\ &o(1)-(ns_n)^{1/2}\frac{\partial}{\partial y}\mathbf{P}\left[|R-y|A\ \text{sign}(R-y)\boldsymbol{X}h_s'\left\{A\ \text{sign}(R-y))\boldsymbol{X}^T\widehat{\boldsymbol{\beta}}\right\}\frac{\widehat{r}(\boldsymbol{X})-r(\boldsymbol{X})}{\pi(A|\boldsymbol{X})}\right]\Big|_{y=r(\boldsymbol{X})}\\
&+(ns_n)^{1/2}s_n^{-1}\int_{-s_n/2}^{s_n/2}E\left[A(R-r(\boldsymbol{X}))\boldsymbol{X}/\pi(A|\boldsymbol{X})|Z(\widehat{\boldsymbol{\beta}})=z\right]dF_{Z(\widehat{\boldsymbol{\beta}})}(z),
\end{aligned}$$

where $Z(\boldsymbol{\beta})$ denotes the random variable A sign$(R-r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}$ and $F_{Z(\boldsymbol{\beta})}$ is its cumulative distribution function. From (b) and since $\boldsymbol{\beta}_n$ is the minimizer of the expected ramp loss, the last term is equal to

$$(ns_n)^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n)\frac{\partial}{\partial\boldsymbol{\beta}}E\left[A(R-r(\boldsymbol{X}))\boldsymbol{X}/\pi(A|\boldsymbol{X})|Z(\boldsymbol{\beta})=0\right]f_{Z(\boldsymbol{\beta})}(0)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_n}+o(1).$$

Thus, the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n)$ holds by noting that

$$\sqrt{ns_n}(\mathbf{P}_n-\mathbf{P})\left[|R-\widehat{r}(\boldsymbol{X})|A\text{ sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}h_s'\left\{A\text{ sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\widehat{\boldsymbol{\beta}}\right\}/\pi(A|\boldsymbol{X})\right]$$

is equivalent to

$$\sqrt{ns_n}(\mathbf{P}_n-\mathbf{P})\left[|R-r(\boldsymbol{X})|A\text{ sign}(R-r(\boldsymbol{X}))\boldsymbol{X}\frac{I(A\text{ sign}(R-r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}_n\in[-s_n/2,s_n/2])}{2s_n\pi(A|\boldsymbol{X})}\right]$$

and therefore,

$$\sqrt{ns_n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n)=\sqrt{ns_n}(\mathbf{P}_n-\mathbf{P})IF_n(R,\boldsymbol{X},A)+o_p(1).$$

The asymptotical normality of $\sqrt{ns_n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n)$ follows from condition (e).

Lastly, we examine the diagnostic statistics for any estimated decision function, denoted as $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$ in (2.10), where $\widehat{f}(\boldsymbol{x})=\boldsymbol{x}^T\widehat{\boldsymbol{\beta}}$ is an estimated rule converging to $f^*(\boldsymbol{x})$ uniformly in x. Note that we split the data into K folds, $\widehat{f}^{(-k)}$ is estimated without the kth part of data and $\widehat{\delta}_{\mathcal{C}}^{(k)}$ is computed using the kth part. Let n_k denote the sample size of the kth part of data and let $\mathbf{P}_{n_k}$ denote the empirical measure for the k part of data. Define by

$$\delta_{\mathcal{C}}^*=\frac{E\left[I(\boldsymbol{X}\in\mathcal{C},Af^*(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X})-I(\boldsymbol{X}\in\mathcal{C},Af^*(\boldsymbol{X})<0)R/\pi(A|\boldsymbol{X})\right]}{E[I(\boldsymbol{X}\in\mathcal{C})]}$$

the subgroup benefit based on the optimal linear rule f^*. Since $\boldsymbol{\beta}_n/\|\boldsymbol{\beta}_n\|\to\boldsymbol{\beta}^*$, from condition (a), we have

$$\boldsymbol{\beta}^{*T}\boldsymbol{X}\boldsymbol{\beta}_n^T\boldsymbol{X}/\|\boldsymbol{\beta}_n\|>0$$

with probability one. Therefore,

$$\delta^*_{\mathcal{C}} = \frac{E\left[I(\boldsymbol{X} \in \mathcal{C}, Af_n(\boldsymbol{X}) > 0)R/\pi(A|\boldsymbol{X}) - I(\boldsymbol{X} \in \mathcal{C}, Af_n(\boldsymbol{X}) < 0)R/\pi(A|\boldsymbol{X})\right]}{E[I(\boldsymbol{X} \in \mathcal{C})]},$$

where $f_n(\boldsymbol{X}) = \boldsymbol{\beta}_n^T \boldsymbol{X}$.

Re-express $\widehat{\delta}_{\mathcal{C}}(\widehat{f}^{(-k)})$ as

$$\widehat{\delta}^{(k)}_{\mathcal{C}} = \frac{\mathbf{P}_{n_k} I(\boldsymbol{X} \in \mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X}) > 0)R/\pi(A|\boldsymbol{X})}{\mathbf{P}_{n_k} I(\boldsymbol{X} \in \mathcal{C})} - \frac{\mathbf{P}_{n_k} I(\boldsymbol{X} \in \mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X}) < 0)R/\pi(A|\boldsymbol{X})}{\mathbf{P}_{n_k} I(\boldsymbol{X} \in \mathcal{C})}.$$

Since $\{I(X \in \mathcal{C}) : \mathcal{C} \in \{\mathcal{C}_1, ..., \mathcal{C}_m\}\}$ and $\left\{Af(\boldsymbol{X}) > 0 : f = \boldsymbol{X}^T\boldsymbol{\beta}\right\}$ are VC-major classes,

$$\begin{aligned} & (\mathbf{P}_{n_k} - \mathbf{P}) I(\boldsymbol{X} \in \mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X}) > 0)R/\pi(A|\boldsymbol{X}) \\ = \;& (\mathbf{P}_{n_k} - \mathbf{P}) I(\boldsymbol{X} \in \mathcal{C}, Af^*(\boldsymbol{X}) > 0)R/\pi(A|\boldsymbol{X}) + o_p(n_k^{-1/2}). \end{aligned}$$

We obtain

$$\begin{aligned} & \widehat{\delta}_{\mathcal{C}}(\widehat{f}^{(-k)}) - \delta^*_{\mathcal{C}} \\ = \;& \frac{(\mathbf{P}_{n_k} - \mathbf{P}) I(\boldsymbol{X} \in \mathcal{C}, Af^*(\boldsymbol{X}) > 0)R/\pi(A|\boldsymbol{X})}{\mathbf{P} I(\boldsymbol{X} \in \mathcal{C})} - \frac{(\mathbf{P}_{n_k} - \mathbf{P}) I(\boldsymbol{X} \in \mathcal{C}, Af^*(\boldsymbol{X}) < 0)R/\pi(A|X)}{\mathbf{P} I(\boldsymbol{X} \in \mathcal{C})} \\ - \;& \frac{E\left[I(\boldsymbol{X} \in \mathcal{C}, Af^*(\boldsymbol{X}) > 0)R/\pi(A|\boldsymbol{X}) - I(\boldsymbol{X} \in \mathcal{C}, Af^*(\boldsymbol{X}) < 0)R/\pi(A|\boldsymbol{X})\right]}{E[I(\boldsymbol{X} \in \mathcal{C})]^2} (\mathbf{P}_{n_k} - \mathbf{P}) I(\boldsymbol{X} \in \mathcal{C}) \\ + \;& \frac{E\left[I(\boldsymbol{X} \in \mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X}) > 0)R/\pi(A|\boldsymbol{X}) - I(\boldsymbol{X} \in \mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X}) < 0)R/\pi(A|\boldsymbol{X})\right]}{E[I(X \in \mathcal{C})]^2} \\ & - \frac{E\left[I(\boldsymbol{X} \in \mathcal{C}, Af_n(\boldsymbol{X}) > 0)R/\pi(A|\boldsymbol{X}) - I(\boldsymbol{X} \in \mathcal{C}, Af_n(X) < 0)R/\pi(A|\boldsymbol{X})\right]}{E[I(\boldsymbol{X} \in \mathcal{C})]^2} \\ + \;& o_p(n_k^{-1/2}). \end{aligned}$$

Using the smooth condition in (b) and the expansion for $\widehat{\boldsymbol{\beta}}^{(-k)}$ around $\widehat{\beta}_n$ from the previous asymptotic proof, we can show that the difference in the last two terms has a convergence rate faster than $n_k^{-1/2}$, given $n_k = o(ns_n)$, and furthermore, when $n_k \to \infty$,

$$\sqrt{n_k}\left(\widehat{\delta}^{(k)}_{\mathcal{C}} - \delta^*_{\mathcal{C}}\right) \to_d \mathcal{G}(\mathcal{C}),$$

where $\mathcal{G}(\mathcal{C})$ is a tight Gaussian process indexed by $\mathcal{C}$ with mean zero. After averaging over all folds and assuming K is fixed, similar argument shows that $\sqrt{n}\left(\widehat{\delta}_{\mathcal{C}} - \delta^*_{\mathcal{C}}\right) \to_d \widetilde{\mathcal{G}}(\mathcal{C})$ for some tight Gaussian process $\widetilde{\mathcal{G}}$, where $\widehat{\delta}_{\mathcal{C}} = \frac{1}{K}\sum_k \widehat{\delta}^{(k)}_{\mathcal{C}}$. Note that these results apply to ABLO $\widehat{f}_L$, or other $\widehat{f}$ estimated from minimizing a weighted hinge loss as in O-learning or predictive modeling.

If f^*_L is also the global optimal rule, that is $f^*_L = f^*$, then $\delta^*_{\mathcal{C}} > 0$ for any $\mathcal{C}$ and any $\boldsymbol{X}$. Therefore, the confidence interval for $\delta^*_{\mathcal{C}}$ will be expected to be within $(0, \infty)$ when n is sufficiently large. We can also construct a test for $H_0 : \delta^*_{\mathcal{C}} \geq 0$ vs $H_a : \delta^*_{\mathcal{C}} < 0$ using this asymptotic distribution.

2.4 Simulation Studies

2.4.1 Simulation Design

For all simulation scenarios, we first generated four latent subgroups of subjects based on 10 feature variables $\boldsymbol{X} = (X^1, \cdots, X^{10})$ informative of optimal treatment choice from a pattern mixture model. Treatment $A = 1$ has a greater average effect for subjects in subgroups 1 and 2, and the alternative treatment -1 has a greater average effect in subgroups 3 and 4. Within each subgroup, $\boldsymbol{X}$ were independently simulated from a normal distribution with different means and standard deviation of one. Two settings were considered. In Setting 1, the means of the feature variables for subjects in the four subgroups were $(1, 0.5, -1, -0.5)$, respectively. In Setting 2, the means were $(1, 0.3, -1, -0.3)$. Five noise variables $\boldsymbol{U} = (U^1, \cdots, U^5)$ not contributing to R were independently generated from the standard normal distribution and included in the analyses in order to assess the robustness of each method in the presence of noise features. The treatments for each subject were randomly assigned to 1 or -1 with equal probability, and the number of subjects in each subgroup was equal.

Three additional feature variables W, V, and S were generated to be directly associated with the clinical outcome R. Here, W is an observed prescriptive variable informative of the optimal treatment, V is a prognostic variable predictive of the outcome but not the optimal treatment, and S is an unobserved prescriptive variable not available in the analysis. The clinical outcome for subjects in the kth subgroup was generated by

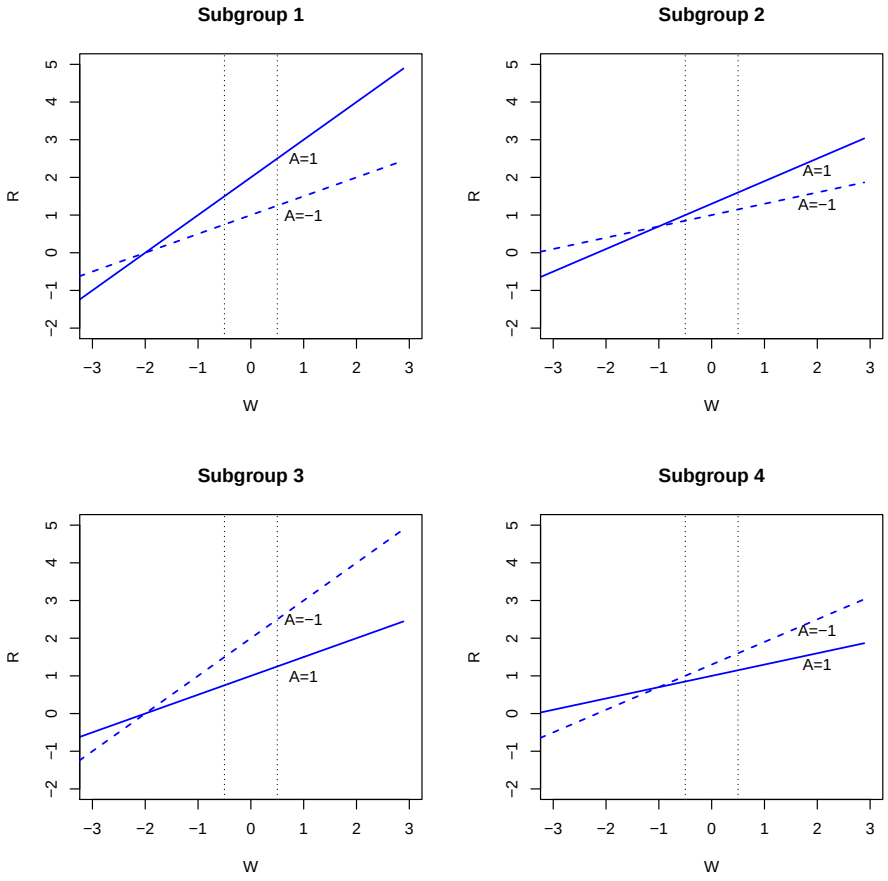
$$R = 1 + I(A = 1)(\delta_{1k} + \alpha_{1k} * W + \beta_{1k} * S) + I(A = -1)(\delta_{2k} + \alpha_{2k} * W + \beta_{2k} * S) + V + e,$$

where $e \sim N(0, 0.25)$, V, W, and S are i.i.d. and follow the standard normal distribution, $\boldsymbol{\delta} = [\delta_{lk}]_{2*4} = \begin{bmatrix} 1 & 0.3 & 0 & 0 \\ 0 & 0 & 1 & 0.3 \end{bmatrix}$, $\boldsymbol{\alpha} = [\alpha_{lk}]_{2*4} = \begin{bmatrix} 1 & 0.6 & 0.5 & 0.3 \\ 0.5 & 0.3 & 1 & 0.6 \end{bmatrix}$, and $\boldsymbol{\beta} = 2\boldsymbol{\alpha}$. Within each group k, there is a qualitative interaction between treatment and W as shown in Figure 2.2.

The benefit function of the theoretical global optimal ITR decision function, denoted as f^*, was computed numerically by simulating the clinical outcome R under both treatment 1 and -1, using all observed feature variables (i.e., $\boldsymbol{X}$, W, and V), and taking the average difference of R under the true optimal and non-optimal treatments using a large independent test set of $N = 100,000$. In practice, this global optimum may not be attained by a linear rule due to the unknown and potentially nonlinear true optimal treatment rule. The theoretical optimal linear rule f_L^* was computed numerically using the observed variables and maximizing the value function in the class of all linear rules under each simulation model (details provided in Appendix Section A.1). The benefit of f_L^* was then computed with a large independent test set of $N = 50,000$.

For each simulated data set, predictive modeling (PM), Q-learning, O-learning, and ABLO were applied to estimate the optimal ITR. For PM, we considered a random forest-based prediction related to the virtual twins approach of Foster et al. (2011). PM first applies random forest on R, including all observed feature variables $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{U}, W, V)$ and treatment assignments. It next

Figure 2.2: Clinical outcome (R) versus W with treatment 1 or -1 in each latent group in the simulation setting. Two vertical dotted lines indicate $W = -0.5$ and $W = 0.5$.



predicts the outcome for the ith subject given $(\boldsymbol{Z}_i, A_i = 1)$ and $(\boldsymbol{Z}_i, A_i = -1)$, denoted as $\widehat{R}_{1i}$ and $\widehat{R}_{-1i}$, respectively. The optimal treatment for the subject is $\text{sign}(\widehat{R}_{1i} - \widehat{R}_{-1i})$. Q-learning was implemented by a linear regression including all the observed feature variables, treatment assignments, and their interactions. Benefit of the estimated optimal ITR under each method and was computed by $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$ in Section 2.2.2.

In the simulations, observed feature variables $\boldsymbol{Z}$ were used in all methods, while the unobserved prescriptive variable S and latent subgroup membership were not included. Linear kernel was used

for O-learning and ABLO. Five-fold cross validation was used to select the tuning parameters C and s. For each method, the optimal treatment selection accuracy and ITR benefit were estimated using two-fold cross validation with equal size of training and testing sets. The training set was used to estimate the ITR and the testing set was used to estimate the ITR benefit and accuracy. Bootstrap was used to estimate the confidence interval of the ITR benefit under the estimated rule. Coverage probabilities were reported to evaluate the performance of the inference procedure. To evaluate performance on subgroups, we partitioned W, V, X^1, and U^1 into three groups based on values in the intervals $(-\infty, -0.5)$, $[-0.5, 0.5]$, or $(0.5, \infty)$. We calculated the HTB test for the candidate variables and tested the difference between the estimated rules and the overall non-personalized rules.

2.4.2 Simulation Results

Results from 500 replicates are summarized in Table 2.1, 2.2, 2.3, Figure 2.3 and 2.4. For both simulation settings, ABLO with linear kernel has the largest optimal treatment selection accuracy regardless of the sample size, and it is also close to the maximal accuracy rate based on the theoretical best linear rule. In addition, ABLO estimates the ITR benefit closest to the true global maximal value of 0.678 on the overall sample, and it is almost identical to the benefit estimated by the theoretical best linear rule when the sample size is large ($N = 800$ training, 800 testing). PM, Q-learning, and O-learning all underestimate the ITR benefit, especially when the sample size is smaller ($N = 400$ training, 400 testing), and thus they do not attain the maximal value of the theoretical optimal linear rule. Based on the empirical standard deviation, we also observe that ABLO is more robust than all other methods. For all methods, as the sample size increases, the treatment selection accuracy increases and the estimated mean benefit is closer to the true

optimal value. Furthermore, the estimated ITR benefit increases as the accuracy rate increases. The coverage probability of the overall benefit of the best linear rule is close to the nominal level of 95% using ABLO, but less than 95% using other methods. The coverages are not nominal for O-learning, Q-learning, and PM, since their benefit estimates are biased when the candidate rules are misspecified (e.g., true optimal rule is not linear). This is because they use a surrogate loss function that does not guarantee convergence to the indicator function in the benefit function $\delta_{\mathcal{C}}(\widehat{f})$.

The performance of estimation of the subgroup ITR benefit shows similar results, whereby ABLO outperforms O-learning, Q-learning, and PM in both settings, especially when $W \in [-0.5, 0.5]$, and $W > 0.5$. Table 2.2 reports the probability of rejecting $H_0 : \delta_{\mathcal{C}_k}(f_L^*) - \delta_{\mathcal{C}_3}(f_L^*) = 0, k = 1$ or 2, using the HTB test with a null distribution of χ^2_2. The rejection rates of the HTB tests of V and U^1 that do not have a difference in ITR benefit across subgroups correspond to the type I error rate. The type I error rates of ABLO are close to 5%, but conservative for the other three methods. To examine the power, we test the effect of W on the benefit across subgroups defined by discretizing W at -0.5 and 0.5. The power of ABLO is much greater than the other three methods especially when the sample size is small. The other three methods underestimate the benefit function, and thus the HTB test is conservative and less powerful.

Lastly, we test the difference in the benefit between the ITRs and the non-personalized rule in the overall sample and the subgroups. Table 2.3 shows that with a sample size of 800, ABLO is the only method that provides a significantly better benefit than the non-personalized rule with a large power ($> 80\%$). When the sample size is large ($N = 1600$), ABLO, Q-learning, and O-learning have a power of $\geq 88\%$. As for the subgroups, the ITR estimated by ABLO is more likely to outperform the non-personalized rule on the subgroups showing a larger true benefit (i.e., when $W > 0.5$).

Additional simulation results varying the strength of the prescriptive feature variable W are described in Appendix Section A.2. In simulation settings where the conditional outcome model is correctly specified and all variables are observed, regression-based methods Q-learning and PM also perform well.

2.5 Application to the STAR*D Study

STAR*D (Rush et al., 2004) was conducted as a multi-site, multi-level, randomized controlled trial designed to compare different treatment regimes for major depressive disorder when patients fail to respond to the initial treatment of Citalopram (CIT) within 8 weeks. The primary outcome, Quick Inventory of Depressive Symptomatology (QIDS) score (ranging from 0 to 27), was measured to assess the severity of depression. A lower QIDS score indicates less symptoms and thus reflects a better outcome. Participants with a total QIDS score under 5 were considered to experience a clinically meaningful response to the assigned treatment and were therefore remitted from future treatments.

The trial had four levels of treatments (e..g, see Figure 2.3 in Chakraborty and Moodie (2013)); we focused on the first two levels. At the first level, all participants were treated with CIT for a minimum of 8 weeks. Participants who had clinically meaningful response were excluded from level-2 treatment. At level-2, participants without remission with level-1 treatment were randomized to level-2 treatment based on their preference to switch or augment their level-1 treatment. Patients who preferred to switch treatment were randomized with equal probability to bupropion (BUP), cognitive therapy (CT), sertraline (SER), or venlafaxine (VEN). Those who preferred augmentation were randomly assigned to CIT+BUP, CIT+buspirone (BUS), or CIT+CT. If a patient had no

preference, s/he was randomized to any of the above treatments.

The clinical outcome (reward) is the QIDS score at the end of level-2 treatment. There were 788 participants with complete feature variable information included in our analysis. We compared two categories of treatments: 1) treatment with selective serotonin reputake inhibitors (SSRIs, alone or in combination): CIT+BUS, CIT+BUP, CIT+CT, and SER; and 2) treatment with one or more non-SSRIs: CT, BUP, and VEN. Feature variables used to estimate the optimal ITR included the QIDS scores measured at the start of level-2 treatment (level 2 baseline), the change in the QIDS score over the level-1 treatment phase, patient preference regarding level-2 treatment, and demographic variables (gender, age, race), and family history of depression. As the randomization to treatment was based on patient preference, we estimated $\pi(A_i|\boldsymbol{X}_i)$ using empirical proportions based on preferring switching or no preference, because patients who preferred augmentation were all treated with an SSRI and were excluded from the analysis.

We applied four methods to estimate the optimal ITR for patients with MDD who did not achieve remission with 8 weeks of treatment with CIT. For all methods, we randomly split the sample into a training and testing set with a 1:1 ratio and repeated the procedure 500 times. The value function and ITR benefits were evaluated on the testing set. PM, Q-learning, O-learning, and ABLO are compared in Figure 2.5 and Table 2.4. The non-personalized rules yield a QIDS score of 10.16 for SSRI and 9.60 for non-SSRI, with a difference of 0.56. The ITR estimated by ABLO yields a QIDS score of 9.32 (sd = 0.23), which is smaller than PM (9.69, sd = 0.38), Q-learning (9.50, sd=0.35), and O-leaning (9.55, sd = 0.41). The overall ITR benefit estimated by ABLO (1.11, sd = 0.46) is much larger than PM (0.38, sd = 0.76), Q-learning (0.77, sd = 0.70), and O-leaning (0.66, sd = 0.82). The ITR benefit based on ABLO is also larger than the non-personalized rule (1.11 versus 0.56).

In Figure 2.5, we also present the performance of assigning all subjects to non-SSRI by a blue dashed line. ABLO is the only method that demonstrates more than 75% of times a greater ITR value and ITR benefit compared to the non-personalized rule. We performed a one-sided HTB test in Section 2.2.3 to compare the value function and benefit of the linear ITR with assigning all to non-SSRI in the overall sample. We observe a difference in benefit of 0.56 points (CI: $(-0.20, \infty)$, p-value= 0.11), and a difference of 0.28 points (CI: $(-0.10, \infty)$, p-value= 0.11) in value function. The confidence intervals and p-values were calculated over repeated cross-validations, which only accounted for the variability of cross-validations given the STAR*D data. Although the differences are not statistically significant, they demonstrate some evidence of improvement by using a personalized treatment strategy that worth future studies.

The final STAR*D linear decision function estimated by ABLO using full data can be expressed as

$$\begin{aligned}\widehat{f}(\boldsymbol{X}) = &- 12.97 + 0.30 * sex + 1.27 * white + 0.79 * black + 2.77 * depression + 0.05 * age \\ &+ 0.26 * qids.start - 3.40 * qids.slope + 2.39 * preference,\end{aligned}$$

and the linear ITR is to treat a patient with SSRI if $\widehat{f} > 0$; otherwise treat with a non-SSRI if $\widehat{f} \leq 0$. The variable "sex" was coded as one for female and "preference" was coded as one for switch and zero for no preference.

Clinical literature suggests that the baseline MDD severity may be a moderator for treatment response (Bower et al., 2013). In addition, baseline MDD severity is highly associated with suicidality; thus, patients with severe baseline MDD (QIDS ≥ 16) represent an important subgroup. We partitioned patients into mild (QIDS ≤ 10), moderate (QIDS $\in [11, 15]$), and severe (QIDS ≥ 16) MDD subgroups. Using ABLO and the HTB test, baseline QIDS score was found to be

significantly associated with ITR benefit: two subgroups show a large positive ITR benefit (2.22 for the mild group and 2.02 for the severe group), whereas the moderate subgroup shows no benefit (ITR benefit $= -0.18$). This result indicates that patients with mild or severe baseline depressive symptoms (high or low QIDS score) might benefit from following the estimated linear ITR. For patients who are moderately depressed (QIDS $\in [11, 15]$), the linear ITR estimated from the overall sample does not adequately fit the data and does not outperform a non-personalized rule. Thus, we re-fit a linear rule using ABLO for the moderate subgroup only. The re-estimated ITR yields a lower average QIDS score of 8.93 (sd $= 0.35$), with a much improved subgroup ITR benefit of 0.60 (sd $= 0.70$). This analysis demonstrates the advantage of the ITR benefit diagnostic measure, the HTB test, and the value of re-fitting the ITR on subgroups showing a lack-of-fit.

2.6 Discussion

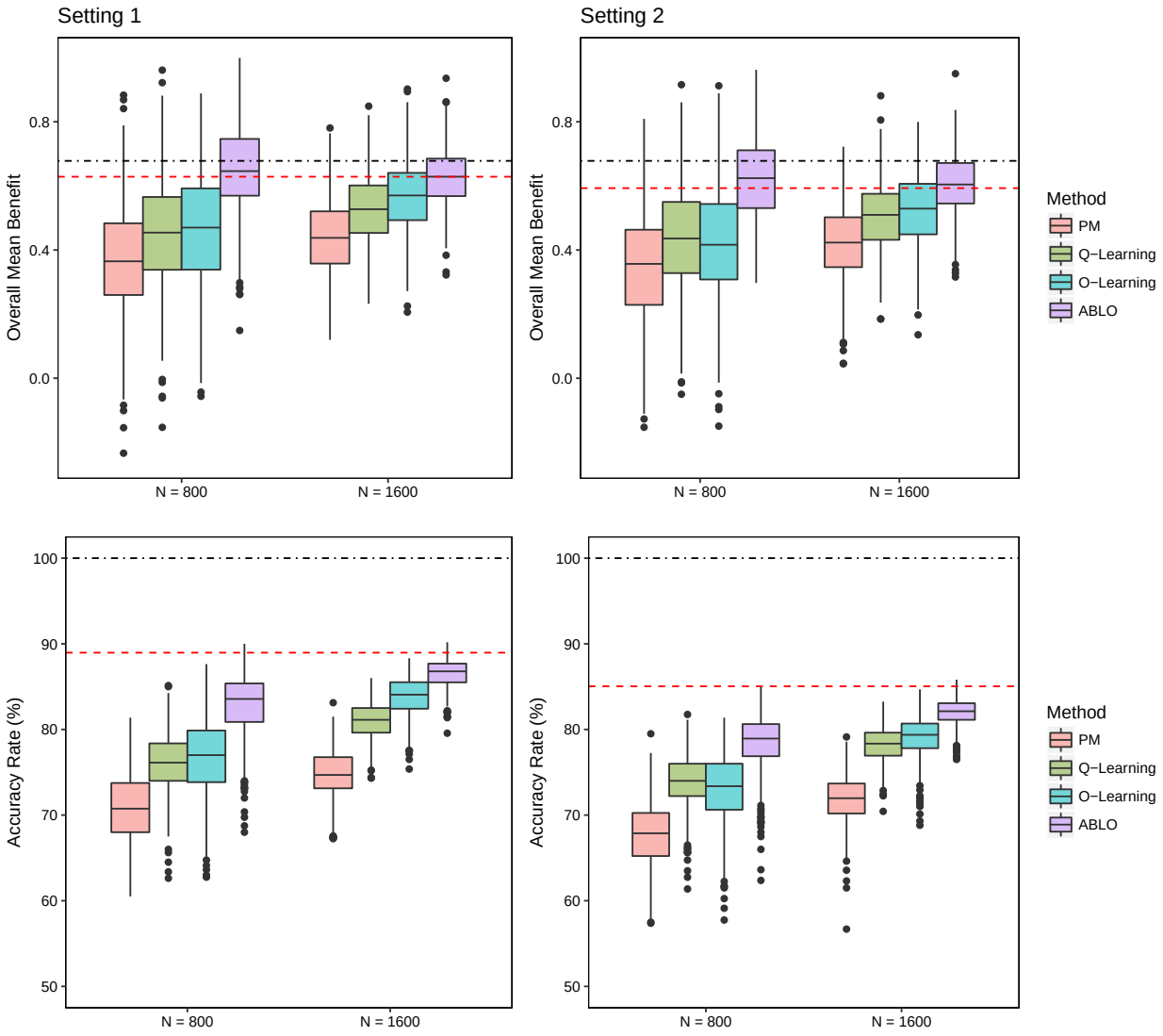
In this work, we propose a diagnostic measure (benefit function) to compare candidate ITRs, a machine learning method (ABLO) to estimate the optimal linear ITR, and several tests for goodness-of-fit. In practice, often not all predictive and prescriptive variables that influence heterogeneous responses to treatment are known and collected. Thus, it is unrealistic to expect that an ITR that benefits each and every individual can be identified. Our practical solution proposes to evaluate the average ITR effect over the entire population and on vulnerable or important subgroups. Although we focus on linear decision functions here, it is straightforward to extend ABLO to other simple decision functions such as polynomial rules by choosing other kernel functions (i.e., polynomial kernel). ABLO can also be applied to observational studies using propensity scores to replace $\pi(A|\boldsymbol{X})$ under the assumption that the propensity score model is correctly specified. We

prove the asymptotic properties of ABLO and identify a condition to avoid the non-regularity issue (in Section 2.3). In practice, when such issue is of concern, adaptive inference (Laber and Murphy, 2011) can be used to construct confidence intervals.

ABLO can consistently estimate the ITR benefit function regardless of misspecification of the rule by drawing a connection with the robust machine learning approach for approximating the zero-one loss. We provide an objective diagnostic measure for assessing optimization. In our method, prescriptive variables mostly contribute to the estimation of the optimal treatment rule while predictive variables mostly contribute to the development of the diagnostic measure and assessment of the benefit of the optimal rule. Future work will consider methods to distinguish these two sets of variables, which potentially overlap.

ABLO is slower than O-learning because it involves iterations of quadratic programming when applying the DCA. In addition, certain simulations show that the algorithm can be slightly sensitive to the initial values in extreme cases (examples provided in Figure A.3 in Appendix). However, our numeric results show that O-learning estimators serve as adequate initial values leading to fast convergence of the DCA. Another limitation is that the current methods only apply to single-stage trials. ABLO can be extended to multiple stage setting following a similar backward multi-stage O-learning in Zhao et al. (2015). The objective function in multi-stage O-learning will be replaced by the ramp loss and the benefit function will be extended with some attention to subjects whose observed treatment sequences are partially consistent with the predicted optimal treatment sequences. In addition, ABLO does not perform variable selection and therefore it cannot distinguish important tailoring variables from noise variables. A possible solution is to replace the $\mathcal{L}2$-norm penalty by $\mathcal{L}1$-norm penalty when minimizing the penalized weighted sum of ramp loss in (2.4).

Figure 2.3: Simulation results: Overall ITR benefit and optimal treatment accuracy rates for the four methods



*:Dotted-dashed lines represent the benefit (top panels) and accuracy (bottom panels) under the theoretical global optimal treatment rule f^*. Dashed lines represent the benefit and accuracy under the theoretical optimal linear rule f_L^*. The methods being compared are (from left to right): PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning.

Table 2.1: Simulation results: mean and standard deviation of the accuracy rate, mean ITR benefit, and coverage probability for estimation of the benefit of the optimal ITR

Setting 1. Four region means $= (1, 0.5, -1, -0.5)$.									
		Overall Benefit		$W < -0.5$		$W \in [-0.5, 0.5]$		$W > 0.5$	
	Accuracy rate	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage
$N = 800$									
PM	0.71 (0.04)	0.37 (0.17)	0.69	0.08 (0.23)	0.97	0.36 (0.23)	0.82	0.67 (0.30)	0.72
Q-learning	0.76 (0.03)	0.45 (0.17)	0.80	0.17 (0.22)	0.97	0.46 (0.23)	0.89	0.73 (0.29)	0.78
O-learning	0.77 (0.05)	0.46 (0.18)	0.82	0.17 (0.24)	0.97	0.46 (0.24)	0.89	0.76 (0.30)	0.80
ABLO	0.83 (0.04)	0.65 (0.14)	0.94	0.30 (0.23)	0.92	0.64 (0.20)	0.96	1.01 (0.24)	0.93
$N = 1600$									
PM	0.75 (0.03)	0.44 (0.12)	0.64	0.11 (0.17)	0.96	0.43 (0.17)	0.80	0.79 (0.20)	0.71
Q-learning	0.81 (0.02)	0.52 (0.11)	0.86	0.18 (0.16)	0.97	0.53 (0.15)	0.92	0.86 (0.19)	0.82
O-learning	0.84 (0.02)	0.57 (0.11)	0.93	0.19 (0.15)	0.97	0.57 (0.16)	0.95	0.94 (0.19)	0.90
ABLO	0.86 (0.02)	0.63 (0.09)	0.96	0.22 (0.15)	0.97	0.63 (0.15)	0.95	1.04 (0.17)	0.94
Best linear rule†	0.890	$\delta^l_{\mathcal{C}} = 0.629$		$\delta^l_{\mathcal{C}} = 0.192$		$\delta^l_{\mathcal{C}} = 0.621$		$\delta^l_{\mathcal{C}} = 1.071$	
Setting 2. Four region means $= (1, 0.3, -1, -0.3)$.									
		Overall Benefit		$W < -0.5$		$W \in [-0.5, 0.5]$		$W > 0.5$	
	Accuracy rate	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage
$N = 800$									
PM	0.68 (0.04)	0.34 (0.17)	0.67	0.10 (0.24)	0.95	0.34 (0.24)	0.83	0.59 (0.30)	0.71
Q-learning	0.74 (0.03)	0.43 (0.16)	0.85	0.16 (0.23)	0.97	0.44 (0.22)	0.92	0.70 (0.28)	0.82
O-learning	0.73 (0.04)	0.42 (0.17)	0.84	0.16 (0.21)	0.98	0.43 (0.24)	0.90	0.68 (0.29)	0.79
ABLO	0.78 (0.03)	0.62 (0.13)	0.95	0.30 (0.21)	0.96	0.62 (0.21)	0.96	0.94 (0.25)	0.92
$N = 1600$									
PM	0.72 (0.03)	0.42 (0.12)	0.69	0.12 (0.17)	0.95	0.42 (0.17)	0.84	0.72 (0.20)	0.73
Q-learning	0.78 (0.02)	0.51 (0.11)	0.89	0.19 (0.16)	0.96	0.52 (0.15)	0.94	0.81 (0.18)	0.85
O-learning	0.79 (0.02)	0.52 (0.11)	0.91	0.19 (0.16)	0.95	0.53 (0.16)	0.93	0.85 (0.19)	0.89
ABLO	0.82 (0.02)	0.61 (0.10)	0.94	0.25 (0.16)	0.94	0.61 (0.15)	0.95	0.96 (0.17)	0.95
Best linear rule†	0.850	$\delta^l_{\mathcal{C}} = 0.593$		$\delta^l_{\mathcal{C}} = 0.200$		$\delta^l_{\mathcal{C}} = 0.583$		$\delta^l_{\mathcal{C}} = 0.996$	
Best global rule†		$\delta_{\mathcal{C}} = 0.678$		$\delta_{\mathcal{C}_1} = 0.285$		$\delta_{\mathcal{C}_2} = 0.647$		$\delta_{\mathcal{C}_3} = 1.109$	

*: The theoretical best linear rule for both settings is $\text{sign}(X_s)$, where $X_s = X^1 + X^2 + \cdots + X^{10}$.

†: The true value of the best linear rule and best global rule is computed from a large independent test data set.

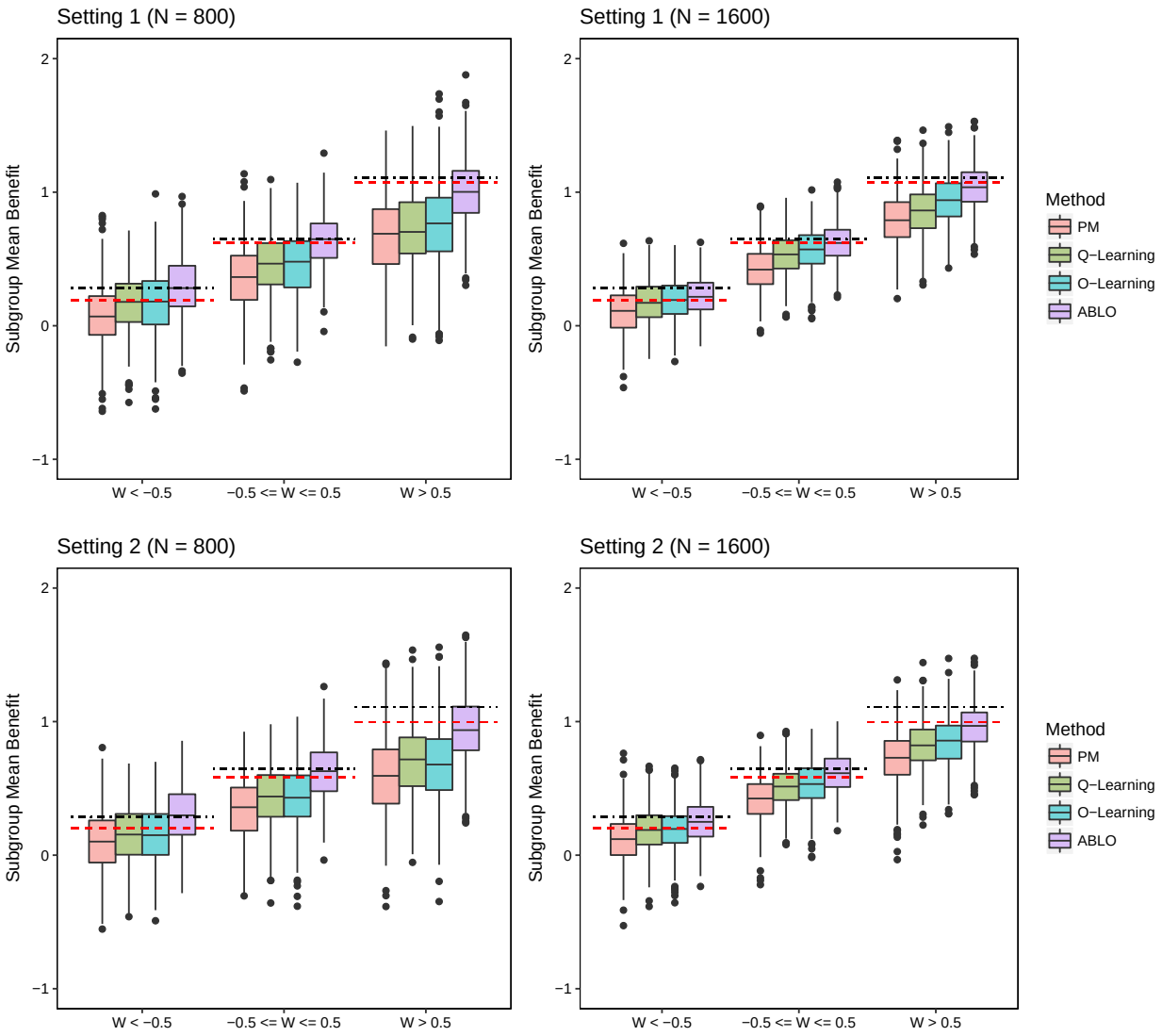
PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning.

Table 2.2: Simulation results: probability of rejecting the null hypothesis that the treatment benefit across subgroups is equivalent by the HTB test

	W	X^1	V	U^1
Setting 1. Four region means $= (1, 0.5, -1, -0.5)$.				
$N = 800$				
PM	0.16	0.05	0.03	0.02
Q-learning	0.18	0.06	0.03	0.03
O-learning	0.21	0.05	0.03	0.03
ABLO	0.42	0.07	0.05	0.06
$N = 1600$				
PM	0.52	0.05	0.05	0.02
Q-learning	0.61	0.05	0.04	0.02
O-learning	0.71	0.04	0.04	0.02
ABLO	0.84	0.05	0.05	0.03
Setting 2. Four region means $= (1, 0.3, -1, -0.3)$.				
$N = 800$				
PM	0.12	0.03	0.02	0.02
Q-learning	0.17	0.04	0.03	0.04
O-learning	0.15	0.03	0.03	0.03
ABLO	0.34	0.06	0.04	0.05
$N = 1600$				
PM	0.42	0.06	0.04	0.03
Q-learning	0.56	0.07	0.04	0.03
O-learning	0.57	0.07	0.03	0.03
ABLO	0.74	0.10	0.04	0.05

*: W has strong signal; X^1 has weak signal; V and U^1 have no signal.

Figure 2.4: Simulation results: Subgroup ITR benefit for the four methods



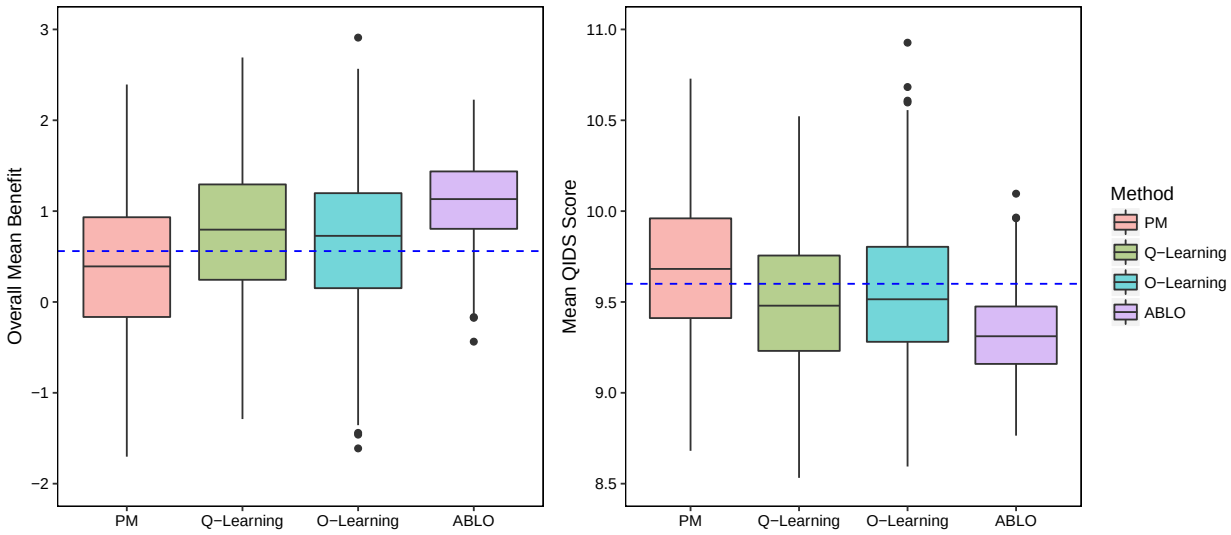
*: Dotted-dashed lines represent the benefit under the theoretical global optimal treatment f^*. Dashed lines represent the benefit under the theoretical optimal linear rule f_L^*. The methods being compared are (from left to right): PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning.

Table 2.3: Simulation results: Comparison of the ITR to the non-personalized universal rule. The proportion of rejecting the null that the ITR has the same benefit as the universal rule* are reported for the overall sample and by subgroups.

	Setting 1. Four region means $= (1, 0.5, -1, -0.5)$.			
	Overall	$W < -0.5$	$W \in [-0.5, 0.5]$	$W > 0.5$
$N = 800$				
PM	0.22	0	0.09	0.33
Q-learning	0.37	0.02	0.20	0.40
O-learning	0.39	0.02	0.20	0.43
ABLO	0.86	0.07	0.47	0.78
$N = 1600$				
PM	0.76	0.02	0.38	0.83
Q-learning	0.92	0.05	0.59	0.90
O-learning	0.95	0.06	0.67	0.94
ABLO	0.99	0.08	0.79	0.98
	Setting 2. Four region means $= (1, 0.3, -1, -0.3)$.			
$N = 800$				
PM	0.18	0.01	0.07	0.27
Q-learning	0.35	0.03	0.17	0.37
O-learning	0.31	0.03	0.17	0.35
ABLO	0.82	0.07	0.43	0.74
$N = 1600$				
PM	0.72	0.03	0.38	0.75
Q-learning	0.88	0.05	0.57	0.86
O-learning	0.90	0.07	0.59	0.86
ABLO	0.99	0.12	0.77	0.97

*: For Setting 1, the mean difference (sd) of the universal rule is 0.09(0.08) for $N = 800$ and 0.07(0.05) for $N = 1600$.
For Setting 2, the mean difference (sd) of the universal rule is 0.11(0.08) for $N = 800$ and 0.08(0.05) for $N = 1600$.

Figure 2.5: STAR*D analysis results: Distribution of the estimated ITR benefit (the higher the better) and QIDS score (the lower the better) at the end of level-2 treatment for the four methods (based on 500 cross-validation runs).



*: The methods being compared are (from left to right): PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning. The blue dashed line indicates non-personalized rule with non-SSRI.

Table 2.4: Results of STAR*D Data Analysis

	QIDS score	ITR benefit	Subgroup ITR benefit by baseline QIDS score		
	Mean(sd)	Mean(sd)	QIDS$\leq$ 10	QIDS$\in$ [11, 15]	QIDS$\geq$ 16
PM	9.69(0.38)	0.38(0.76)	1.29(0.82)	-0.10(1.02)	0.40(1.67)
Q-learning	9.50(0.35)	0.77(0.70)	2.08(0.68)	-0.17(0.92)	1.09(1.62)
O-learning	9.55(0.41)	0.66(0.82)	1.58(0.92)	-0.23(0.95)	1.20(1.84)
ABLO	9.32(0.23)	1.11(0.46)	2.22(0.45)	-0.18(0.51)	2.02(1.12)

*: lower QIDS score indicates a better outcome; higher benefit indicates a better outcome.

Chapter 3

Composite Interaction Tree for Learning Optimal Individualized Treatment Rules and Subgroups

3.1 Overview

In this chapter, we propose a tree-based learning method, composite interaction tree (CITree), to simultaneously estimate the optimal individualized treatment rules and identify subgroups with small and large benefit. In Section 3.2, we propose the statistical method and discuss the algorithm of CITree in detail. In Section 3.3, we conduct two simulation studies to investigate the performance of HTB test and compare the performance of the proposed CITree method to existing tree-based method. In Section 3.4, we apply the method to the REVAMP data and fit a CITree for the whole sample. Finally, we end this chapter with conclusions and discussions in Section 3.5.

3.2 Methodologies

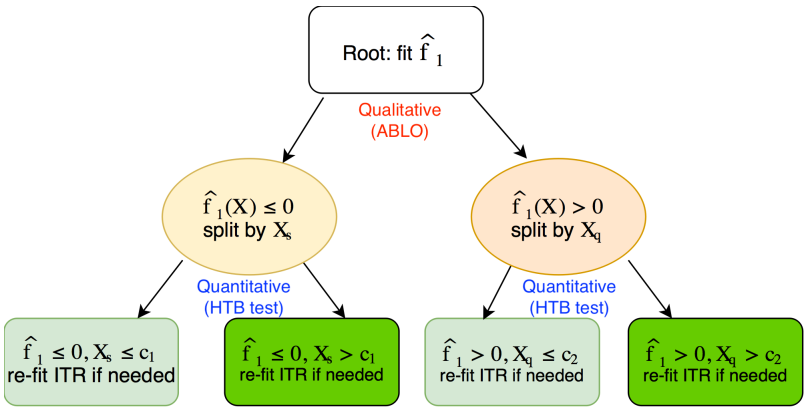
In this Chapter, we still focus on single stage two-arm randomized trials and use the same notation as described in Section 2.2. In clinical practice, linear decision rules are useful due to their transparency and simplicity for interpretation. However, many existing methods may not yield an ITR with the maximal value function within the class of linear rules, because they focus on optimizing some surrogate objective function for treatment benefit (2.2). Note that optimizing the surrogate function only guarantees the optimality when there is no restriction on the functional form of the rules (i.e., nonparametric rules). In Chapter 2, an asymptotically best linear O-learning (ABLO) was proposed to guarantee that the estimated decision rule is optimal among all linear rules by replacing the zero-one loss in (2.10) with the ramp-loss and estimating the optimal linear ITR by minimizing a penalized loss function (2.4).

Note that ABLO estimates a "global" linear ITR that maximizes the value function applied to the overall sample. The linear rule may be overly simplified so that it does not lead to the optimal rule for some subpopulations. Under the ramp loss, the empirical benefit function (2.10) can be used to identify subgroups with a poor fit under a "global" linear rule and provide an opportunity to re-fit a "local", piece-wise linear rule using subjects in the subgroups. Thus, here we propose the composite interaction tree (CITree) to automatically detect inadequate fit on subgroups and discover subgroups exhibiting heterogeneous ITR benefit. The central idea of CITree is to improve the performance of the estimated ITR by re-fitting rules in subgroups with poor performance.

Typically, a tree-based method recursively partitions parent nodes into two child nodes using a criterion, such as Gini index for CART or G-statistic for interaction tree (IT). Therefore, the tree growing step is the same from one level to the next and the parent/child nodes have similar

characteristics. In order to detect both qualitative and quantitative interactions, CITree consists of two types of parent/child nodes corresponding to qualitative and quantitative splits. Briefly, for nodes at odd levels, we use ABLO to estimate an optimal ITR using all feature variables to detect qualitative interaction. For nodes at even levels, we perform a heterogeneity of ITR benefit (HTB) test to partition subjects in a parent node into two child nodes with homogeneous ITR benefit, and detect quantitative interaction and lack-of-fit. The motivation for CITree is that the overall ITR is more likely to be non-optimal for subgroups with a lower benefit. Therefore, the overall goal is to maximize the value function of the overall population by re-fitting the ITR on certain subpopulations. Unlike CART which is applied to data with known labels, CITree is an unsupervised learning that estimates the optimal ITR to assign treatment to future subjects.

Figure 3.1: Diagram of an example Composite Interaction Tree (CITree)*



*: For rounded rectangulars, CITree fits $\hat{f}$ to identify qualitative interaction and estimate the optimal linear ITR using all feature variables (achieve homogeneous optimal treatment). For ellipses, CITree searches each feature variable to identify significant quantitative interaction in ITR benefit (achieve homogeneous benefit).

Consider the root node as level 1 in an example CITree in Figure 3.1. The first step is a qualitative split partitioning subjects into subgroups of homogeneous optimal treatment rules. Specifically, we estimate an optimal linear ITR using ABLO on the overall sample and obtains a

linear decision function $\widehat{f}_1(\boldsymbol{X})$. Based on the predicted optimal treatment using $\widehat{f}_1(\boldsymbol{X})$, subjects will be partitioned into two nodes at level 2, where on the left are subjects whose optimal treatment is predicted to be -1, e.g., $\{i : \widehat{f}_1(\boldsymbol{X}_i) \leq 0\}$, and on the right are those whose optimal treatment is predicted to be the complementary treatment, e.g., $\{i : \widehat{f}_1(\boldsymbol{X}_i) > 0\}$. The second step is a quantitative split partitioning each of the nodes at level 2 into two child nodes with homogeneous ITR benefit. Specifically, we partition the sample space according to a covariate greater than a threshold, $X_s > c$, to maximize the difference in ITR benefit between groups in the resulting child nodes, $\mathcal{C}_1 : X_s \leq c$ and $\mathcal{C}_2 : X_s > c$, by performing an HTB test. The null hypothesis is

$$H_0 : \delta_{\mathcal{C}_1}(f) - \delta_{\mathcal{C}_2}(f) = 0,$$

and the HTB statistic is defined as $\text{HTB} = [\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) - \widehat{\delta}_{\mathcal{C}_2}(\widehat{f})]^2 / \text{Var}(\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) - \widehat{\delta}_{\mathcal{C}_2}(\widehat{f}))$, where the subgroup benefit is estimated by (2.10) and the variance of $\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) - \widehat{\delta}_{\mathcal{C}_2}(\widehat{f})$ given $\widehat{f}$ is estimated based on a formula in Appendix B.1. We can show that the conditional distribution given $\widehat{f}$ of the test statistic computed from an independent sample follows a $\chi^2(1)$ under H_0 (such an HTB test is implemented in the honest CITree as described subsequently). The HTB test is a measure to evaluate ITR benefit heterogeneity between two child nodes to detect quantitative interactions.

In step 2, we search for the threshold c among grid points $c = \{c_1, \cdots, c_J\}$ (e.g., quantiles of the observed samples) and the covariate space X_s for $s = 1, \cdots, p$, that has the smallest p-value. To protect against overfitting, we use Benjamin and Hochberg procedure (Benjamini and Hochberg, 1995) to control for the false discovery rate (FDR). Given $\widehat{f}_1(\boldsymbol{X})$, a covariate and cut-point, we partition each node into child nodes. For example, in Figure 3.1, four child nodes at level 3 are $\{\widehat{f}_1(\boldsymbol{X}) \leq 0; X_s \leq c_1\}$, $\{\widehat{f}_1(\boldsymbol{X}) \leq 0; X_s > c_1\}$, $\{\widehat{f}_1(\boldsymbol{X}) > 0; X_q \leq c_2\}$, and $\{\widehat{f}_1(\boldsymbol{X}) > 0; X_q > c_2\}$. Note that X_s and X_q are selected independently based on different subgroups, so they can be

different. Each node not only indicates the estimated optimal treatment, but also partitions patients based on the magnitude of the benefit gained by following the estimated optimal rule, and chooses the most significant cut point after adjusting for the FDR. If no HTB test is significant at the controlled FDR level, CITree stops partitioning the parent node.

The next step is to consider re-fitting the ITR using only subjects in each child node and denote the decision functions as $\widehat{f}_2$. If the value function is improved by the re-fitted ITR, CITree further partitions the current node into subgroups of different optimal treatment based on $\widehat{f}_2$. CITree stops growing when no HTB test is significant, or the number of observations in the node is less than a pre-specified number, or the re-fitted ITR does not improve the value function (stopping rules are described in Algorithm 1). The resulting ITR for a K terminal nodes tree takes the form of $\sum_k I(\boldsymbol{X} \in \widehat{\mathcal{R}}_k)\widehat{f}_k(\boldsymbol{X})$, where $\widehat{\mathcal{R}}_k$ indicates subjects in the kth terminal node following a particular path on the tree, and $\widehat{f}_k(\boldsymbol{X}) = \widehat{\beta}_{0k} + \boldsymbol{X}^T\widehat{\boldsymbol{\beta}}_k$ is the final linear ITR fitted from ABLO for the node k. The fitted ITR contains binary partitioning of the covariate space through $\widehat{\mathcal{R}}_k$ and a linear combination of covariates through $\widehat{f}_k(\boldsymbol{X})$ to balance interpretability and flexibility. Algorithm 1 illustrates the procedure.

An advantage of CITree is the simultaneous detection of qualitative and quantitative interactions if both are present. Step 1 identifies qualitative interaction to estimate the optimal treatment using ABLO and step 2 identifies quantitative interaction through HTB test. Note that after performing step 1 to partition subjects into homogeneous groups with similar optimal treatment, step 2 identifies quantitative heterogeneity of ITR effects and diagnoses potential subgroups with a low benefit, which suggests a poor fit of the ITR estimated from the whole sample from upper level. By performing a re-fit on such subgroups, CITree always leads to a higher or equal value function compared to using the full sample. The step of re-fitting ITR offers patients whose optimal

Algorithm 1 Composite Interaction Tree (CITree)

Fit an ITR for the whole sample using ABLO and obtain an initial rule $\widehat{f}_1$.

Step 1: for $l = 1, 3, 5, \cdots$

- Partition each parent node at level l into two child nodes: $\{\widehat{f}_k \leq 0\}$ and $\{\widehat{f}_k > 0\}$.

Step 2: for $l = 2, 4, 6, \cdots$

- Search for the most significant HTB test among covariate space at grid points.
- Split each parent node into two child nodes if the adjusted HTB test is significant for variable X_s at split point c: $\{X_s \leq c\ \}$ and $\{X_s > c\}$.
- Re-fit $\widehat{f}$ for each child node if it improves the value function.

Stopping rules

- If number of subjects in the node is less than a pre-specified number, the node will not be split.
- If no HTB test is significant at even level of nodes, the node will not be split.
- If re-fitting ITR at the odd level of nodes does not improve the value function, the node will not be split.

The final ITR is a piece-wise linear function denoted as $\sum_K I(\boldsymbol{X} \in \widehat{\mathcal{R}}_k)\widehat{f}_k(\boldsymbol{X})$.

treatments are estimated incorrectly at an upper level an opportunity to rectify and search for their true optimal treatment. Another advantage of CITree is to use a linear combination of feature variables rather than one variable (as done in IT or QUINT) to build a tree, which may lead to identifying the optimal rule in fewer steps and with improved accuracy.

CITree uses the same sample for two types of splits. When the sample size is large and overfitting is of primary concern, honest CITree can be used where the HTB test is performed on a hold-out sample not involved in estimating ITRs. Specifically, honest CITree separates the whole sample into two subsets. The training set is only involved to estimate ITR, while the ITR benefit and HTB test are computed using the hold-out sample. This procedure guarantees using an honest measure of the performance (Athey and Imbens, 2016) to determine the tree growth.

3.3 Simulation Studies

We perform extensive simulation studies to evaluate the CITree algorithm. In the first simulation study, we examine whether the algorithm can successfully identify quantitative interaction between covariates and treatment. In a simple simulation setting, we evaluate the performance of HTB test by type I and type II error rates. The type I error rate represents the probability that CITree performs a quantitative split in a group of subjects with no heterogeneous ITR benefit. The type II error rate represents the probability of failure to detect a true ITR heterogeneity. In the second simulation study, we compare the overall performance of CITree with existing tree-based machine learning methods by assessing the overall ITR benefit in a more complex tree setting, and investigate to what extent that CITree can recover the true tree structure.

3.3.1 Simulation Study 1

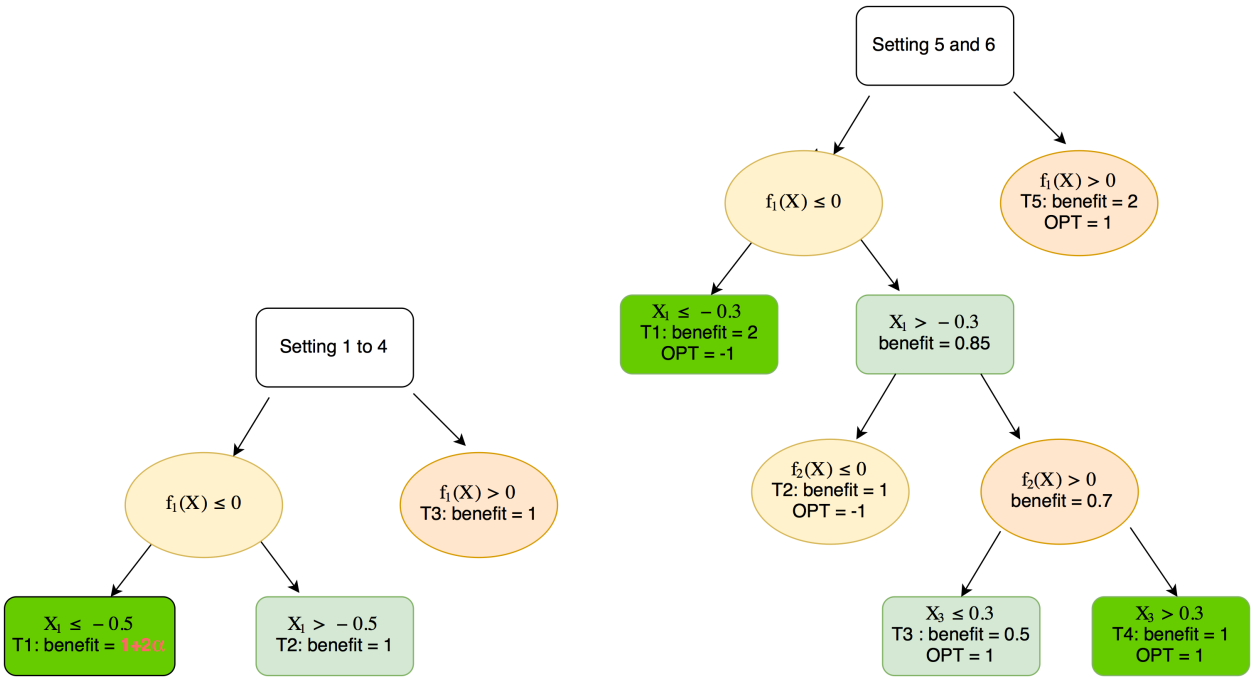
Simulation data sets were generated based on the tree structure in the left panel of Figure 3.2. The simulation setting is inspired by real world applications where patients consist of heterogeneous subgroups with both quantitative and qualitative interactions. In this setting, treatment 1 is more beneficial for subjects in the right node and treatment -1 is more beneficial for subjects in the left node. Within the subgroup where treatment -1 benefits more, there exists a subgroup ($X_1 \leq -0.5$) that has a greater treatment effect comparing to the rest of subjects ($X_1 > -0.5$). The true optimal treatment decision function is $f_1(\boldsymbol{X}) = X_1 + X_2$, where X_1 and X_2 are feature variables generated i.i.d from $N(0, 1)$. For the left node, a quantitative interaction was simulated between X_1 and the optimal treatment, which leads to two child nodes, one with a larger ITR benefit, and the other with a smaller ITR benefit. We compared four settings in the simulations, where A is randomly assigned as 1 or -1 with equal probability, and the clinical outcome R is generated by

$$R = \eta(\boldsymbol{X}) + A\phi(\boldsymbol{X}) + \epsilon, \quad \epsilon \sim N(0, 0.25),$$
$$\phi(\boldsymbol{X}) = 0.5I(f_1(\boldsymbol{X}) > 0) - I(f_1(\boldsymbol{X}) \leq 0)\big[0.5 + \alpha I(X_1 \leq -0.5)\big]. \tag{3.1}$$

In the above model, the first term $\eta(\boldsymbol{X})$ is the main effect not contributing to defining the true ITR; and the second term $A\phi(\boldsymbol{X})$ indicates that the optimal ITR should have the same sign as f_1. For setting 1 and 2, $\eta(\boldsymbol{X}) = X_1 - 0.5X_2$ and no other feature variables were considered in the analysis. For setting 3 and 4, $\eta(\boldsymbol{X}) = X_1 - 0.5X_2 + 0.5X_3$, variables X_3 and X_4 were generated i.i.d following $N(0, 1)$ and not contributing to the optimal ITR. In these two settings, all four feature variables were considered when fitting CITree. Parameter α in model (3.1) determines the effect size of the quantitative interaction, which is set to be 0.5 for setting 1 and 3, and 1 for setting 2 and 4. True benefits for subjects in terminal nodes under the true optimal rule are summarized in

Figure 3.2.

Figure 3.2: CITree structure for generating data of simulation studies (left panel: simulation study 1; right panel: simulation study 2)



*: For setting 1 and 3, $\alpha = 0.5$; for setting 2 and 4, $\alpha = 1$.

In each simulation, honest CITree randomly partitioned the whole sample into two subsets in step 1 and step 2 of the CITree algorithm. We examined different sample size ratio for step 1 and step 2 (2:1, 1:1, and 1:2) in order to assess the performance of the fitted ITR and HTB test. Step 1 was applied to the first data set to obtain $\widehat{f}_1$, which was used to assign subjects in the second set into two nodes at level 2 based on their predicted optimal treatment. Thus subjects in step 1 and step 2 were independent, and HTB tests were evaluated on two nodes at level 2. For each feature variable, we searched from the values of 1/4 quantile to 3/4 quantile, with an increase of 0.1. We split a node when at least one of the HTB tests is significant at the controlled FDR level. By the

simulation design, we can use the left node to evaluate the type II error and the right node to evaluate the type I error rate. The sample size was $N = 500, 1,000$, and $2,000$. We controlled FDR at rate 0.05 and 0.1. To evaluate the performance of ITR, we simulate an independent validation data set for each setting with $N = 10,000$. We report the overall benefit and accuracy rate on the validation set.

Results from $1,000$ replicates were summarized in Table 3.1. The type I error rates of HTB test are adequately controlled at the FDR rate for CITree. For honest CITree, type I error rates of HTB test are controlled at the FDR rate for setting 1 and setting 2. For setting 3 and 4, the type I errors were better controlled when the sample size for step 2 is larger. The power of both CITree and honest CITree increase as the sample size and the effect size of quantitative interaction increase. Given the same total sample size, the power of HTB test increases as the sample size of step 2 increases, which leads to a greater power of CITree over honest CITree. However, a smaller sample size of step 1 will affect the performance of fitted ITR. Table 3.1 shows that the accuracy rates and the overall benefits of the honest CITree decrease as the subset of step 1 gets smaller. To balance the performance of step 1 and step 2, we used a 1:1 ratio in the rest of analyses.

When HTB tests were significant, we report the rate of choosing the correct variable for splitting. Given the correct variable was selected, we also report the rate of choosing the correct split point. Similar to Dusseldorp and Van Mechelen (2014), the correct split point is defined as being in the interval $[c - 0.2, c + 0.2]$, where c is the true split point. For setting 2 and 4 with a large benefit heterogeneity between terminal nodes 1 and 2 and a sample size of 1,000 or 2,000, both CITree and honest CITree can almost always identify the quantitative interaction, choose the right variable, and at the right split point. In other settings of reduced benefit heterogeneity and sample size, the ability to identify the correct tree structure is modest.

Table 3.1: Simulation study 1 results: type I error rate and power for HTB tests

	N	Type I error		Power		Correct variable		Correct point		Accuracy	Benefit
FDR		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1		
Honest CITree				sample sizes for step 1 and 2 are 2 to 1							
Setting 1	500	0.033	0.062	0.308	0.409	0.938	0.939	0.682	0.708	0.985 (0.013)	1.239 (0.194)
	1000	0.043	0.072	0.528	0.642	0.992	0.988	0.842	0.841	0.989 (0.010)	1.242 (0.136)
	2000	0.031	0.069	0.879	0.921	0.998	0.998	0.981	0.979	0.990 (0.015)	1.244 (0.102)
Setting 2	500	0.033	0.066	0.860	0.919	0.995	0.991	0.945	0.944	0.983 (0.014)	1.498 (0.202)
	1000	0.046	0.071	0.993	0.997	1	1	0.983	0.983	0.987 (0.015)	1.498 (0.143)
	2000	0.026	0.065	1	1	1	1	1	1	0.988 (0.017)	1.500 (0.110)
Setting 3	500	0.071	0.141	0.275	0.373	0.800	0.780	0.641	0.639	0.980 (0.013)	1.218 (0.209)
	1000	0.069	0.115	0.490	0.616	0.920	0.890	0.834	0.827	0.985 (0.012)	1.224 (0.149)
	2000	0.057	0.116	0.83	0.903	0.983	0.970	0.924	0.924	0.989 (0.015)	1.235 (0.108)
Setting 4	500	0.073	0.138	0.798	0.879	0.961	0.944	0.887	0.889	0.978 (0.014)	1.477 (0.217)
	1000	0.069	0.115	0.985	0.994	0.996	0.995	0.983	0.982	0.985 (0.010)	1.483 (0.156)
	2000	0.052	0.116	1	1	1	1	0.997	0.997	0.988 (0.011)	1.495 (0.109)
Honest CITree				sample sizes for step 1 and 2 are 1 to 1							
Setting 1	500	0.037	0.071	0.410	0.535	0.976	0.966	0.812	0.805	0.982 (0.014)	1.220 (0.153)
	1000	0.025	0.053	0.768	0.840	0.995	0.994	0.932	0.932	0.988 (0.009)	1.234 (0.111)
	2000	0.036	0.072	0.967	0.980	1	1	0.993	0.993	0.990 (0.012)	1.238 (0.080)
Setting 2	500	0.041	0.077	0.965	0.982	0.999	0.999	0.974	0.975	0.980 (0.014)	1.478 (0.159)
	1000	0.025	0.057	1	1	1	1	0.998	0.998	0.986 (0.013)	1.490 (0.118)
	2000	0.029	0.063	1	1	1	1	1	1	0.989 (0.010)	1.498 (0.082)
Setting 3	500	0.072	0.129	0.376	0.496	0.888	0.849	0.769	0.755	0.975 (0.014)	1.208 (0.171)
	1000	0.060	0.115	0.666	0.765	0.953	0.945	0.880	0.874	0.984 (0.010)	1.222 (0.120)
	2000	0.052	0.100	0.962	0.981	0.991	0.988	0.974	0.973	0.988 (0.012)	1.235 (0.089)
Setting 4	500	0.065	0.131	0.947	0.970	0.993	0.987	0.938	0.938	0.974 (0.015)	1.466 (0.181)
	1000	0.054	0.114	1	1	0.999	0.999	0.998	0.998	0.982 (0.010)	1.481 (0.124)
	2000	0.053	0.104	1	1	1	1	1	1	0.987 (0.009)	1.495 (0.091)
Honest CITree				sample sizes for step 1 and 2 are 1 to 2							
Setting 1	500	0.040	0.072	0.532	0.624	0.996	0.984	0.851	0.847	0.977 (0.017)	1.205 (0.132)
	1000	0.025	0.046	0.892	0.922	0.998	0.997	0.962	0.963	0.985 (0.010)	1.230 (0.097)
	2000	0.030	0.067	0.994	0.996	1	1	0.998	0.998	0.989 (0.013)	1.236 (0.070)
Setting 2	500	0.042	0.070	0.993	0.999	1	0.999	0.989	0.989	0.974 (0.017)	1.462 (0.140)
	1000	0.023	0.046	1	1	1	1	1	1	0.983 (0.011)	1.488 (0.099)
	2000	0.034	0.069	1	1	1	1	1	1	0.987 (0.011)	1.494 (0.073)
Setting 3	500	0.071	0.127	0.511	0.618	0.937	0.908	0.808	0.806	0.966 (0.018)	1.187 (0.148)
	1000	0.058	0.109	0.818	0.881	0.979	0.967	0.926	0.926	0.979 (0.010)	1.215 (0.104)
	2000	0.061	0.102	0.994	0.997	0.998	0.997	0.988	0.988	0.986 (0.008)	1.231 (0.076)
Setting 4	500	0.066	0.129	0.983	0.993	0.997	0.997	0.972	0.972	0.964 (0.019)	1.446 (0.156)

	1000	0.063	0.106	1	1	1	1	0.999	0.999	0.978 (0.010)	1.474 (0.107)
	2000	0.061	0.103	1	1	1	1	1	1	0.985 (0.009)	1.489 (0.079)
CITree											
Setting 1	500	0.032	0.054	0.753	0.838	0.997	0.996	0.920	0.920	0.990 (0.010)	1.242 (0.107)
	1000	0.027	0.050	0.983	0.994	1	1	0.985	0.985	0.991 (0.012)	1.245 (0.082)
	2000	0.032	0.063	1	1	1	1	0.997	0.997	0.990 (0.021)	1.240 (0.066)
Setting 2	500	0.030	0.051	1	1	1	1	0.997	0.997	0.988 (0.008)	1.500 (0.112)
	1000	0.028	0.046	1	1	1	1	1	1	0.990 (0.010)	1.503 (0.083)
	2000	0.031	0.068	1	1	1	1	1	1	0.988 (0.018)	1.498 (0.070)
Setting 3	500	0.049	0.102	0.648	0.776	0.965	0.938	0.885	0.882	0.986 (0.008)	1.239 (0.117)
	1000	0.038	0.086	0.950	0.974	0.991	0.991	0.967	0.966	0.990 (0.008)	1.241 (0.086)
	2000	0.039	0.089	1	1	1	1	0.999	0.999	0.990 (0.016)	1.240 (0.068)
Setting 4	500	0.060	0.104	1	1	0.999	0.999	0.994	0.994	0.985 (0.008)	1.498 (0.123)
	1000	0.036	0.086	1	1	1	1	1	1	0.989 (0.009)	1.499 (0.088)
	2000	0.040	0.082	1	1	1	1	1	1	0.990 (0.012)	1.501 (0.067)

3.3.2 Simulation Study 2

In this study, we investigate the performance of different methods in a more complex tree setting. We compared CITree and honest CITree (1:1) with MIDAs (Laber and Zhao, 2015), interaction trees (IT) and QUINT. Unlike other tree-based methods (e.g., IT or QUINT), which first fit a large tree and then prune the tree to avoid overfitting, the CITree does not require pruning since splitting only occurs when the HTB test is significant using an independent testing sample at controlled FDR rate (i.e., honest CI Tree).

In our real data application described in Section 3.4, we found that although the benefit of a global linear rule was greater than the effect of the non-personalized rule on the overall sample, it was worse for some subgroups. In many cases, the performance of the ITR can be improved by refitting the rule within specific subgroups. In order to mimic such real data cases, simulation data sets were generated based on a tree structure in the right panel of Figure 3.2, where treatment 1 is more beneficial for subjects in terminal nodes T3, T4 and T5, and treatment -1 is better for

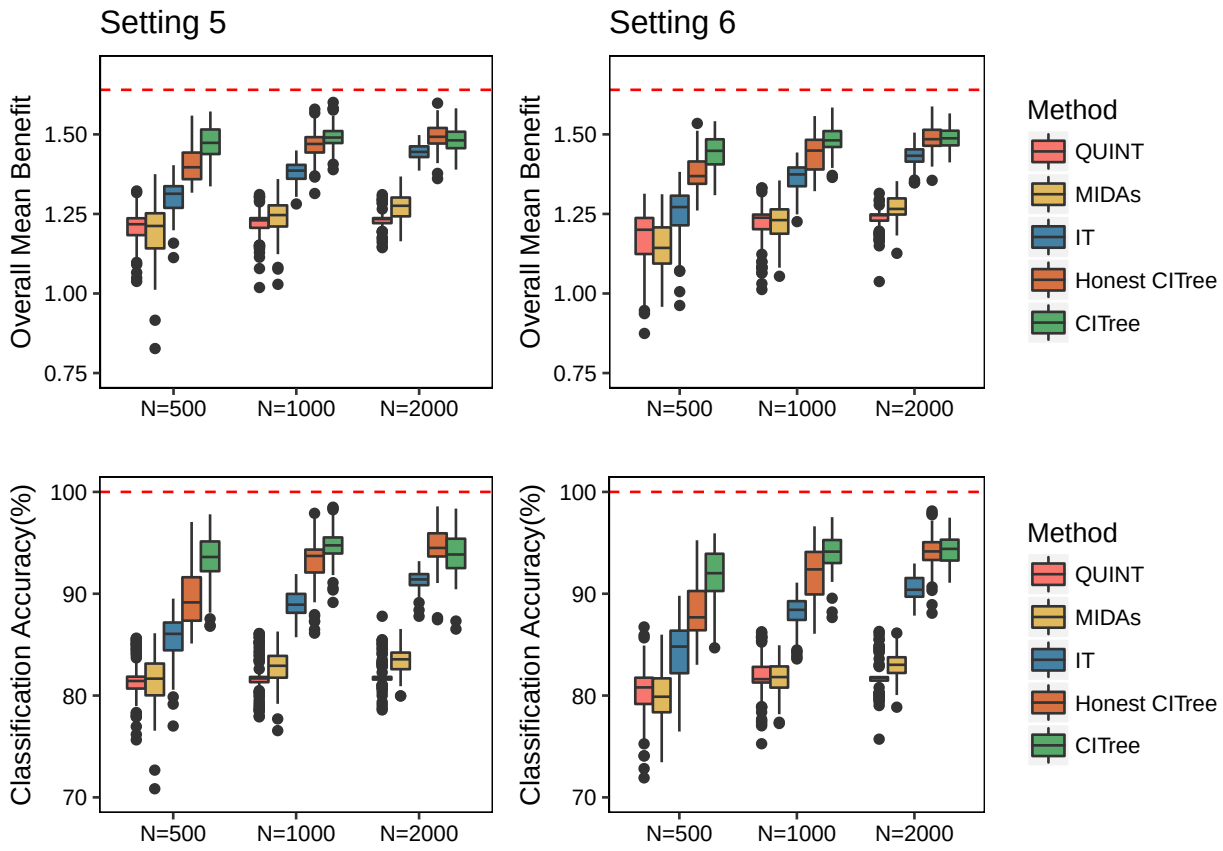
subjects in terminal nodes T1 and T2. ITR benefits are the largest for subjects in terminal nodes T1 and T5. Similar to simulation study 1, we generated four feature variables, X_1 to X_4, and a continuous outcome R by (3.1), but with

$$\begin{aligned}\phi(\boldsymbol{X}) &= I(f_1(\boldsymbol{X})>0)-I(f_1(\boldsymbol{X})\le 0)\Big\{I(X_1\le -0.3)\\ &+ 0.5I(X_1>-0.3)\Big[I(f_2(\boldsymbol{X})\le 0)-I(f_2(\boldsymbol{X})>0)[I(X_3>0.3)+0.5I(X_3\le 0.3)]\Big]\Big\}.\end{aligned}$$

We define $f_1(\boldsymbol{X}) = X_1 + X_2 - 0.5$ and $f_2(\boldsymbol{X}) = X_3 + X_4$. In setting 5, $\eta(\boldsymbol{X}) = X_1 - 0.5X_2 + 0.5X_3 - 0.5X_4$. In setting 6, $\eta(\boldsymbol{X}) = X_1 - 0.5X_2 + 0.5X_3 - 0.5X_4 + 0.5X_5$, where X_5 and X_6 are also i.i.d. with a standard normal distribution. In setting 6, both X_5 and X_6 were included when fitting CITree.

For each simulated data set, CITree, honest CITree, IT, MIDAs, and QUINT were applied to estimate optimal ITR and estimate its benefit. QUINT partitions the overall population into three classes, where treatment -1 is more beneficial to subjects in class 1, treatment 1 is more beneficial to subjects in class 2 and treatment effects are similar for subjects in class 3. Assignment of optimal treatment for individuals depends on the treatment effect of the class they belong to. For those predicted to be in class 3 where the difference of two treatments is not significant, the optimal treatment is determined by the sign of the average effect in the terminal node where the subject belongs to. Interaction tree is a classification/regression tree (CART) where the splitting criterion is based on a t-test of difference between treatment groups. MIDAs is a decision tree method, which splits the parent node into child nodes depending on whether the value function will dramatically increase. For the honest CITree, we randomly partitioned all subjects to two subsets with 1:1 ratio and perform step 1 and step 2 of the algorithm on separate data sets. Other design features of the simulation study is the same as simulation 1.

Figure 3.3: Overall performance of five methods in simulation study 2



*: The red dashed lines represent values based on true optimal rules

Results from 100 replicates are summarized in Table 3.2, 3.3, 3.4, 3.5, and 3.6, and Figure 3.3, 3.4, and 3.5. For both settings, CITree has the best overall performance (Table 3.2 , reported empirical mean and standard deviation) with the highest optimal treatment allocation accuracy and the largest overall benefit when the sample size is 500 or 1,000. As the sample size increases to 2,000, honest CITree performs almost the same as CITree, but with significantly shorter computation time. CITree is the only method that achieves a greater than 90% treatment allocation accuracy when the sample size is 500. QUINT and MIDAs perform worse than the other methods, with less than 85%

Table 3.2: Simulation study 2 results: comparing overall performance of five methods

		Optimal Treatment Allocation Accuracy Rate				
Setting	N	QUINT	MIDAs	IT	Honest CITree	CITree
Setting 5	500	0.815 (0.020)	0.815 (0.027)	0.856 (0.023)	0.898 (0.030)	0.936 (0.023)
	1000	0.818 (0.018)	0.827 (0.017)	0.889 (0.013)	0.931 (0.023)	0.947 (0.016)
	2000	0.820 (0.014)	0.835 (0.013)	0.913 (0.010)	0.946 (0.019)	0.938 (0.020)
Setting 6	500	0.804 (0.027)	0.798 (0.025)	0.841 (0.028)	0.884 (0.027)	0.917 (0.027)
	1000	0.819 (0.022)	0.818 (0.017)	0.881 (0.017)	0.917 (0.029)	0.941 (0.019)
	2000	0.819 (0.016)	0.830 (0.013)	0.905 (0.012)	0.941 (0.018)	0.942 (0.015)
		Overall ITR Benefit				
Setting	N	QUINT	MIDAs	IT	Honest CITree	CITree
Setting 5	500	1.207 (0.054)	1.196 (0.090)	1.301 (0.052)	1.405 (0.059)	1.473 (0.049)
	1000	1.218 (0.045)	1.240 (0.055)	1.383 (0.037)	1.466 (0.044)	1.493 (0.036)
	2000	1.230 (0.027)	1.273 (0.040)	1.444 (0.024)	1.493 (0.041)	1.483 (0.035)
Setting 6	500	1.174 (0.089)	1.146 (0.081)	1.256 (0.079)	1.381 (0.056)	1.441 (0.055)
	1000	1.218 (0.064)	1.222 (0.059)	1.366 (0.044)	1.439 (0.058)	1.485 (0.042)
	2000	1.235 (0.035)	1.269 (0.041)	1.434 (0.031)	1.486 (0.042)	1.488 (0.033)

*: The true overall benefit for both settings is 1.64.

treatment estimation accuracy and the lowest estimated overall benefit. When the total sample size is $2,000$, the estimated overall benefits of the CITree and honest CITree are close to the true benefit of 1.64, and the treatment allocation accuracy is near 95%. Based on the simulation designs, the true terminal node membership of each subject can be determined. Thus, we further examined the optimal treatment allocation accuracy for subjects in each terminal node in Table 3.3. For subjects in terminal nodes T2, T3, and T4, who are most likely to be mis-allocated, both CITree and honest CITree perform much better than other methods in terms of accuracy. MIDAs performs relatively well on T3 and T4 comparing to CITree and honest CITree. Honest CITree and CITree perform adequately for subjects in T2 with accuracy greater than 72%, while QUINT and MIDAs perform the worst with the accuracy rates ranging from 20% to 45%. CITree and Honest CITree have a substantially greater benefit on T3 and T4 when $N = 1,000$ and $2,000$ than other methods due to

re-fitting ITR on these subgroups. QUINT, honest CITree and CITree perform well for subjects in terminal node 1 (T1) with accuracy greater than 96%. IT and MIDAs are the only two methods with lower than 95% accuracy rates for T1 when $N = 500$. MIDAs, IT, honest CITree and CITree have a better performance (accuracy rates over 89.5%) in terminal 5 (T5) comparing to QUINT (accuracy rates around 86%).

In Figure 3.4 and 3.5, we visualize the estimated optimal treatment for each patient in the validation set averaged over 100 repetitions. The optimal treatment boundaries (solid blue lines) were projected onto feature variables X_1 and X_2 for terminal nodes T1, T2, and T5, and onto X_3 and X_4 for T2, T3, and T4. In both figures, the color of dots represents the average estimated optimal treatment (with treatments coded as "1" or "-1") for that subject in the validation set by a certain method based on 100 simulated data sets. The blue solid line is the true optimal treatment boundary for qualitative interaction. The blue dashed line is the true boundary of quantitative interaction (high vs low ITR benefit). Subjects between the solid and dashed lines form the subgroup which is most easily mis-allocated by a sub-optimal method. CITree and honest CITree perform better as shown in Figure 3.4, because they use a combination of feature variables to perform qualitative split. Other methods only consider one variable for each split and therefore do not capture a non-rectangular boundary. In Figure 3.5, our methods also perform better than alternatives, due to performing a quantitative split. An advantage of CITree over other trees that only fits qualitative interaction is that: subgroup with small benefits is the group that most likely to have an inaccurate predicted optimal treatment. For example, MIDAs uses the overall value function in all subjects as criterion, thus may not have adequate fit on this subgroup. In contrast, CITree specifically identifies this subgroup with poor fit and re-fit an ITR locally using data from this subgroup only.

To examine the complexity of the trees fitted by five methods, we report number of terminal nodes for each method in Table 3.4. The average number of terminal nodes for honest CITree and CITree is around 5, which is the true number of nodes. The average number of tree levels is also close to the the truth (5 levels). QUINT and MIDAs also generate relatively simple trees with an average number of 6 to 7 terminal nodes. IT has the most complicated tree, after pruning it still terminates with 10 to 25 terminal nodes (more complicated with larger sample size). With a relatively simple tree, CITree has better performance than IT with a more complicated structure.

To further explore performance on recovering the tree structure, we report positive predictive value (PPV) and subgroup benefit for the terminal nodes predicted by CITree and honest CITree. The PPV of a terminal node represents the proportion of subjects who truly belong to the node among those who are predicted to be in a terminal node. For T3 and T4, the sample size in both nodes are small and the difference in benefit are low, thus the results with T3 and T4 combined are reported. When the sample size is large at $1,000$ or $2,000$, for both CITree and honest CITree, the PPV of T1 is high at a rate above 95%; the PPV of T2 and T3 + T4 is above 85%; the PPV of T5 is the lowest, where some subjects who belong to T2, T3 and T4 are misallocated to T5. When sample size is small ($N = 500$), the PPV of CITree is greater than that of the honest CITree for each predicted terminal. The subgroup benefit of the predicted terminal nodes is consistent with the PPVs, where subjects in T1 performs the best. The subgroup benefit of predicted T5 is lower than the true value of 2.0, which suggests that some subjects from low benefit nodes were misallocated to this terminal node.

In addition, we examined whether the correct variable was selected for splitting from level 2 to level 3. When the sample size is $1,000$, more than 95% times, the correct variable X_1 was selected and more than 95% times an adequate split point (true value -0.3, considered as adequate in the

range $[-0.5, -0.1]$) was identified. When sample size is $2,000$, both CITree and honest CITree can select the correct variable at an adequate cut point.

Regarding the computational speed, the honest CITree takes 15 seconds and CITree takes 1.5 minutes for $N = 1,000$ on a PC with 3.8GHz CPU. The honest CITree is much faster than the CITree since it only uses half of the sample in each step. They are both faster than QUINT (2 minutes), which used five bootstrapped samples in the bootstrap-based pruning procedure. The computation of searching the best HTB test for CITree is fast given the derived variance formula. The computational time of estimating ITR for CITree after level 1 is also fast since the sample size reduces dramatically in lower levels of the tree. QUINT was implemented via the R package "quint" (Dusseldorp et al., 2016). R code for IT is available at `http://biopharmnet.com/subgroup-analysis-software/`. R code for MIDAs was from the authors. R codes for fitting CITree are available upon request.

3.4 Application to the REVAMP Study

REVAMP (Kocsis et al., 2009) is a two-phase, 12-week randomized trial, which aimed to compare the efficacy of combining psychotherapy with medication to medication alone among chronic MDD patients who didn't fully respond to initial treatment of an antidepressant medication. Among 808 patients who entered phase I, 491 patients didn't achieve remission and therefore entered phase II. They were randomly assigned to receive (1) continued pharmacotherapy and augmentation with cognitive behavioral analysis system of psychotherapy (MEDS+CBASP), (2) continued pharmacotherapy and augmentation with brief supportive psychotherapy (MEDS+BSP), or (3) continued optimized pharmacotherapy (MEDS) alone with 2:2:1 ratio.

The clinical outcome of interest is the 24-item Hamilton Depression Rating Scale (HAM-D) total score at the end of phase II treatment. A lower HAM-D score indicates less symptoms and thus a better response. We compared two categories of treatments: 1)Medication plus psychotherapy (MEDS+THERAPY); and 2) Medication alone (MEDS). Five feature variables were used to estimate the optimal ITR: gender(Male=1; Female=0), age (median = 47), Quick Inventory of Depression Symptoms (QIDS, a score to measure the symptoms of depression, a lower score indicates better clinical outcome) score at the end of phase I (median = 8), the relative change of QIDS score over phase I (median = -0.44, calculated as change of QIDS score over phase I divided by baseline QIDS score), and current alcohol use (Yes=1). We estimated $\pi(A_i|X_i)$ from the data by the empirical proportions of the treatment. There were 418 participants with complete feature variable information included in our analysis, where 336 received medication and therapy, and 82 patients received medication only.

We applied CITree, IT, QUINT, and MIDAs to estimate the optimal ITR for MDD patients who did not achieve remission with phase I treatment. For all methods, we randomly split the sample into a training set and testing set with 1:1 ratio and repeated the procedure 100 times. QUINT did not identify any individual variable that has a qualitative interaction with the treatment, therefore it returns a non-personalized treatment rule, that is, to treat all patients with medication and therapy. The mean HAM-D score is 12.50 for MEDS+THERAPY and 12.81 for MEDS only group. Three other methods estimated an individualized treatment rule. Figure 3.6 compares the overall performance of CITree to IT and MIDAs. The ITR estimated by CITree yields an average HAM-D score of 12.17 (sd=1.41), which is smaller than IT (12.97, sd=1.33), and MIDAs (12.54, sd=0.95), indicating a better performance of CITree (Table 3.7). The overall ITR benefit estimated by CITree (0.96, sd=2.43) is much larger than that based on IT (-0.64, sd=2.26) and MIDAs (0.21, sd=2.48),

and also greater than QUINT (an overall benefit of 0.31). Both IT and MIDAs perform worse than the non-personalized rule.

We present the fitted CITree on the full data in Figure 3.7. Subjects are first partitioned into two groups with predicted optimal treatment as MEDS + THERAPY or MEDS only based on the sign of first level decision rule, $\widehat{f}_1(\boldsymbol{X})$. For the right node with MEDS only as the optimal treatment, subjects are then partitioned into a high benefit group (QIDS change> -0.21) and a low benefit group (QIDS change≤ -0.21) based on the relative change of QIDS score at phase 1. CITree stops growing for the high benefit node (T4) and re-fit $\widehat{f}_2(\boldsymbol{X})$ for the low benefit group and forms another two terminal nodes T2 and T3. On the right side of the tree, no heterogeneity in benefit was found and CITree stops growing after the first split. The final optimal treatment decision rule omitting variables with negligible coefficients can be summarized as $-I(\widehat{f}_1(\boldsymbol{X}) \leq 0) + I(\widehat{f}_1(\boldsymbol{X}) > 0)\Big[I(QIDs.change > -0.21) + I(QIDs.change \leq -0.21) \times \text{sign}(\widehat{f}_2(\boldsymbol{X}))\Big]$, where $\widehat{f}_1(\boldsymbol{X}) = -0.5 + alcohol$ and $\widehat{f}_2(\boldsymbol{X}) = 0.5 - gender$. For the first level ITR, current alcohol use dominates the optimal ITR, and in level 3 gender is the most important feature variable. Thus current alcohol use and gender are identified to manifest a qualitative interaction. A recent study by Gunter et al. (2011) also showed alcohol dependence is useful for selecting treatments in a Nefazodone CBASP Trial for major depression. Relative change of QIDS score manifests a quantitative interaction to distinguish patients with a high and low benefit.

CITree divides patients into four terminal nodes. For subjects who don't use alcohol (T1), adding therapy is more beneficial than treated by medication alone. For those with current alcohol use and phase I QIDS relative change > -0.21, medication alone is more beneficial compared to adding therapy. Males with current alcohol use and larger relative change of phase I QIDS decrease are more likely to achieve a better treatment outcome (lower HAMD score) with

MEDS+THERAPY; while females with current alcohol use and larger relative change of phase I QIDS decrease are more likely to perform better by MEDs only.

3.5 Discussion

In this work, we propose a tree-based learning method (CITree) to estimate interpretable ITRs and simultaneously identify subgroups with large benefit to guide treatment decision making for the patients. CITree performs qualitative interaction and quantitative interaction splits and retains the re-fitted linear rule in each node only if the new rule improves the subgroup and overall benefit, so it is guaranteed to outperform ABLO. The proposed algorithm fits a linear rule in each node to estimate an ITR and thus the resulting rule is piece-wise linear. When desirable, polynomial rule can also be considered. We have shown that using a linear combination of feature variables may be more powerful in identifying qualitative interaction comparing to exploring a single variable in turn. QUINT and IT do not have this feature, which may have contributed to their inferior performance both in simulations and real data application. Other possible reasons for their inadequate performance include using surrogate objective function for splitting criterion (e.g., t-test) and imbalanced treatment group size.

The composite interaction tree method has several limitations. First, it is slower comparing to ABLO, since it requires re-fitting ABLO at different child nodes. Faster computational algorithm is worth investigating in the future, e.g., parallel computing for finding child nodes in different parent nodes. In addition, like other tree-based methods, the composite interaction tree could be unstable depending on the first few splits (although the fitted tree was fairly stable in REVAMP example). An extension is to consider ensemble methods in line with random forest. However, such methods

may lose the advantage on interpretability. Lastly, an important extension is to develop a dynamic tree-based method for multi-stage studies such as sequential multi-stage randomization trials.

Table 3.3: Simulation study 2 results: classification accuracy among subjects truly belong to each terminal node

Setting	N	Method	T1	T2	T3	T4	T5
Setting 5	500	QUINT	0.981 (0.044)	0.374 (0.337)	0.665 (0.343)	0.784 (0.280)	0.865 (0.061)
		MIDAs	0.932 (0.056)	0.315 (0.170)	0.779 (0.174)	0.803 (0.151)	0.912 (0.044)
		IT	0.949 (0.038)	0.591 (0.157)	0.642 (0.270)	0.787 (0.185)	0.927 (0.049)
		Honest CITree	0.969 (0.018)	0.740 (0.122)	0.674 (0.215)	0.755 (0.193)	0.965 (0.021)
		CITree	0.974 (0.015)	0.788 (0.099)	0.841 (0.143)	0.932 (0.100)	0.975 (0.014)
	1000	QUINT	0.990 (0.025)	0.283 (0.357)	0.753 (0.356)	0.856 (0.265)	0.870 (0.049)
		MIDAs	0.956 (0.036)	0.252 (0.143)	0.854 (0.131)	0.866 (0.113)	0.924 (0.038)
		IT	0.968 (0.022)	0.685 (0.102)	0.666 (0.206)	0.853 (0.099)	0.941 (0.025)
		Honest CITree	0.973 (0.014)	0.765 (0.099)	0.835 (0.141)	0.928 (0.124)	0.975 (0.014)
		CITree	0.978 (0.012)	0.802 (0.076)	0.892 (0.094)	0.974 (0.039)	0.977 (0.012)
	2000	QUINT	0.996 (0.008)	0.242 (0.352)	0.755 (0.372)	0.907 (0.223)	0.873 (0.045)
		MIDAs	0.975 (0.021)	0.200 (0.124)	0.872 (0.125)	0.899 (0.090)	0.936 (0.023)
		IT	0.982 (0.013)	0.742 (0.074)	0.720 (0.155)	0.886 (0.066)	0.953 (0.019)
		Honest CITree	0.979 (0.010)	0.795 (0.088)	0.898 (0.081)	0.970 (0.056)	0.976 (0.015)
		CITree	0.974 (0.011)	0.822 (0.070)	0.787 (0.170)	0.916 (0.116)	0.980 (0.011)
Setting 6	500	QUINT	0.978 (0.047)	0.437 (0.350)	0.605 (0.359)	0.685 (0.333)	0.846 (0.079)
		MIDAs	0.911 (0.049)	0.304 (0.166)	0.780 (0.191)	0.791 (0.154)	0.895 (0.046)
		IT	0.933 (0.046)	0.594 (0.180)	0.603 (0.268)	0.756 (0.217)	0.913 (0.062)
		Honest CITree	0.966 (0.020)	0.725 (0.119)	0.594 (0.204)	0.706 (0.205)	0.961 (0.028)
		CITree	0.973 (0.013)	0.747 (0.112)	0.773 (0.169)	0.875 (0.158)	0.965 (0.022)
	1000	QUINT	0.988 (0.036)	0.325 (0.359)	0.729 (0.353)	0.838 (0.268)	0.864 (0.061)
		MIDAs	0.941 (0.040)	0.248 (0.155)	0.820 (0.160)	0.842 (0.132)	0.925 (0.034)
		IT	0.968 (0.020)	0.659 (0.118)	0.643 (0.228)	0.835 (0.116)	0.936 (0.029)
		Honest CITree	0.970 (0.017)	0.750 (0.103)	0.772 (0.175)	0.873 (0.155)	0.966 (0.018)
		CITree	0.975 (0.013)	0.790 (0.096)	0.879 (0.082)	0.961 (0.053)	0.975 (0.014)
	2000	QUINT	0.995 (0.013)	0.236 (0.344)	0.763 (0.357)	0.877 (0.263)	0.879 (0.039)
		MIDAs	0.963 (0.029)	0.200 (0.124)	0.854 (0.129)	0.898 (0.081)	0.938 (0.023)
		IT	0.981 (0.013)	0.719 (0.091)	0.693 (0.176)	0.869 (0.089)	0.950 (0.017)
		Honest CITree	0.976 (0.011)	0.793 (0.074)	0.876 (0.086)	0.962 (0.053)	0.973 (0.016)
		CITree	0.974 (0.012)	0.816 (0.065)	0.848 (0.123)	0.946 (0.064)	0.978 (0.010)

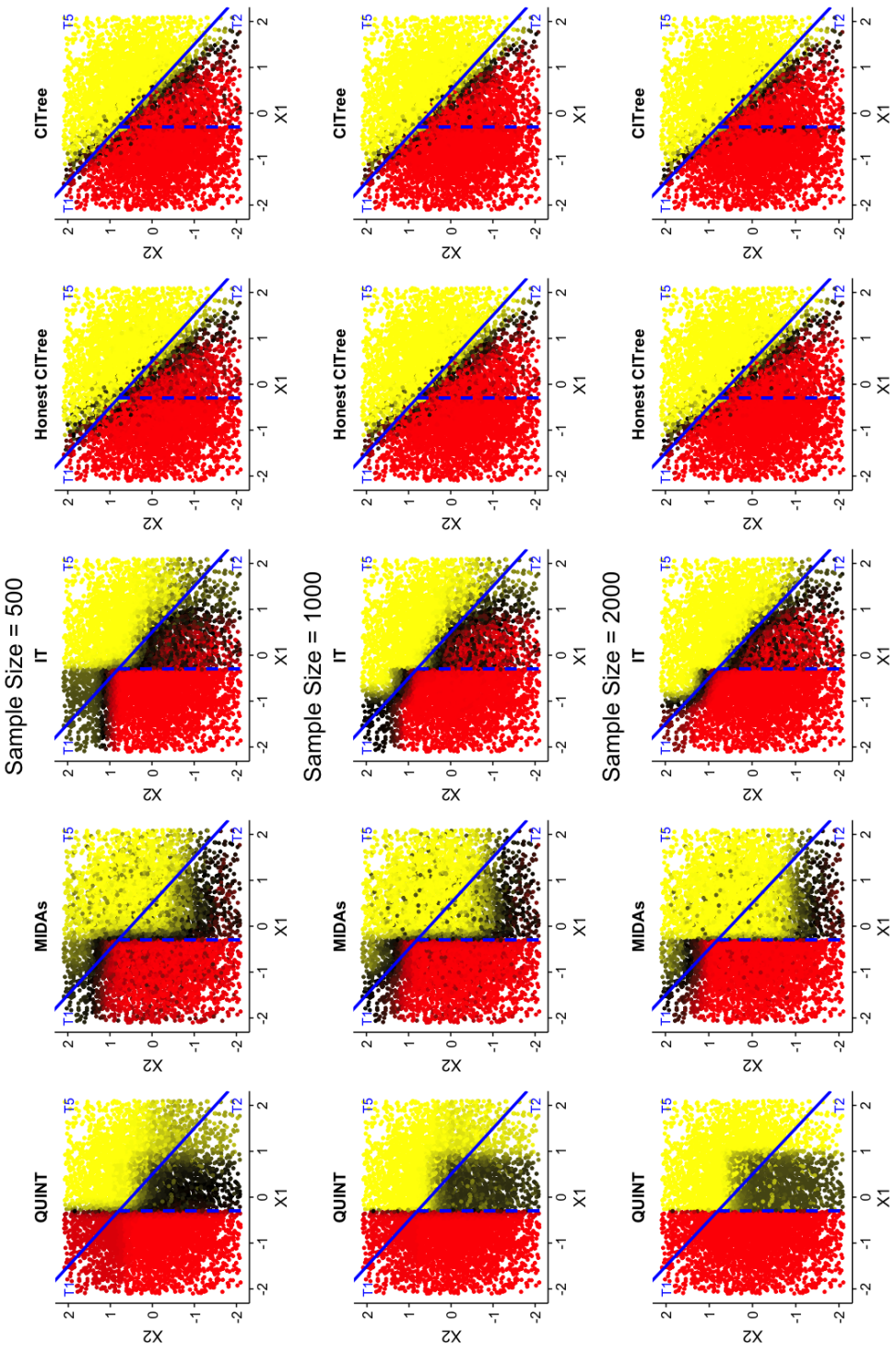
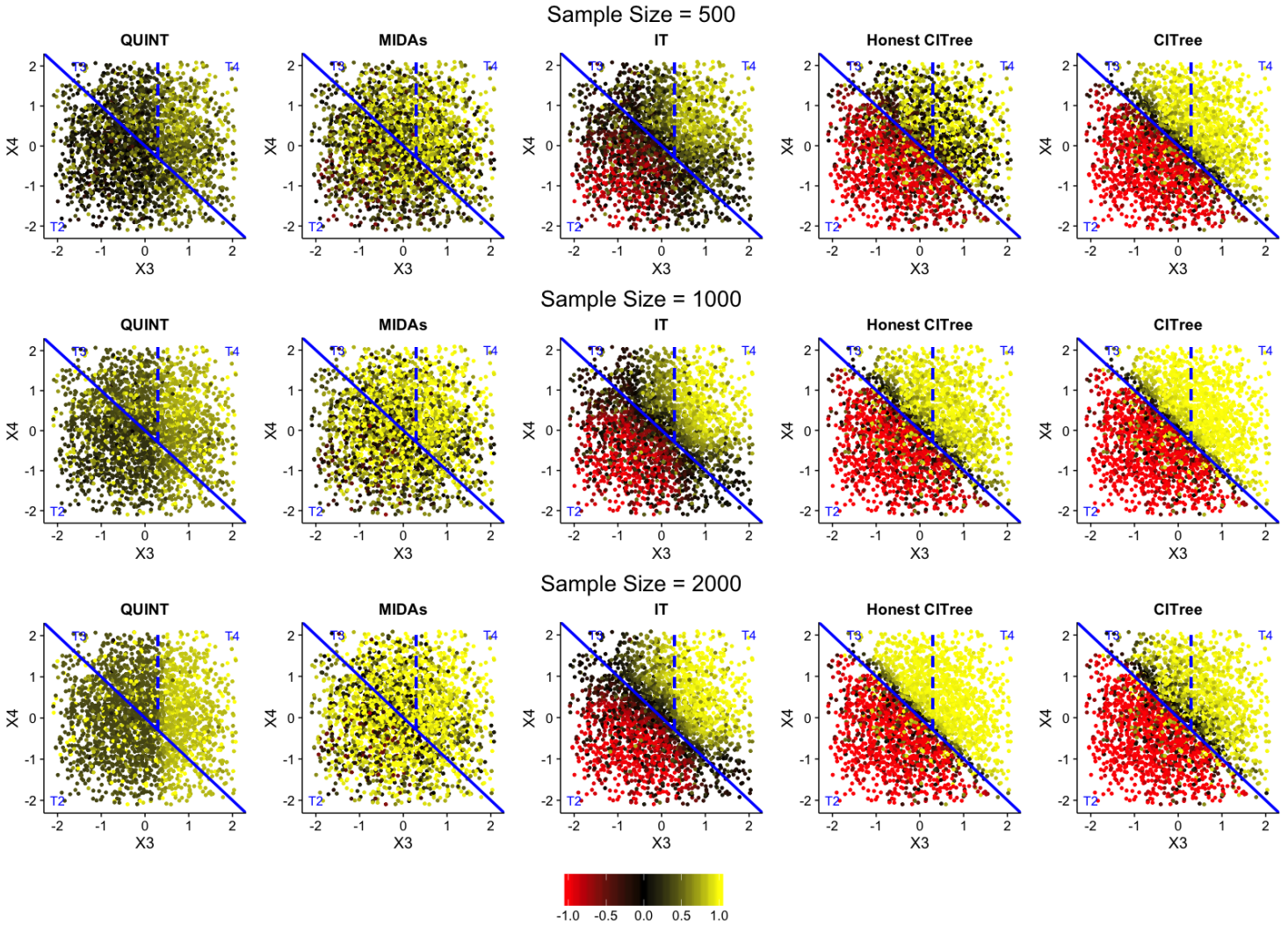


Figure 3.4: Simulation 2 results for subjects in terminal nodes T1, T2, and T5*

*: The blue solid line is the true optimal treatment boundary for qualitative interaction. The blue dashed line is the true boundary of quantitative interaction (high vs low ITR benefit). The color of each subject represents the average predicted optimal treatment for a subject in the validation set based on the ITR fitted from 100 simulated data sets. The true optimal treatment of subjects below the blue solid line is -1 (red color) and above the line is 1 (yellow color). Subjects between the solid and dashed blue line is the subgroup most easily to be mis-allocated by a sub-optimal method.

Figure 3.5: Simulation 2 results for subjects in terminal nodes T2, T3, and T4*



*: The blue solid line is the true optimal treatment boundary for qualitative interaction. The blue dashed line is the true boundary of quantitative interaction (high vs low ITR benefit). The color of each subject represents the average predicted optimal treatment for a subject in the validation set based on the ITR fitted from 100 simulated data sets. The true optimal treatment of subjects below the blue solid line is -1 (red color) and above the line is 1 (yellow color). Subjects between the solid and dashed blue line is the subgroup most easily to be mis-allocated by a sub-optimal method.

Table 3.4: Simulation study 2 results: total nodes or levels of tree based methods

		QUINT	MIDAs	IT	Honest CITree		CITree	
Setting	N	Nodes	Nodes	Nodes	Levels	Nodes	Levels	Nodes
Setting 5	500	6.09 (0.93)	5.80 (0.91)	9.26 (1.54)	4.81 (0.87)	4.18 (1.18)	5.30 (0.61)	4.99 (1.03)
	1000	6.01 (1.00)	6.32 (1.03)	14.55 (2.13)	5.34 (0.54)	5.22 (1.09)	5.47 (0.72)	5.65 (1.55)
	2000	5.89 (0.99)	6.58 (0.79)	22.60 (3.86)	5.42 (0.61)	5.72 (1.48)	5.68 (0.91)	6.47 (1.84)
Setting 6	500	6.10 (0.85)	6.22 (0.89)	9.95 (1.86)	4.73 (1.00)	4.30 (1.55)	5.41 (0.89)	5.43 (1.67)
	1000	5.98 (1.08)	6.69 (0.99)	16.05 (2.89)	5.34 (0.86)	5.20 (1.36)	5.56 (0.80)	5.73 (1.52)
	2000	5.64 (0.98)	6.98 (0.90)	24.13 (4.27)	5.57 (0.76)	5.89 (1.58)	5.77 (1.13)	6.45 (2.01)

Table 3.5: Simulation study 2 results: PPVs of predicted terminal nodes by CITree

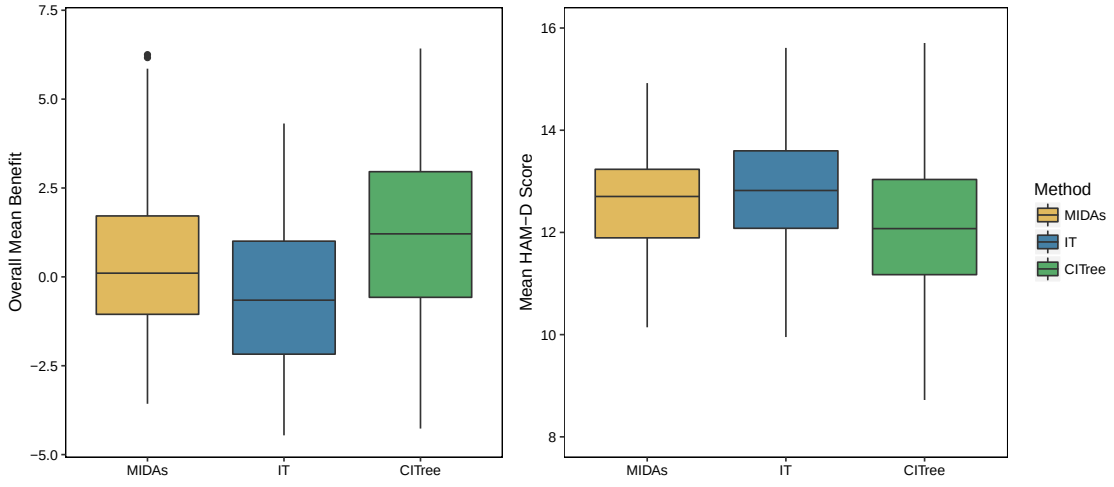
Setting	N	Method	T1	T2	T3+T4	T5
Setting 5	500	Honest CITree	0.919 (0.126)	0.761 (0.234)	0.832 (0.198)	0.748 (0.072)
		CITree	0.969 (0.056)	0.875 (0.137)	0.934 (0.063)	0.779 (0.072)
	1000	Honest CITree	0.970 (0.047)	0.871 (0.136)	0.945 (0.059)	0.759 (0.061)
		CITree	0.977 (0.037)	0.918 (0.096)	0.966 (0.040)	0.774 (0.060)
	2000	Honest CITree	0.980 (0.011)	0.930 (0.043)	0.970 (0.044)	0.768 (0.073)
		CITree	0.982 (0.010)	0.809 (0.156)	0.923 (0.074)	0.777 (0.071)
Setting 6	500	Honest CITree	0.911 (0.131)	0.700 (0.232)	0.800 (0.163)	0.740 (0.067)
		CITree	0.964 (0.050)	0.851 (0.100)	0.890 (0.092)	0.754 (0.063)
	1000	Honest CITree	0.956 (0.080)	0.850 (0.117)	0.904 (0.091)	0.753 (0.063)
		CITree	0.976 (0.041)	0.907 (0.057)	0.952 (0.042)	0.773 (0.069)
	2000	Honest CITree	0.977 (0.017)	0.915 (0.045)	0.961 (0.029)	0.767 (0.063)
		CITree	0.982 (0.008)	0.855 (0.138)	0.938 (0.063)	0.781 (0.061)

Table 3.6: Simulation study 2 results: subgroup benefits of predicted terminal nodes by CITree

Setting	N	Method	T1	T2	T3+T4	T5
Setting 5	500	Honest CITree	1.959 (0.400)	0.610 (0.406)	0.996 (1.132)	1.521 (0.311)
		CITree	1.964 (0.233)	0.957 (0.316)	0.811 (0.305)	1.638 (0.207)
	1000	Honest CITree	1.958 (0.225)	0.804 (0.357)	0.852 (0.427)	1.550 (0.203)
		CITree	1.957 (0.149)	0.928 (0.240)	0.768 (0.272)	1.598 (0.176)
	2000	Honest CITree	1.956 (0.149)	0.904 (0.246)	0.774 (0.289)	1.569 (0.200)
		CITree	1.977 (0.098)	0.932 (0.154)	0.631 (0.210)	1.591 (0.157)
Setting 6	500	Honest CITree	1.959 (0.558)	0.380 (0.531)	0.814 (0.666)	1.427 (0.296)
		CITree	1.990 (0.303)	0.893 (0.322)	0.766 (0.434)	1.513 (0.204)
	1000	Honest CITree	1.915 (0.323)	0.710 (0.338)	0.786 (0.389)	1.567 (0.271)
		CITree	1.964 (0.163)	0.903 (0.252)	0.804 (0.330)	1.630 (0.185)
	2000	Honest CITree	1.937 (0.173)	0.855 (0.267)	0.768 (0.310)	1.542 (0.181)
		CITree	1.983 (0.114)	0.940 (0.158)	0.718 (0.219)	1.601 (0.155)

*: True benefits of Setting 5 are 2.00 for T1, 0.91 for T2, 1.95 for T5 and 0.70 for T3+T4. True benefits of Setting 6 are 2.00 for T1, 0.89 for T2, 1.97 for T5 and 0.77 for T3+T4.

Figure 3.6: Overall performance of three methods in the REVAMP Study

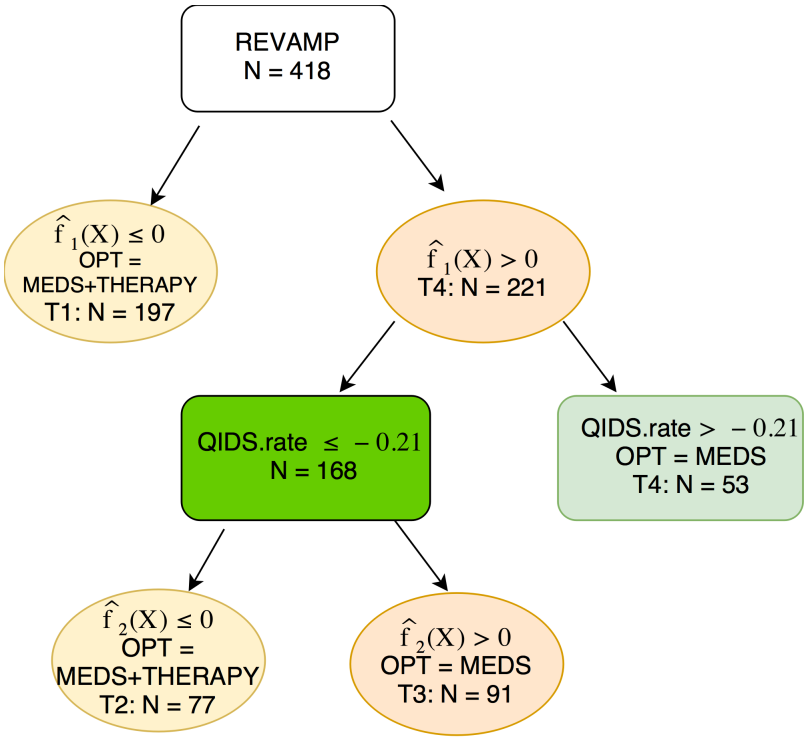


*: Distribution of estimated ITR benefit (the higher the better) and QIDS score (the lower the better) at the end of stage 2 treatment for three methods (based on 100 cross-validation)

Table 3.7: Overall performance of three methods in the REVAMP Study

Method	MIDAs	IT	CITree
Value	12.54 (0.95)	12.97 (1.33)	12.17 (1.41)
Benefit	0.21 (2.48)	-0.64 (2.26)	0.96 (2.43)

Figure 3.7: CITree for optimal individualized treatment decision (the REVAMP Study)



Chapter 4

Integrative Learning to Synthesize Individualized Treatment Rules Across Multiple Trials

4.1 Overview

In this chapter, we propose an integrative learning method to estimate the optimal individualized treatment rule by synthesizing evidence across multiple trials. In Section 4.2, we introduce the rationale and algorithm of the proposed integrative machine learning method. In Section 4.3, we derive the underlying Bayesian rules for the proposed method. In Section 4.4, we perform extensive simulation studies to show improvement as compared to existing machine learning methods for single studies. In Section 4.5, we apply the proposed method to the EMBARC trial using information from HEAL trial. Finally, we end this chapter with discussions and possible extensions in Section

4.6.

4.2 Methodologies

In Chapter 2 and 3, we focus on proposing methods for learning simple ITRs on a single randomized controlled trial. However, an ITR learned by a single study aims to optimize the performance on the study, which may not generalize to a future sample. A method that can use information from multiple studies may provide an ITR that represents a more general population and therefore has better reproducibility. Because not all of the same feature variables are collected in each study, one cannot directly combine study data and perform a single analysis. Taking weighted averages of the estimated ITR decision functions also does not apply because each ITR is obtained by maximizing the empirical value function, which has a complex, nonlinear relationship with the decision function. In this section, we propose novel integrative learning methods to combine ITRs from RCTs that collect different subsets of baseline feature variables.

4.2.1 Integrative Learning for High-Resolution ITR Using Coarsened ITRs

First consider how to enhance learning a high-resolution ITR dependent on a rich set of features in a comprehensive trial by using coarsened ITRs that have been learned from smaller trials with subsets of feature variables. Let $(\boldsymbol{X}_i, A_i, R_i), i = 1, ..., n$, be the richest set of baseline feature variables, treatments, and treatment outcomes collected on the ith participant in a comprehensive trial, denoted as study 0. The randomization probability of ith participant is denoted as $P(A_i|\boldsymbol{X}_i) = \pi_i$. For other K trials, only a subset of features, denoted by $\boldsymbol{Z}_k$, is collected by design. The observed data from these trials consist of $(\boldsymbol{Z}_{jk}, A_{jk}, R_{jk})$ for the jth patient in the kth trial, and $P(A_{jk}|\boldsymbol{Z}_{jk}) = \pi_{jk}$ for its randomization probability, with $j = 1, ..., n_k, k = 1, ..., K$.

The richest feature set $\boldsymbol{X}$ defines the high-resolution ITR (denoted by $\text{sign}(f_0(\boldsymbol{X}))$), in terms of conditional response given $\boldsymbol{X}$. Features $\boldsymbol{Z}_k$ as subsets of $\boldsymbol{X}$ define coarser, low-resolution ITR (denoted by $\text{sign}(f_k(\boldsymbol{Z}_k))$), in terms of conditional response given $\boldsymbol{Z}_k$. Our key rationale is that each low-resolution f_k still maintains certain optimality when applied to patients in the comprehensive trial but only using features available in $\boldsymbol{Z}_k$. Although f_k is fit by maximizing the average value of the conditional response given coarsened groups (stratified by conditioning on $\boldsymbol{Z}_k$), such that it may not be optimal for every individual in the finest group stratified by $\boldsymbol{X}$, f_k is informative of f_0 in the following sense: randomly select a patient from subpopulations stratified by different values of $\boldsymbol{Z}_k$, and apply to this patient the finest optimal ITR f_0 based on $\boldsymbol{X}$. We should expect that most times, the treatment recommendations from f_0 should be the same as those based on f_k because f_k should maximize the conditional response given $\boldsymbol{Z}_k$ averaged over all subjects (definition of value function). Based on this rationale, we propose the following method to integrate evidence from the smaller/auxiliary trials to improve learning the optimal ITR from the comprehensive trial.

First, for $k = 1, ..., K$, we use outcome weighted learning (e.g. ABLO) and kth study data to obtain the optimal rule dependent on the kth study features, denoted as $\text{sign}(\widetilde{f}_k(\boldsymbol{Z}_k))$. The corresponding benefit function is denoted by $\tilde{\delta}(\tilde{f}_k(\boldsymbol{Z}_k))$. To simplify the notation of benefit function, we will use $\widetilde{\delta}_k(\boldsymbol{Z}_k)$ to represent $\tilde{\delta}(\tilde{f}_k(\boldsymbol{Z}_k))$ in the following sections. That is, for any future subject including the subject in the high-resolution study (study 0) whose value for the feature variable $\boldsymbol{Z}_k$ is $\boldsymbol{z}_k$, the learned rule based on the kth study concludes that the optimal treatment assignment should be 1 if $\widetilde{f}_k(\boldsymbol{z}_k) > 0$ and -1, otherwise. Under the this optimal treatment assignment, the expected benefit is $\widetilde{\delta}_k(\boldsymbol{z}_k)$.

Hence, for each subject i in study 0, in addition to the observed trial data $(A_i, \boldsymbol{X}_i, R_i)$, we also know from the kth study evidence that his/her optimal treatment rule should not be significantly

different from $\text{sign}(\widetilde{f}_k(\boldsymbol{Z}_{ik}))$ and that the expected benefit should not be far from $\widetilde{\delta}_k(\boldsymbol{Z}_{ik})$, where $\boldsymbol{Z}_{ik}$ is the set of the feature variables used in the kth study so is part of $\boldsymbol{X}_i$. Next, we integrate this external information to learn the optimal treatment rule using the data from study 0. Specifically, we introduce an additional regularization to incorporate $\widetilde{f}_k$ as prior information in the learning approach. The regularized value function to be maximized to estimate f_0 is

$$\frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i}I(A_i f_0(\boldsymbol{X}_i)>0)+\left\{\sum_{k=1}^{K}\frac{\lambda_k}{n}\sum_{i=1}^{n}\frac{\widetilde{\delta}_{ik}}{\pi_i}I(f_0(\boldsymbol{X}_i)\widetilde{f}_k(\boldsymbol{Z}_{ik})>0)\right\}, \tag{4.1}$$

where λ_k is a tuning parameter to be selected from data, and $\widetilde{\delta}_{ik}=\widetilde{\delta}_k(\boldsymbol{Z}_{ik})$. A straightforward choice is to let δ_{ik} be the approximate benefit of using $\widetilde{f}_k$ to assign treatments available based on ABLO. Note that the first term in (4.1) is the empirical value function associated with f_0. The second term is a regularization term such that the more consistent the signs of f_0 and $\widetilde{f}_k$, the larger the value. Thus, the regularization ensures that the derived treatment rule is consistent with the learned rule from the kth trial.

Note that maximizing (4.1) is equivalent to maximizing

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{R_i}{\pi_i}I(A_i f_0(\boldsymbol{X}_i)>0)+\left\{\sum_{k=1}^{K}\frac{\lambda_k}{2}\frac{\widetilde{\delta}_{ik}}{\pi_i}\text{sign}(A_i f_0(\boldsymbol{X}_i))\text{sign}(A_i\widetilde{f}_k(\boldsymbol{Z}_{ik}))+\sum_{k=1}^{K}\frac{\lambda_k}{2}\frac{\widetilde{\delta}_{ik}}{\pi_i}\right\}\right],$$

which can be rewritten (up to a constant) as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{R}_i}{\pi_i}I(A_i f_0(\boldsymbol{X}_i)>0)$$

after defining $\widetilde{R}_i=R_i+\sum_{k=1}^{K}\lambda_k\widetilde{\delta}_{ik}\left\{2I(A_i\widetilde{f}_k(\boldsymbol{Z}_{ik})>0)-1\right\}$. Thus, maximizing (4.1) is equivalent to maximizing another empirical version of the value function, but the outcome for individual i is augmented as $\widetilde{R}_i$ instead of R_i. Hence, this optimization can be solved by minimizing a surrogate ramp loss function, as in ABLO. The tuning parameter λ_k can be selected by cross-validation with a maximal value function. $\widetilde{\delta}_{ik}$ can be estimated by parametric or non-parametric models using

the kth trial data and predicting the estimated benefit on the subjects of study 0. Specifically, in this work, we apply a linear regression model on the kth trial data with assigned treatment, baseline feature variables and their interactions to predict clinical outcome for each subject with both treatment 1 and -1 given its feature variables. The benefit is then estimated by the difference between the predicted outcome given the estimated optimal treatment and the predicted outcome given the alternative treatment for subjects in study 0. Note that λ_k can be tuned separately for different coarsened ITR k. In practice, since δ_{ik} will be estimated separately for each study k, one can keep λ_k the same across different studies to ease computational burden on choosing tuning parameters.

4.2.2 Integrative Learning for Coarsened ITRs Using High-Resolution ITR

In clinical settings with limited resources (e.g., when cost constraint does not permit the collection of brain imaging data), an ITR that depends only on easily assessed clinical variables can be practically useful. Thus, we propose an integrative learning approach to enhance coarsened ITRs by using evidence generated from existing studies with high-resolution ITR.

The key rationale is that the average benefit of the high-resolution ITR f_0 evaluated in a subgroup stratified by $\boldsymbol{Z}_k$ should be the same as the average benefit of the coarsened ITR f_k. First, we obtain the optimal treatment rule, denoted by $\text{sign}(\widetilde{f}_0(\boldsymbol{X}))$, and its benefit function, denoted by $\widetilde{\delta}_0(\boldsymbol{X})$, using study 0. To incorporate this information to improve the rule learning in study $k, k = 1, ..., K$. For each k, we consider estimating f_k by maximizing the following objective function similar to (4.1), but exchange the roles of f_0 and f_k as

$$\frac{1}{n_k}\sum_{j=1}^{n_k}\frac{R_{jk}}{\pi_{jk}}I(A_{jk}f_k(\boldsymbol{Z}_{jk})>0)+\frac{\lambda_k}{n}\sum_{i=1}^{n}\frac{\widetilde{\delta}_i}{\pi_i}I(f_k(\boldsymbol{Z}_{ik})\widetilde{f}_0(\boldsymbol{X}_i)>0), \tag{4.2}$$

where λ_k is a tuning constant and weight $\widetilde{\delta}_i = \widetilde{\delta}_0(\boldsymbol{X}_i)$. Again, δ_i can be the benefit of assigning treatment to subject i using $\widetilde{f}_0$ estimated from ABLO. In (4.2), the first component is the empirical value for each ITR f_k involving the kth trial data. The second component encourages f_k and $\widetilde{f}_0$ to yield the same treatment recommendation (same sign) as much as possible, especially for subjects with a large benefit $\widetilde{\delta}_i$.

To solve the optimization for kth trial, we can rewrite (4.2), up to a constant, as

$$\frac{1}{n_k}\sum_{j=1}^{n_k}\frac{R_{jk}}{\pi_{jk}}I(A_{jk}f_k(\boldsymbol{Z}_{jk})>0)+\frac{\lambda_k}{n}\sum_{i=1}^{n}\frac{\widetilde{\delta}_i}{\pi_i}\Big(2I(A_i\widetilde{f}_0(\boldsymbol{X}_i)>0)-1\Big)\,I(A_if_k(\boldsymbol{Z}_{ik})>0).$$

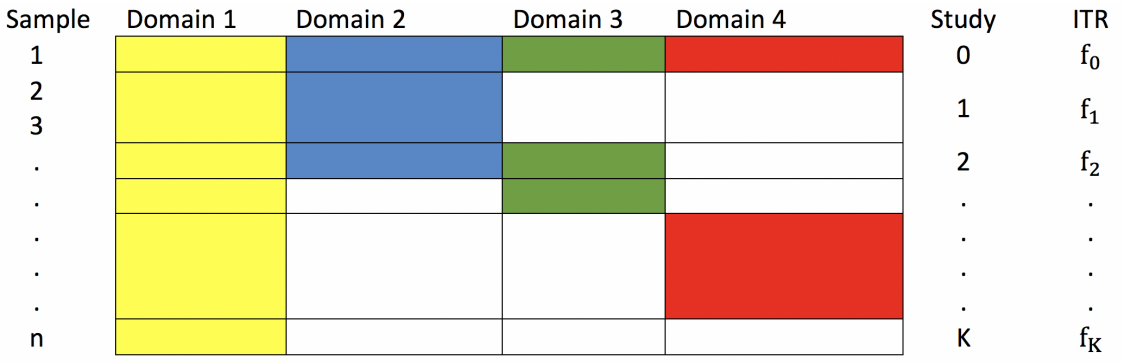
We again augment the original data set by combining data from both study k and study 0; however, for subject i in study 0, instead of using R_i as the reward outcome in the augmentation term, we use $\tilde{R}_i = \lambda_k n_k n^{-1}\widetilde{\delta}_i(2I(A_i\widetilde{f}_0(\boldsymbol{X}_i)>0)-1)$, which is less variable than R_i. For optimization, we just augment the data from kth study, $(\boldsymbol{Z}_{jk}, A_{jk}, R_{jk})$, with additional data from study 0 with $(\boldsymbol{Z}_{ik}, A_i, \tilde{R}_i)$. Therefore, we can apply the standard learning method to learn the optimal ITR using this integrated data set. Additionally, λ_k weighs the importance of this augmentation and will be chosen by cross-validation.

4.2.3 Extension to Blockwise Feature Domains in a Single Trial

With the development of new data collection techniques, multiple sources of data may be collected from different domains. In practice, it is common that the data entails blockwise feature entries (Xiang et al., 2013), due to cost or primitive protocol of data collection. For example, in the EMBARC trial, some subjects have complete clinical, neuropsychiatric measures and brain imaging measures, while others are lack of the entire domain of brain imaging data.

Assume that the blockwise feature pattern is completely random and noninformative. The data

Figure 4.1: An example for data collected with blockwise feature domains



*: white color indicates the measures for that domain are missing, other colors indicate the measures that are complete. In this case, $\boldsymbol{X}$ includes feature variables from domain 1, 2, 3, and 4; $\boldsymbol{Z}_1$ includes variables from domain 1 and 2; $\boldsymbol{Z}_2$ includes variables from domain 1, 2, and 3; $\boldsymbol{Z}_K$ includes only variables from domain 1.

from this single study can be treated as collected from multiple trials based on the domains of feature variables that are collected. More specifically, study 0 consists all subjects with complete feature variable measures, and subjects that have the same block of feature variables measured will be considered as from an independent trial. In this way, the total number of small studies, K, is the total number of block patterns. For example, in Figure 4.1, there are K sub-studies. In study 0 with complete feature variables $\boldsymbol{X}$, using methods in Section 4.2.1, one can learn a high-resolution ITR $\widetilde{f}_0(\boldsymbol{X})$; while in study k, a coarsened ITR $\widetilde{f}_k(\boldsymbol{Z}_k)$ will be estimated given the available covariates $\boldsymbol{Z}_k$. Compared to using study data 0 alone, the integrative learning offers an opportunity to use all collected data efficiently. Hence, our proposed method to integrate treatment rules learned in K studies can be used to improve learning optimal treatment rule using the data from study 0.

4.3 Theoretical Results

In this section, we derive the underlying Bayesian rules of the integrative learning methods. To this end, we assume that the asymptotic limits of $\widetilde{f}_k(\boldsymbol{Z}_k)$ and $\widetilde{\delta}_k(\boldsymbol{Z}_k)$ in (4.1) are, $f_k(\boldsymbol{Z}_k)$ and $\delta_k(\boldsymbol{Z}_k)$,

respectively. Then for given λ_k's, the Bayesian rule for the integrative learning for high-resolution ITR maximizes

$$E_0\left[\left\{R+\sum_{k=1}^{K}\lambda_k\delta_k(\boldsymbol{Z}_k)\left(2I(Af_k(\boldsymbol{Z}_k)>0)-1\right)\right\}\frac{I(Af_0(\boldsymbol{X})>0)}{\pi(A|\boldsymbol{X})}\right]$$

over $f_0(\boldsymbol{X})$, where $E_0(\cdot)$ denotes the expectation under the joint distribution of $(R,\boldsymbol{X})$ in study 0, and $\pi(a|\boldsymbol{X})=P(A=a|\boldsymbol{X})$ in study 0. Using the standard result from the outcome weighted learning (Zhao et al., 2012), we immediately obtain that the Bayesian rule, denoted by $f_0^*(\boldsymbol{X})$, satisfies

$$\begin{aligned}
&\operatorname{sign}(f_0^*(\boldsymbol{X}))\\
=\;&\operatorname{sign}\left(E_0\left\{R+\sum_{k=1}^{K}\lambda_k\delta_k(\boldsymbol{Z}_k)\left(2I(Af_k(\boldsymbol{Z}_k)>0)-1\right)\Big|A=1,\boldsymbol{X}\right\}\right.\\
&\left.-E_0\left\{R+\sum_{k=1}^{K}\lambda_k\delta_k(\boldsymbol{Z}_k)\left(2I(Af_k(\boldsymbol{Z}_k)>0)-1\right)\Big|A=-1,\boldsymbol{X}\right\}\right)\\
=\;&\operatorname{sign}\left(E_0\left\{R+\sum_{k=1}^{K}\lambda_k\delta_k(\boldsymbol{Z}_k)\left(2I(f_k(\boldsymbol{Z}_k)>0)-1\right)\Big|A=1,\boldsymbol{X}\right\}\right.\\
&\left.-E_0\left\{R+\sum_{k=1}^{K}\lambda_k\delta_k(\boldsymbol{Z}_k)\left(2I(f_k(\boldsymbol{Z}_k)<0)-1\right)\Big|A=-1,\boldsymbol{X}\right\}\right)\\
=\;&\operatorname{sign}\left(E_0[R|A=1,\boldsymbol{X}]-E_0[R|A=-1,\boldsymbol{X}]+\sum_{k=1}^{K}\lambda_k\delta_k(\boldsymbol{Z}_k)\operatorname{sign}(f_k(\boldsymbol{Z}_k))\right).
\end{aligned}$$

Note that $\delta_k(\boldsymbol{Z}_k)\operatorname{sign}(f_k(\boldsymbol{Z}_k))=E_0[R_k|A_k=1,\boldsymbol{Z}_k]-E_0[R_k|A_k=-1,\boldsymbol{Z}_k]$. Thus, the Bayesian rule is

$$\operatorname{sign}\left(E_0[R|A=1,\boldsymbol{X}]-E_0[R|A=-1,\boldsymbol{X}]+\sum_{k=1}^{K}\lambda_k(E_0[R_k|A_k=1,\boldsymbol{Z}_k]-E_0[R_k|A_k=-1,\boldsymbol{Z}_k])\right).$$

In other words, the Bayesian rule is to essentially combine all conditional treatment effects given $\boldsymbol{Z}_k$ from all the auxiliary studies with treatment effect given $\boldsymbol{X}$. Here, λ_k is used to weigh the effect size from the kth study. The first term has the same sign as the theoretical optimal high-resolution ITR using the rich feature set $\boldsymbol{X}$. The second term aggregates coarsened treatment

recommendations and benefits that are fitted from the smaller feature sets $\boldsymbol{Z}_k$. Therefore, the integrated ITR converges to an ITR that prescribes treatment with larger weighted conditional benefits across trials. Thus, in this sense the integrated ITR borrows information from all trials. Under the assumption that the direction of the expected treatment difference given $\boldsymbol{X}$ in study 0 is the same as the average treatment difference given $\boldsymbol{Z}_k$, $\widetilde{f}_0$ is Fisher consistent.

Furthermore, we remark that when λ_k is close to zero, the derived Bayesian rule approximates the true optimal treatment rule. However, in finite samples of study 0, the optimal rule estimated from one single study is likely to be different from the true optimal rule for some subjects due to limited information available to learn a fully nonparametric rule. Therefore, by allowing λ_k to be non-zero and determined in a data-driven way, one can use additional information from other studies to correct the finite sample bias in study 0 and improve the precision by including more samples available in auxiliary studies.

Using the same argument, we obtain that the Bayesian rule for low-resolution study, say study k, maximizes

$$\begin{aligned} & E_k\left[\frac{R_k}{\pi_k(A_k|\boldsymbol{Z}_k)} I(A_k f_k(\boldsymbol{Z}_k)>0)\right] + \lambda_k E_0\left[\frac{\delta_0(\boldsymbol{X})}{\pi(A|\boldsymbol{X})}\text{sign}(Af_0(\boldsymbol{X}))I(Af_k(\boldsymbol{Z}_k)>0)\right] \\ = \; & E_k\left[\frac{R_k}{\pi_k(A_k|\boldsymbol{Z}_k)} I(A_k f_k(\boldsymbol{Z}_k)>0)\right] + \lambda_k E_k\left[\frac{\delta_0(\boldsymbol{X})}{\pi(A|\boldsymbol{X})}\frac{g_0(\boldsymbol{Z}_k)}{g_k(\boldsymbol{Z}_k)}\text{sign}(Af_0(\boldsymbol{X}))I(Af_k(\boldsymbol{Z}_k)>0)\right], \end{aligned}$$

where $E_k(\cdot)$ is the expectation with respect to $(R_k, \boldsymbol{Z}_k)$ in study k, $g_0(\cdot)$ and $g_k(\cdot)$ are the density functions for $\boldsymbol{Z}_k$ in study 0 and study k, respectively. This expression, up to a constant independent of f_k, is equivalent to,

$$\begin{aligned} & E_k\left[\{E_k[R_k|A_k=1,\boldsymbol{Z}_k]-E_k[R_k|A_k=-1,\boldsymbol{Z}_k]\}\,\text{sign}(f_k(\boldsymbol{Z}_k))\right] \\ + \; & \lambda_k E_k\left[\delta_0(\boldsymbol{X})\frac{g_0(\boldsymbol{Z}_k)}{g_k(\boldsymbol{Z}_k)}\text{sign}(f_0(\boldsymbol{X}))\text{sign}(f_k(\boldsymbol{Z}_k))\right]. \end{aligned}$$

Therefore, the Bayesian rule for this integrative learning is

$$\begin{aligned}\text{sign}(f_k^*(\boldsymbol{Z}_k)) &= \text{sign}\Big(E_k[R_k|A_k=1,\boldsymbol{Z}_k]-E_k[R_k|A_k=-1,\boldsymbol{Z}_k] \\ &+ \lambda_k\frac{g_0(\boldsymbol{Z}_k)}{g_k(\boldsymbol{Z}_k)}(E_0[R|A=1,\boldsymbol{Z}_k]-E_0[R|A=-1,\boldsymbol{Z}_k])\Big),\end{aligned}$$

which is a weighted combination of the conditional treatment effects given $\boldsymbol{Z}_k$ from study k and study 0. In particular, λ_k gives the weight to integrate evidence from study 0. Additionally, if the conditional treatment effects are in the same direction or even have the same magnitude across the trials, the above Bayesian rule is the same as the optimal rule for study k. However, using the combined treatment effects to estimate the integrative rule has a better precision in finite sample due to including more data from study 0.

As a remark, $\delta_k(\boldsymbol{Z}_k)$ or $\delta(\boldsymbol{X})$ can be replaced by any other functions of the feature variables, or even a constant of one. In this case, the Bayesian rules are not necessarily based on the combined conditional treatment effects, but is the combination of the treatment effect in the study of interest and the evidence from other trials depending on the optimal rules learned in those trials.

4.4 Simulations

4.4.1 Simulation Design

The simulation model is inspired by real-world applications where multiple RCTs are conducted at different locations or different time (Justice et al., 1999). In the simulation studies, the patient population consists of a mixture of patients. For all trials, we assume that within each mixing subgroup the conditional distribution of R given $\boldsymbol{X}$ is the same, but the mixing proportion is different across trials. For example, a trial conducted in one location may consist of more elderly patients than another trial in another location, but the distribution of outcome given age and other

feature variables follow the same distribution within the subgroup of elder or younger patients. In the simulation model, three feature variables, (X_1, X_2, X_3), informative of optimal treatment choice, were generated from four latent subgroups of subjects with probabilities (p_1, p_2, p_3, p_4). Within each subgroup, X_1, X_2, and X_3 were independently simulated from a normal distribution with different means, $(1, 0.5, -1, -0.5)$, and standard deviation of one. The treatment for each subject was randomly assigned to 1 or -1 with equal probability. The clinical outcome for subjects in the kth subgroup was generated by

$$R = 1 + I(A = 1)(\delta_{1k} + \alpha_{1k} * W) + I(A = -1)(\delta_{2k} + \alpha_{2k} * W) + \alpha_{1k} * S + 0.5 * V * A + \epsilon,$$

where $\epsilon \sim N(0, 0.25)$, V, W, and S are i.i.d. and follow the standard normal distribution, $\boldsymbol{\delta} = [\delta_{lk}]_{2*4} = \begin{bmatrix} 0.8 & 0.3 & 0 & 0 \\ 0 & 0 & 0.8 & 0.3 \end{bmatrix}$, $\boldsymbol{\alpha} = [\alpha_{lk}]_{2*4} = \begin{bmatrix} 1 & 0.6 & 0.5 & 0.3 \\ 0.5 & 0.3 & 1 & 0.6 \end{bmatrix}$. Treatment $A = 1$ has a greater average effect for subjects in subgroups 1 and 2, and the alternative treatment -1 has a greater average effect in subgroups 3 and 4. Here, W is an observed prescriptive variable within each subgroup, V has qualitative interaction with treatment and therefore it is directly informative of the selection of optimal treatment, and S is an unobserved prognostic variable with the same main effect within each subgroup. One noise variable N_1 not contributing to the clinical outcome was independently generated from the standard normal distribution. In order to apply our method, we generated one high-resolution data set (data set 0), including all the observed feature variables $\boldsymbol{X} = (X_1, X_2, X_3, W, V, N_1)$, and one low-resolution data set (data set 1), with only $\boldsymbol{Z}_1 = (X_1, X_2, X_3, V)$. To mimic real-world situation, we set the probabilities of belonging to each latent subgroups as $(0.4, 0.1, 0.4, 0.1)$ for data set 0, and $(0.2, 0.3, 0.2, 0.3)$ for data set 1.

In order to test the generalizability of the proposed method, we evaluate the estimated ITR on different target populations, aiming to simulate different settings of patient recruitment in

the future target population where the ITR will be implemented. The total population consists of a mixture of patients, with the same conditional distribution of R given $\boldsymbol{X}$ within subgroup as described above. Each setting has mixing proportion, specifically, the ratio of the validation settings are $(0.25, 0.25, 0.25, 0.25)$, $(0.2, 0.2, 0.3, 0.3)$, $(0.5, 0, 0.5, 0)$, and $(0, 0.5, 0, 0.5)$ for settings 1 to 4. Validation setting 3 and 4 represent extreme cases when only subgroups with large or small benefits are collected. To maintain the simplicity and interpretability of the ITRs, we only consider linear rules in the following.

4.4.2 Simulation Results

4.4.2.1 Integrative learning for improving high-resolution ITR using coarsened ITRs

In this section, we evaluate the performance of the integrative learning for high-resolution ITR improved by coarsened ITRs. For each simulation, data set 1 is used to learn a coarsened ITR with only 4 variables, namely X_1, X_2, X_3, and V. The ITR learned by integrative learning uses all the 6 observed feature variables $(X_1, X_2, X_3, W, V, N_1)$ and includes the coarsened ITR as prior information in the learning algorithm. The weight $\widetilde{\delta}_{i1}$ of subject i is estimated by a linear regression model using the data set 1 including treatment assignment, feature variables and their interactions, and predict the estimated benefit on the study 0. In the real data applications, when estimating the benefit is not possible (e.g., only the coarsened ITR is available but not the individual-level data of study 1), we assign the same weight for each individual in the high-resolution data set, which is equivalent to setting δ as one for each subject. We compared the integrative learning results (both weighted by estimated benefit and equally weighted) to ABLO using only data set 0 (ABLO-H) with feature variables $\boldsymbol{X}$ or data set 1 (ABLO-L) with feature variables $\boldsymbol{Z}_1$. The sample size was 200 for both data set 0 and 1. Cross validation was used to select the tuning parameters λ.

To evaluate the performance of ITRs, we simulate independent validation data with four different settings as described above, each setting with $N = 10,000$. We report the overall benefit, value and classification accuracy on the four validation settings.

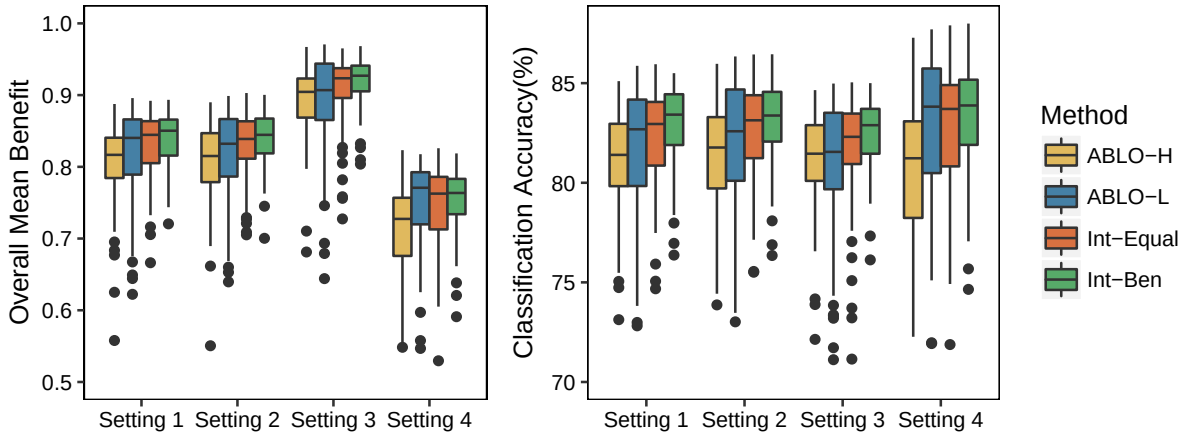
Table 4.1: Mean and standard deviation of overall benefit, value and accuracy rates using integrative learning for high-resolution ITRs comparing to ITRs by ABLO on single studies

		ABLO-H	ABLO-L	Int-Equal	Int-Ben
Benefit	Setting 1	0.806(0.053)	0.819(0.063)	0.830(0.045)	0.841(0.036)
	Setting 2	0.807(0.052)	0.818(0.060)	0.830(0.045)	0.841(0.035)
	Setting 3	0.894(0.046)	0.893(0.068)	0.909(0.047)	0.920(0.031)
	Setting 4	0.711(0.068)	0.747(0.065)	0.745(0.055)	0.752(0.045)
Value	Setting 1	0.668(0.027)	0.675(0.033)	0.680(0.023)	0.686(0.019)
	Setting 2	0.665(0.026)	0.670(0.030)	0.676(0.022)	0.682(0.018)
	Setting 3	0.833(0.024)	0.832(0.035)	0.840(0.025)	0.846(0.017)
	Setting 4	0.501(0.034)	0.519(0.032)	0.518(0.028)	0.522(0.022)
Accuracy	Setting 1	0.810(0.027)	0.817(0.033)	0.823(0.024)	0.830(0.019)
	Setting 2	0.812(0.028)	0.820(0.033)	0.825(0.025)	0.831(0.020)
	Setting 3	0.812(0.022)	0.808(0.038)	0.818(0.026)	0.825(0.017)
	Setting 4	0.806(0.037)	0.827(0.039)	0.827(0.032)	0.832(0.028)

*: ABLO-H is estimated by ABLO using high-resolution data only;
ABLO-L is estimated by ABLO using low-resolution data only;
Int-Equal is estimated by integrative learning for high-resolution ITR with equal weights for all subjects;
Int-Ben is estimated by integrative learning for high-resolution ITR weighted by estimated benefit.

Results from 100 replicates are summarized in Table 4.1 and Figure 4.2. For all the validation settings, ITRs estimated by integrative learning using both trial data outperform ITRs learned by ABLO using single trial data with larger optimal treatment selection accuracy, larger overall benefit and estimated value function. With more information used to estimate an ITR, the ITRs by integrative learning are more likely to be generalized to a future sample. Also, integrative learning weighted by estimated benefit performs better than using equal weights in the simulation since the estimated benefits are also useful information in learning the ITR. By integrating useful

Figure 4.2: Overall ITR benefit and optimal treatment allocation accuracy for the four methods.



*: ABLO-H is estimated by ABLO using high-resolution data only;
ABLO-L is estimated by ABLO using low-resolution data only;
Int-Equal is estimated by integrative learning for high-resolution ITR with equal weights for all subjects;
Int-Ben is estimated by integrative learning for high-resolution ITR weighted by estimated benefit.

information, the benefit and value function estimated by integrative learning are also more efficient compared to ABLO with smaller standard deviations.

4.4.2.2 Integrative learning for improving coarsened ITRs using high-resolution ITR

In this section, we evaluate the performance of the integrative learning for coarsened ITR using high-resolution ITR. The simulation data sets we generated are exactly the same as the above section. The only difference is that the ITRs learned by integrative learning is a low-resolution ITR without variable W and N_1. We assume the those two variables are not observed in the validation settings. Therefore, the high-resolution ITR cannot be used to predict the optimal ITR in the validation set. For each simulation, data set 0 is used to learn a high-resolution ITR. And then the high-resolution ITR was used as prior information to integrate the coarsened ITR. We compared

the integrative learning results (both weighted by estimated benefit and equally weighted) to the coarsened ITR by ABLO using data set 1 (ABLO-L) with low-resolution feature variables $\boldsymbol{Z}_1$. We were not able to compare the result with ABLO-H using data set 0, because we assumed complete feature variable information was not observed in the validation sets and only low-resolution ITRs were fitted in the simulation.

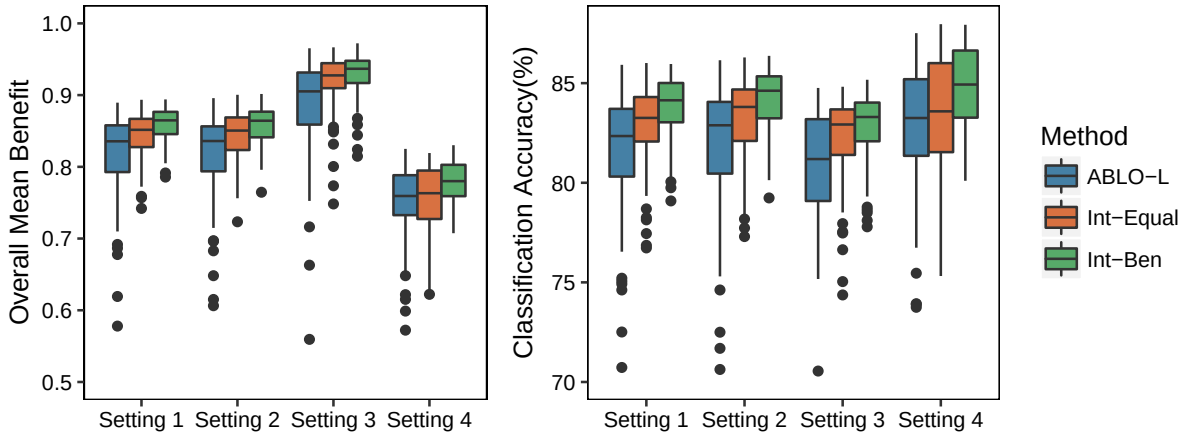
Table 4.2: Mean and standard deviation of overall benefit, value and accuracy rates using integrative learning for low-resolution ITRs comparing to ITRs by ABLO on single studies

		ABLO-L	Int-Equal	Int-Ben
Benefit	Setting 1	0.820(0.057)	0.844(0.033)	0.858(0.023)
	Setting 2	0.820(0.057)	0.843(0.036)	0.858(0.025)
	Setting 3	0.888(0.066)	0.919(0.039)	0.928(0.029)
	Setting 4	0.748(0.060)	0.757(0.046)	0.778(0.030)
Value	Setting 1	0.675(0.030)	0.688(0.018)	0.695(0.013)
	Setting 2	0.671(0.029)	0.683(0.018)	0.690(0.013)
	Setting 3	0.829(0.034)	0.845(0.020)	0.849(0.016)
	Setting 4	0.519(0.030)	0.524(0.022)	0.534(0.014)
Accuracy	Setting 1	0.817(0.030)	0.830(0.021)	0.838(0.016)
	Setting 2	0.820(0.032)	0.833(0.021)	0.842(0.015)
	Setting 3	0.805(0.035)	0.823(0.021)	0.828(0.017)
	Setting 4	0.827(0.036)	0.834(0.029)	0.848(0.021)

*: ABLO-L is estimated by ABLO using low-resolution data only;
Int-Equal is estimated by integrative learning for low-resolution ITR with equal weights for all subjects;
Int-Ben is estimated by integrative learning for low-resolution ITR weighted by estimated benefit.

Simulation results from 100 repetitions are summarized in Table 4.2 and Figure 4.3. For all the validation settings, ITRs learned by integrative learning (Int-Equal and Int-Ben) using data sets from both trials outperform the coarsened ITR learned by ABLO using single data set of low-resolution feature variables. Also, integrative learning weighted by estimated benefit performs better than integrative learning using equal weights, indicating the estimated benefit does help

Figure 4.3: Overall ITR benefit and optimal treatment allocation accuracy for the three methods.



*: ABLO-L is estimated by ABLO using low-resolution data only;
Int-Equal is estimated by integrative learning for low-resolution ITR with equal weights for all subjects;
Int-Ben is estimated by integrative learning for low-resolution ITR weighted by estimated benefit.

in learning the integrative ITR. By borrowing useful information, the integrative learning is also more efficient comparing to ABLO-L by reducing the standard deviations for about 50% in estimation of overall benefit and value function. By borrowing information from high-resolution ITR, the integrative ITR retains the simplicity by including only four feature variables, but improves the performance by incorporating additional information from another RCT with more feature variables.

Additional simulation studies were performed in Appendix Section C.1, where simulation models are slightly different across different trials.

4.5 Application to the EMBARC trial

EMBARC (Trivedi et al., 2016) is a two-phase, multi-site, randomized trial, which was designed to discover biosignatures associated with response to treatment for MDD. In the first stage, participants were randomized to an 8-week treatment of sertraline (one kind of SSRI, Selective Serotonin Reuptake Inhibitor) or placebo. We will focus on the first stage in our analysis. The primary outcome, Hamilton Depression Rating Scale (HAM-D) was measured at week 8 to assess the severity of depression. The clinical outcome (reward) we use in the analysis is the change in HAM-D score from baseline to week 8, where a larger value of the change corresponds to better treatment response (a greater reduction in symptom). There are 242 participants with complete baseline clinical feature variables and outcome. The clinical feature variables are gender, race, age, baseline HAM-D score, and baseline Quick Inventory of Depressive Symptomatology (QIDS) score. Among all 242 patients, only 138 have neuroimaging and behavioral phenotyping measures. In this example, 138 patients with complete clinical measures, behavioral phenotyping and neuroimaging measures were considered as study 0. We performed univariate analysis as a screening step by a linear regression model including the feature variable, assigned treatment and their interaction to assess all tier 1 behavioral phenotyping and brain imaging measures. The most informative behavioral phenotyping measure "Effect of Flanker interference on accuracy" and functional magnetic resonance imaging (fMRI) measure "Pregenual cingulate (seed) to right amygdala conflict adaptation coupling" were selected with p-value of the interaction term less than 0.1. Flanker accuracy effect is a measure of interference effects, with higher scores indicating an increased interference effects (i.e., reduced cognitive control). The selected fMRI measure is the difference of the connectivities between Pregenual cingulate and Right amygdala during incongruent minus congruent frames, dur-

ing the emotion recognition task (ERT), which is a standard neuropsychiatric test. Both variables were included when estimating a high-resolution ITR using EMBARC data.

Healing Emotion After Loss (HEAL, Shear et al., 2016) is the low-resolution study we used to improve the performance of ITR for EMBARC study. HEAL is a single-phase, multi-site, randomized trial. Patients with complicated grief were randomized to receive citalopram (also one kind of SSRI), placebo, citalopram+psychotherapy, or placebo+psychotherapy. Since a large proportion of patients with complicated grief are also suffering from MDD, we use the subgroup of patients with current MDD and were randomized to either citalopram or placebo in the analysis. Based on the study design, the primary outcome for HEAL was the change of QIDS score from baseline to week 12. There are 74 patients satisfied our inclusion criteria and used in the analysis. To learn a coarsened ITR, we used age, gender and baseline QIDS score in the HEAL study as the feature variables.

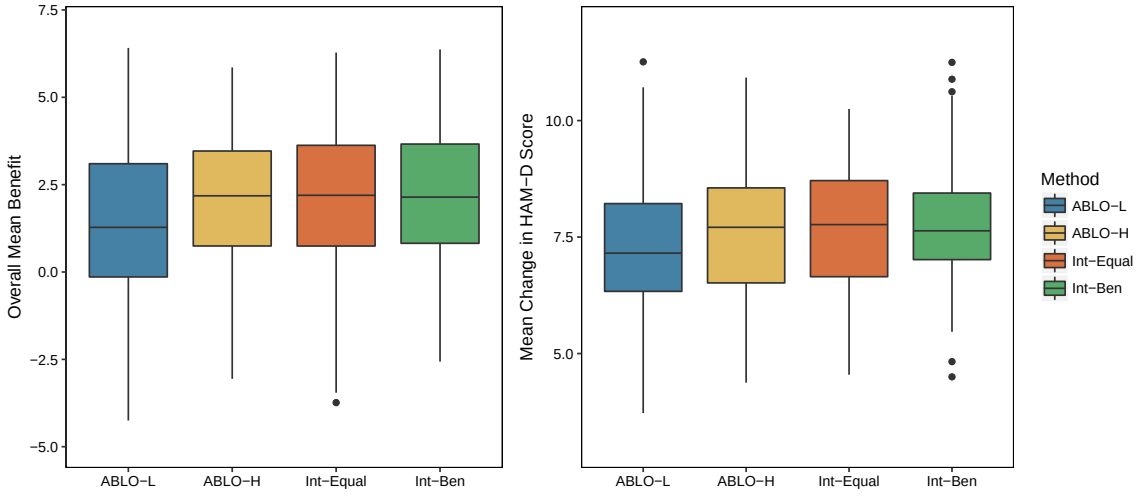
We applied the same four methods as the simulation studies to estimate the optimal high-resolution ITR (SSRI or placebo) for patients with complete feature variable information in the EMBARC study (132 subjects). For all methods, we randomly split the EMBARC sample into a training and testing set with a 2:1 sample size ratio and repeated the procedure 100 times. We bootstrapped the HEAL sample to learn a coarsened ITR for every repetition. The value function and ITR benefits were evaluated on the testing set. Results for ABLO using the HEAL data (ABLO-L), ABLO using the EMBARC complete feature variables (ABLO-H), integrative learning with equal weight (Int-Equal) and integrative learning with estimated benefit (Int-Ben) are compared in Table 4.3 and Figure 4.4. The non-personalized rules yield a change in HAM-D score of 7.48 for SSRI and 5.84 for placebo, with a difference of 1.64. The ITR estimated by Int-Ben yields a HAM-D change score of 7.71 (sd = 1.27), which is larger than Int-Equal (7.64, sd = 1.36),

ABLO-H (7.63, sd = 1.40), and ABLO-L (7.24, sd=1.60). The overall ITR benefit estimated by Int-Ben (2.24, sd = 1.98) is larger than Int-Equal (2.04, sd = 2.18), ABLO-H (2.04, sd = 1.98), and ABLO-L (1.33, sd = 2.49). The ITR benefit based on Int-Ben is also much larger than the non-personalized rule (2.24 versus 1.64). The estimation of value function is more efficient by using Int-Ben (sd = 1.27) comparing to ABLO-H (sd = 1.40), while the estimation of ITR benefit is equally efficient for Int-Ben (sd = 1.98) and ABLO-H (sd = 1.98) with a larger estimation of overall benefit.

Table 4.3: Overall performance of the four methods in EMBARC study using HEAL ITR

Method	ABLO-L	ABLO-H	Int-Equal	Int-Ben
Value	7.236(1.602)	7.634(1.399)	7.640(1.359)	7.712(1.266)
Benefit	1.331(2.485)	2.041(1.975)	2.042(2.179)	2.239(1.982)

Figure 4.4: Overall performance of the four methods in EMBARC study using HEAL ITR

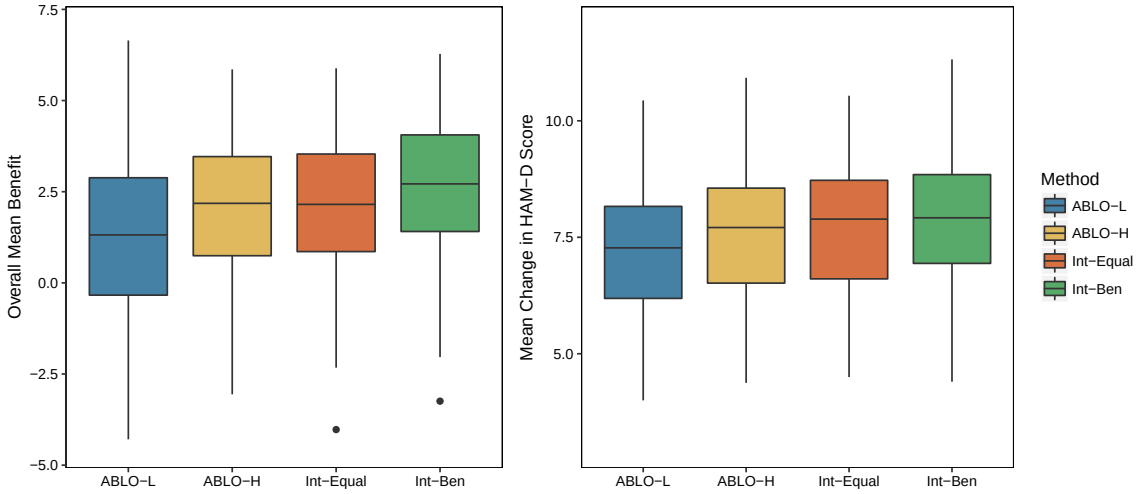


*:Distribution of estimated ITR benefit (the higher the better) and change in HAM-D score (the larger the better) at the end of stage 1 treatment for four methods (based on 100 cross-validation)

Table 4.4: Overall performance of the four methods to handle blockwise feature domain data in EMBARC study

Method	ABLO-L	ABLO-H	Int-Equal	Int-Ben
Value	7.217(1.438)	7.634(1.399)	7.672(1.351)	7.949(1.371)
Benefit	1.253(2.206)	2.041(1.975)	2.113(1.968)	2.682(2.024)

Figure 4.5: Overall performance of the four methods to handle blockwise feature domain data in EMBARC study



*:Distribution of estimated ITR benefit (the higher the better) and change in HAM-D score (the larger the better) at the end of stage 1 treatment for four methods (based on 100 cross-validation)

Next, we present an application to handle blockwise feature domain data when learning an ITR. In EMBARC, 104 subjects have complete clinical measures but do not have complete neuroimaging and behavioral phenotyping measures. In the following, we will use these 104 subjects as the study 1 with low-resolution covariates to improve the performance of the high-resolution ITR learned from the remaining 138 subjects with complete feature variables. The coarsened ITR learned

from study 1 includes all the clinical measures at baseline, i.e., age, gender, race, baseline QIDs score, and HAM-D score, while the high-resolution ITR includes the selected fMRI and behavioral phenotyping measure. We repeated the same procedure as using the HEAL study for 100 times and reported the results in Table 4.4 and Figure 4.5. The ITR benefit of Int-Ben (2.68, sd=2.02) is larger than that using HEAL ITR (2.24, sd=1.98). The coarsened ITR learned from EMBARC includes more feature variables and more subjects than HEAL, which lead to a greater improvement in performance of the high-resolution ITR.

4.6 Discussion

In this work, we propose integrative learning methods to estimate integrative ITRs with combined information across multiple trials. In practice, it is rare for RCTs to collect exactly the same feature variables, given different hypotheses and goals. Our method does not require all the trials to collect the same subsets of feature variables, which allows more flexibility comparing to integrative analysis that combines multiple data sets into one data set. Although here we focused on linear decision rules, our method can be generalized to any nonparametric form of decision rules or using other loss functions to replace zero-one loss for optimization (4.1) and (4.2). Our integrative methods improve the efficiency and reproducibility not only for high-resolution ITRs but also for coarsened ITRs by using a regularized value function to incorporate information from other related studies and a data-driven method to determine how much evidence each study contributes to the integrative ITR.

Here a linear regression is used to estimate the benefit of ITRs estimated from other studies. Other nonparametric or machine learning methods can be considered. In addition, future studies

can be developed to explore other methods to properly weigh the evidence of RCTs, such as using Composite Interaction Tree (CITree, Qiu and Wang, 2018). Another extension is to apply the integrative learning methods to observational studies, where one difficulty is how to properly weigh the evidence of RCTs versus observational studies when both types of studies are present. Due to the virtue of randomization in RCTs, an unbiased estimator of expected clinical outcome given a subset of features is also an unbiased estimator of the expected clinical outcome given comprehensive features, which may not be true for observational studies with unknown confounders. This causes another difficulty in applying integrative learning to observational studies. In this work, we derived the underlying Bayesian rules for the integrative learning method, which shows that our integrative rule maximizes a weighted combination of conditional treatment effects given differential subsets of feature variables. An interesting theoretical work would be to show that the integrative learning method is more efficient compared to the ITRs learned from single trials.

There are several limitations for the proposed method. First, integrative learning may not improve performance when subjects from multiple trials follow substantially different distributions, especially when f_0 and f_k yield different treatment recommendations for a large number of subjects. However, our tuning parameters are chosen adaptively based on data to prevent integrative learning from deteriorating performance in this setting. For example, when the target population is the first trial, which has substantially different distribution from the second trial, we expect that the tuning parameter λ will be chosen close to 0, and therefore the second trial does not assist in improving the performance of the ITR. When the target population is different from all trials with observed data, the performance gain depends on the nature of differences and we performed a few sensitivity analyses to examine performance (see for example, Table C.1 and C.2). Another limitation of the integrative learning is that it requires the same two treatments are assessed across multiple trials.

It is interesting to extend the current integrative learning method to handle multiple treatments using evidence from direct and indirect comparisons, which is similar to network meta-analysis (Tonin et al., 2017).

Bibliography

An, L. T. H., Tao, P. D., and Muu, L. D. (1996). Numerical solution for optimization over the efficient set by D.C. optimization algorithms. *Operations Research Letters*, 19(3):117–128.

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

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 57(1):289–300.

Blatt, D., Murphy, S., and Zhu, J. (2004). *A-learning for approximate planning*. Technical Report 04-63, The Methodology Center, Pennsylvania State University, State College.

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., Knowles, S., Cuijpers, P., Andersson, G., Christensen, H., et al. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ*, 346:f540.

Carini, C., Menon, S. M., and Chang, M. (2014). *Clinical and Statistical Considerations in Personalized Medicine*. CRC Press.

Chakraborty, B. and Moodie, E. (2013). *Statistical methods for dynamic treatment regimes*. Springer.

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., Leucht, S., Ruhe, H. G., Turner, E. H., Higgins, J. P., et al. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*, 391(10128):1357–1366.

Collobert, R., Sinz, F., Weston, J., and Bottou, L. (2006). Trading convexity for scalability. In *Proceedings of the 23rd international conference on Machine learning*, pages 201–208. ACM.

Dusseldorp, E., Doove, L., and Van Mechelen, I. (2016). Quint: An r package for the identification of subgroups of clients who differ in which treatment alternative is best for them. *Behavior research methods*, 48(2):650–663.

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.

Ertefaie, A., Shortreed, S., and Chakraborty, B. (2016). Q-learning residual analysis: application to the effectiveness of sequences of antipsychotic medications for patients with schizophrenia. *Statistics in medicine*, 35(13):2221–2234.

Fan, C., Lu, W., Song, R., and Zhou, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(5):1565–1582.

Foster, J. C., Taylor, J. M., and Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in medicine*, 30(24):2867–2880.

Fu, H., Zhou, J., and Faries, D. E. (2016). Estimating optimal treatment regimes via subgroup identification in randomized control trials and observational studies. *Statistics in medicine*, 35(19):3285–3302.

Gunter, L., Zhu, J., and Murphy, S. (2011). Variable selection for qualitative interactions. *Statistical Methodology*, 8(1):42–55.

Haidich, A.-B. (2010). Meta-analysis in medical research. *Hippokratia*, 14(Suppl 1):29–37.

Huang, Y. and Fong, Y. (2014). Identifying optimal biomarker combinations for treatment selection via a robust kernel method. *Biometrics*, 70(4):891–901.

Huynh, N. N. and McIntyre, R. S. (2008). What are the implications of the STAR* D trial for primary care? a review and synthesis. *Primary care companion to the Journal of clinical psychiatry*, 10(2):91–96.

Jakobsen, J. C., Katakam, K. K., Schou, A., Hellmuth, S. G., Stallknecht, S. E., Leth-Møller, K., Iversen, M., Banke, M. B., Petersen, I. J., Klingenberg, S. L., et al. (2017). Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. a systematic review with meta-analysis and trial sequential analysis. *BMC psychiatry*, 17:58.

Jones, A. P., Happé, F. G., Gilbert, F., Burnett, S., and Viding, E. (2010). Feeling, caring, knowing: different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 51(11):1188–1197.

Justice, A. C., Covinsky, K. E., and Berlin, J. A. (1999). Assessing the generalizability of prognostic information. *Annals of internal medicine*, 130(6):515–524.

Keller, M. B., McCullough, J. P., Klein, D. N., Arnow, B., Dunner, D. L., Gelenberg, A. J., Markowitz, J. C., Nemeroff, C. B., Russell, J. M., Thase, M. E., et al. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *New England Journal of Medicine*, 342(20):1462–1470.

Kocsis, J. H., Gelenberg, A. J., Rothbaum, B. O., Klein, D. N., Trivedi, M. H., Manber, R., Keller, M. B., Leon, A. C., Wisniewski, S. R., Arnow, B. A., et al. (2009). Cognitive behavioral analysis system of psychotherapy and brief supportive psychotherapy for augmentation of antidepressant nonresponse in chronic depression: the REVAMP trial. *Archives of General Psychiatry*, 66(11):1178–1188.

Koo, J.-Y., Lee, Y., Kim, Y., and Park, C. (2008). A bahadur representation of the linear support vector machine. *The Journal of Machine Learning Research*, 9:1343–1368.

Laber, E. and Zhao, Y. (2015). Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514.

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

Laber, E. B. and Murphy, S. A. (2011). Adaptive confidence intervals for the test error in classification. *Journal of the American Statistical Association*, 106(495):904–913.

Lavori, P. W. and Dawson, R. (2004). Dynamic treatment regimes: practical design considerations. *Clinical trials*, 1(1):9–20.

Lipkovich, I., Dmitrienko, A., Denne, J., and Enas, G. (2011). Subgroup identification based on

differential effect searcha recursive partitioning method for establishing response to treatment in patient subpopulations. *Statistics in medicine*, 30(21):2601–2621.

Liu, C., Xu, X., and Hu, D. (2015). Multiobjective reinforcement learning: A comprehensive overview. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(3):385–398.

Liu, Y., Wang, Y., Kosorok, M. R., Zhao, Y., and Zeng, D. (2014). Robust hybrid learning for estimating personalized dynamic treatment regimens. *arXiv preprint arXiv:1611.02314*.

Ma, S., Huang, J., and Song, X. (2011). Integrative analysis and variable selection with multiple high-dimensional data sets. *Biostatistics*, 12(4):763–775.

McAllester, D. A. and Keshet, J. (2011). Generalization bounds and consistency for latent structural probit and ramp loss. In *Neural Information Processing Systems*, pages 2205–2212.

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.

Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234.

Pelham, W. E. and Fabiano, G. A. (2008). Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, 37(1):184–214.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2):1180–1210.

Qiu, X. and Wang, Y. (2018). Composite interaction tree for simultaneous learning optimal individualized treatment rules and subgroups. *Submitted*.

Qiu, X., Zeng, D., and Wang, Y. (2017). Estimation and evaluation of linear individualized treatment rules to guarantee performance. *Biometrics*, In Press.

Rich, B., Moodie, E. E., Stephens, D. A., and Platt, R. W. (2010). Model checking with residuals for g-estimation of optimal dynamic treatment regimes. *The international journal of biostatistics*, 6(2):Article 12.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., Thase, M. E., Nierenberg, A. A., Quitkin, F. M., Kashner, T. M., et al. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Controlled Clinical Trials*, 25(1):119–142.

Rush, A. J., Trivedi, M. H., Stewart, J. W., Nierenberg, A. A., Fava, M., Kurian, B. T., Warden, D., Morris, D. W., Luther, J. F., Husain, M. M., et al. (2011). Combining medications to enhance depression outcomes (CO-MED): acute and long-term outcomes of a single-blind randomized study. *American Journal of Psychiatry*, 168(7):689–701.

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):640–661.

Shankman, S. A., Campbell, M. L., Klein, D. N., Leon, A. C., Arnow, B. A., Manber, R., Keller,

M. B., Markowitz, J. C., Rothbaum, B. O., Thase, M. E., et al. (2013). Dysfunctional attitudes as a moderator of pharmacotherapy and psychotherapy for chronic depression. *Journal of psychiatric research*, 47(1):113–121.

Shear, M. K., Reynolds, C. F., Simon, N. M., Zisook, S., Wang, Y., Mauro, C., Duan, N., Lebowitz, B., and Skritskaya, N. (2016). Optimizing treatment of complicated grief: a randomized clinical trial. *JAMA psychiatry*, 73(7):685–694.

Su, X., Tsai, C.-L., Wang, H., Nickerson, D. M., and Li, B. (2009). Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, 10:141–158.

Sysko, R., Sha, N., Wang, Y., Duan, N., and Walsh, B. T. (2010). Early response to antidepressant treatment in bulimia nervosa. *Psychological medicine*, 40(6):999–1005.

Tonin, F. S., Rotta, I., Mendes, A. M., and Pontarolo, R. (2017). Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. *Pharmacy Practice (Granada)*, 15(1):943.

Trivedi, M. H. and Daly, E. J. (2008). Treatment strategies to improve and sustain remission in major depressive disorder. *Dialogues in clinical neuroscience*, 10(4):377–384.

Trivedi, M. H., McGrath, P. J., Fava, M., Parsey, R. V., Kurian, B. T., Phillips, M. L., Oquendo, M. A., Bruder, G., Pizzagalli, D., Toups, M., et al. (2016). Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *Journal of Psychiatric Research*, 78:11–23.

Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., Norquist, G., Howland, R. H., Lebowitz, B., McGrath, P. J., et al. (2006). Evaluation of outcomes with

citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *American Journal of Psychiatry*, 163(1):28–40.

Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge university press.

Wallace, M. P., Moodie, E. E., and Stephens, D. A. (2016). Model assessment in dynamic treatment regimen estimation via double robustness. *Biometrics*, 72(3):855–864.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. PhD thesis, University of Cambridge England.

Xiang, S., Yuan, L., Fan, W., Wang, Y., Thompson, P. M., and Ye, J. (2013). Multi-source learning with block-wise missing data for alzheimer's disease prediction. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 185–193. ACM.

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.

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.-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.

Zhu, R., Zeng, D., and Kosorok, M. R. (2015). Reinforcement learning trees. *Journal of the American Statistical Association*, 110(512):1770–1784.

Zhu, R., Zhao, Y.-Q., Chen, G., Ma, S., and Zhao, H. (2017). Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics*, 73(2):391–400.

Appendix A

Appendices for Chapter 2

A.1 Computing the Theoretical Optimal Linear Rule

Here we derive the theoretical optimal linear rule f_L^* in the class of all linear rules $f \in \mathcal{L}$ under our simulation settings in Section 2.4. Let G be the latent class identifier in the simulations. Define $G|(\boldsymbol{X}, W, V, A, \boldsymbol{U}) = G|\boldsymbol{X}$ as the class number, which only depends on $\boldsymbol{X} = (X^1, X^2, \cdots, X^p)$, where $X^j|G = k \sim N(\mu_k, 1)$ for $j = 1, \cdots, p$, and $k = 1, 2, 3, 4$. For a given treatment decision rule f, the expected value function under the decision rule is

$$
\begin{aligned}
& E\left[\frac{R}{\pi(A|\boldsymbol{X})}\left\{I(Af(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0)\right\}\right] \\
= \;& E\left[I(f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0)\left\{E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A = 1) - E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A = -1)\right\}\right] \\
+ \;& E\big\{E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A = -1)\big\}.
\end{aligned}
$$

Because $E\left\{E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A = -1)\right\}$ is a constant which doesn't depend on f, maximizing the expected value function is equivalent to maximizing $E\left\{I(f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0)\Omega(\boldsymbol{X}, W)\right\}$, where

under the simulation model for $E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A)$ we can obtain

$$\begin{aligned}\Omega(\boldsymbol{X}, W) &= P(G=1|\boldsymbol{X})\{\delta_1 + (\alpha_{11}-\alpha_{21})W\} + P(G=2|\boldsymbol{X})\{\delta_2 + (\alpha_{12}-\alpha_{22})W\} \\ &+ P(G=3|\boldsymbol{X})\{-\delta_1 + (\alpha_{13}-\alpha_{23})W\} + P(G=4|\boldsymbol{X})\{-\delta_2 + (\alpha_{14}-\alpha_{24})W\}.\end{aligned}$$

Next, we show that V and $\boldsymbol{U}$ are independent of optimal linear decision rule f_L^*. Let $f_L^*(\boldsymbol{X}, W)$ maximizes the value function in class $\mathcal{L}$. For any fixed V and $\boldsymbol{U}$,

$$\begin{aligned}E\{I[f_L^*(\boldsymbol{X}, W) > 0]\Omega(\boldsymbol{X}, W)\} &\geq E[I\{f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0\}\Omega(\boldsymbol{X}, W)] \\ &= E[I\{f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0\}\Omega(\boldsymbol{X}, W)|V, \boldsymbol{U}].\end{aligned}$$

Therefore, take expectation of the inequality to obtain

$$\begin{aligned}E[E\{I(f_L^*(\boldsymbol{X}, W) > 0)\Omega(\boldsymbol{X}, W)\}] &\geq E[E\{I(f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0)\Omega(\boldsymbol{X}, W)|V, \boldsymbol{U}\}] \\ &= E[I\{f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0\}\Omega(\boldsymbol{X}, W)].\end{aligned}$$

Thus we can ignore the independent noise variables while maximizing the value function.

Under linear transformation,

$$\boldsymbol{X} \to (\frac{X_s}{\sqrt{p}}, \widetilde{x_2}, \cdots, \widetilde{x_p}),$$

where $X_s = X^1 + X^2 + \cdots + X^p$, and $\widetilde{x_2}, \cdots, \widetilde{x_p}$ are orthogonal to X_s, the objective function becomes

$$\int\int I\{f(X_s, \widetilde{x_2}, \cdots, \widetilde{x_p}, W) > 0\}\,\Omega(X_s, W) e^{-\frac{X_s^2}{2p} - \frac{\tilde{x}_2^2}{2} \cdots - \frac{\tilde{x}_s^2}{2}} dX_s f(W) dW d\widetilde{x}_2 \cdots d\widetilde{x}_s,$$

where

$$\begin{aligned}\Omega(X_s, W) = e^{\mu_1 X_s - \frac{p\mu_1^2}{2}}\{\delta_1 + (\alpha_{11}-\alpha_{21})W\} + e^{\mu_2 X_s - \frac{p\mu_2^2}{2}}\{\delta_2 + (\alpha_{12}-\alpha_{22})W\} \\ + e^{\mu_3 X_s - \frac{p\mu_3^2}{2}}\{-\delta_1 + (\alpha_{13}-\alpha_{23})W\} + e^{\mu_4 X_s - \frac{p\mu_4^2}{2}}\{-\delta_2 + (\alpha_{14}-\alpha_{24})W\}.\end{aligned}$$

Because $(\widetilde{x_2}, \cdots, \widetilde{x_p})$ are independent noise variables, as shown before, the optimal linear rule only depends on X_s and W. The objective function is thus equivalent to

$$\int\int I\left\{f(X_s, W) > 0\right\} \Omega(X_s, W) dX_s f(W) dW.$$

As $X_s \sim \frac{1}{4}N(\mu_1, p) + \frac{1}{4}N(\mu_2, p) + \frac{1}{4}N(\mu_3, p) + \frac{1}{4}N(\mu_4, p)$ and $W \sim N(0, 1)$, where μ_k is the mean of X^p in the kth class. Monte Carlo method can be applied to find the optimal linear rule f_L^*.

A.2 Additional Simulation Results

We performed additional simulations to vary the strength of the informative feature variable W, such that its effects in different settings are $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 1 & 0.3 & 0.6 \\ 0.5 & 0.5 & 0.3 & 0.6 \end{bmatrix}$.

Results from 500 replicates are summarizied in Table A.1, Figure A.1, and A.2. ABLO with linear kernel has the highest optimal treatment classification accuracy regardless of the sample size for both settings, and also estimates the ITR benefit closest to the true global maximal value of 0.705 on the overall sample. PM, Q-learning, and O-learning underestimate the ITR benefit, especially when the sample size is smaller ($N = 400$ training, 400 testing). Thus they do not achieve the maximal value of the theoretical optimal linear rule. The performance of estimating subgroup ITR benefit is similar to the overall sample. ABLO outperforms other methods with subgroup ITR benefit closer to the true global maximal value (e.g., in groups $W \in [-0.5, 0.5]$ and $W > 0.5$).

A.3 Sensitivity to the Starting Values of ABLO

To evaluate the sensitivity of the algorithm to starting values, we include the algorithm convergence path of two example datasets in terms of value function and weighted ramp loss function. In Figure

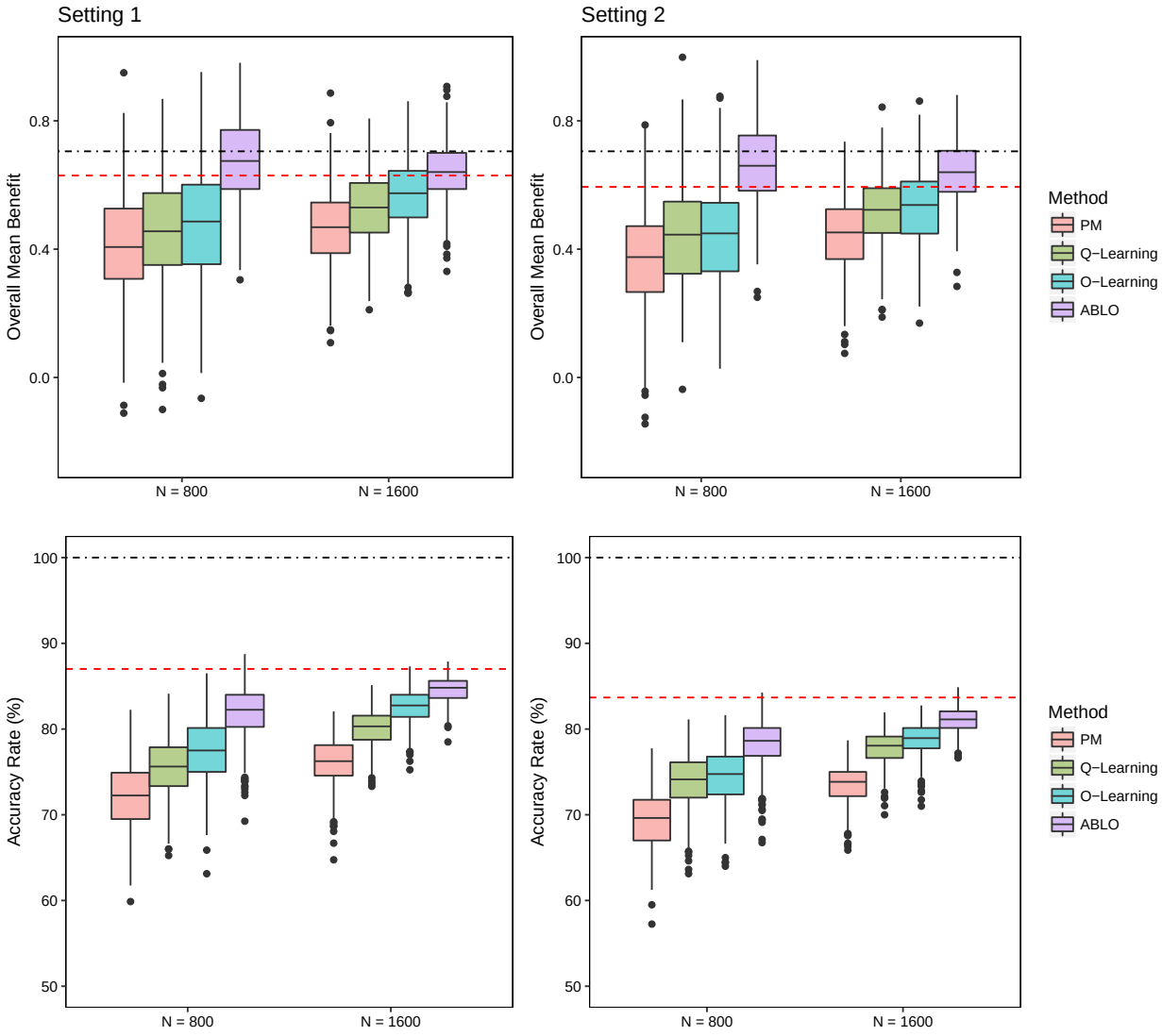
A.3, lines indicate convergence paths given different initial values. In the first example dataset, the algorithm converges to the same value function and ramp loss. However, the algorithm converges fastest if starting with O-leaning estimates. In the second dataset, the algorithm is more sensitive to different starting values, but the one starting with O-learning estimates performs the best, which is also the proposed starting values for ABLO.

Table A.1: Simulation results: mean and standard deviation of the accuracy rate, mean benefit, and coverage probability for estimation of the benefit of the optimal ITR

Setting 1. Four region means $= (1, 0.5, -1, -0.5)$.									
		Overall Benefit		$W < -0.5$		$W \in [-0.5, 0.5]$		$W > 0.5$	
	Accuracy rate	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage
$N = 800$									
PM	0.72 (0.04)	0.41 (0.17)	0.77	0.16 (0.24)	0.97	0.40 (0.23)	0.87	0.69 (0.30)	0.76
Q-learning	0.75 (0.03)	0.46 (0.17)	0.83	0.21 (0.24)	0.97	0.46 (0.23)	0.89	0.72 (0.27)	0.77
O-learning	0.77 (0.04)	0.48 (0.17)	0.85	0.20 (0.24)	0.97	0.48 (0.23)	0.92	0.76 (0.28)	0.82
ABLO	0.82 (0.03)	0.68 (0.13)	0.94	0.36 (0.22)	0.91	0.66 (0.21)	0.95	0.99 (0.24)	0.93
$N = 1600$									
PM	0.76 (0.03)	0.47 (0.12)	0.72	0.15 (0.17)	0.96	0.46 (0.16)	0.85	0.80 (0.21)	0.70
Q-learning	0.80 (0.02)	0.53 (0.11)	0.86	0.21 (0.16)	0.97	0.53 (0.16)	0.92	0.85 (0.18)	0.80
O-learning	0.83 (0.02)	0.57 (0.10)	0.94	0.23 (0.16)	0.97	0.57 (0.15)	0.95	0.93 (0.19)	0.90
ABLO	0.85 (0.01)	0.64 (0.09)	0.96	0.27 (0.15)	0.93	0.64 (0.14)	0.95	1.02 (0.16)	0.94
Best linear rule*	0.870	$\delta^l_{\mathcal{C}} = 0.630$		$\delta^l_{\mathcal{C}} = 0.200$		$\delta^l_{\mathcal{C}} = 0.624$		$\delta^l_{\mathcal{C}} = 1.063$	
Setting 2. Four region means $= (1, 0.3, -1, -0.3)$.									
		Overall Benefit		$W < -0.5$		$W \in [-0.5, 0.5]$		$W > 0.5$	
	Accuracy rate	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage
$N = 800$									
PM	0.69 (0.03)	0.37 (0.17)	0.75	0.15 (0.24)	0.97	0.36 (0.23)	0.87	0.60 (0.29)	0.75
Q-learning	0.74 (0.03)	0.44 (0.16)	0.85	0.20 (0.22)	0.98	0.45 (0.23)	0.91	0.68 (0.26)	0.82
O-learning	0.74 (0.03)	0.45 (0.16)	0.86	0.21 (0.22)	0.98	0.44 (0.23)	0.91	0.69 (0.27)	0.84
ABLO	0.78 (0.03)	0.67 (0.13)	0.92	0.39 (0.20)	0.92	0.65 (0.21)	0.95	0.95 (0.24)	0.94
$N = 1600$									
PM	0.73 (0.02)	0.44 (0.11)	0.76	0.17 (0.18)	0.96	0.44 (0.15)	0.89	0.72 (0.19)	0.76
Q-learning	0.78 (0.02)	0.51 (0.11)	0.90	0.23 (0.16)	0.98	0.52 (0.16)	0.94	0.79 (0.19)	0.83
O-learning	0.79 (0.02)	0.53 (0.11)	0.91	0.24 (0.17)	0.96	0.53 (0.16)	0.94	0.82 (0.18)	0.88
ABLO	0.81 (0.01)	0.64 (0.10)	0.92	0.33 (0.16)	0.92	0.63 (0.15)	0.93	0.95 (0.16)	0.96
Best linear rule	0.837	$\delta^l_{\mathcal{C}} = 0.594$		$\delta^l_{\mathcal{C}} = 0.222$		$\delta^l_{\mathcal{C}} = 0.585$		$\delta^l_{\mathcal{C}} = 0.974$	
Best global rule		$\delta_{\mathcal{C}} = 0.705$		$\delta_{\mathcal{C}_1} = 0.373$		$\delta_{\mathcal{C}_2} = 0.647$		$\delta_{\mathcal{C}_3} = 1.109$	

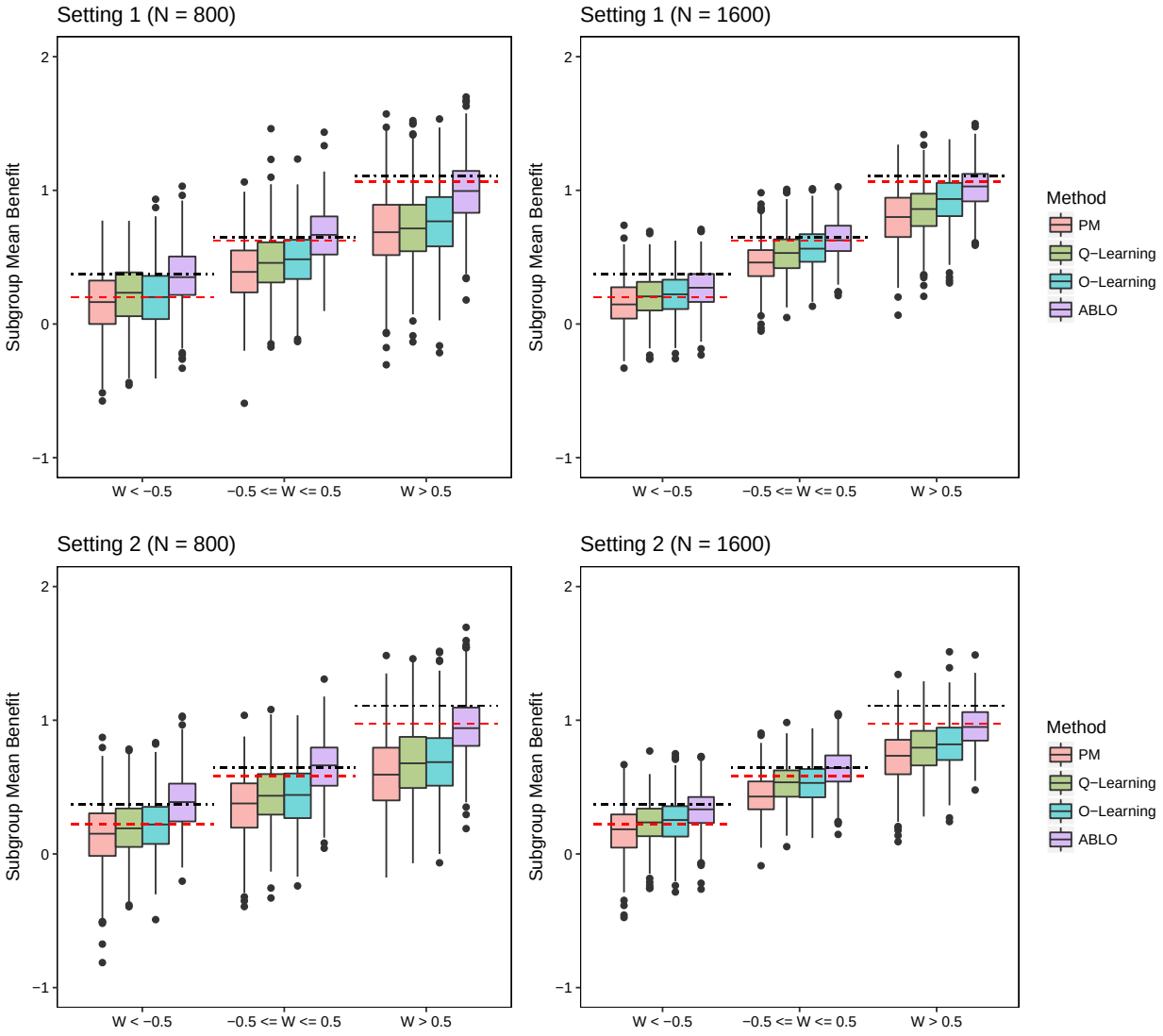
*: The theoretical best linear rule for setting 1 is $\text{sign}(X_s * 0.98 + W * 0.19 - 0.03)$, where $X_s = X^1 + X^2 + \cdots + X^{10}$.
The theoretical best linear rule for setting 2 is sign($X_s * 0.90 + W * 0.39 - 0.19$). PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning.

Figure A.1: Simulation results: Overall ITR benefit and accuracy rates for the four methods



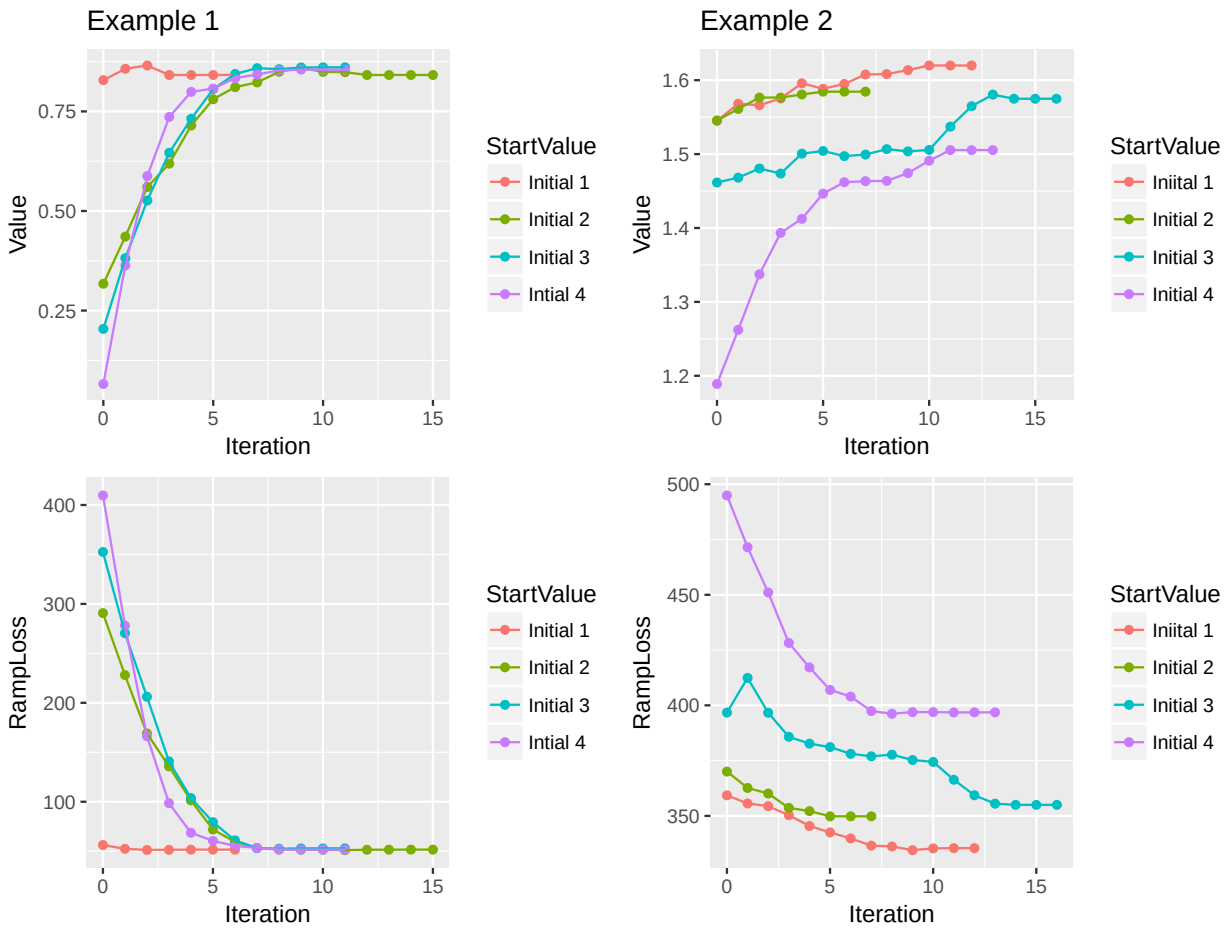
*: Dotted-dashed lines represent the benefit (top panels) and accuracy (bottom panels under the theoretical global optimal treatment f^*. Dashed lines represent the benefit and accuracy under the theoretical optimal linear rule f_L^*. The methods being compared are (from left to right): PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning. This figure appears in color in the electronic version of this article.

Figure A.2: Simulation results: Subgroup ITR benefit for the four methods



*:Black dotted-dashed lines represent the benefit under the theoretical global optimal treatment f^*. Red dashed lines represent the benefit under the theoretical optimal linear rule f_L^*. The methods being compared are (from left to right): PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning. This figure appears in color in the electronic version of this article.

Figure A.3: Performance of the algorithm on two example datasets evaluated by value function and penalized weighted sum of ramp loss



*: Initial 1 starts $\boldsymbol{\beta}$ from the estimates obtained by O-learning; Initial 2 starts from $\boldsymbol{\beta} = \mathbf{0}$; Initial 3 starts from $\boldsymbol{\beta} = (1, \cdots, 1, -1, \cdots, -1)^T$, where half of the components are 1 and the other half are -1; Initial 4 starts from $\boldsymbol{\beta} = (1, 0, \cdots, 0)^T$. This figure appears in color in the electronic version of this article.

Appendix B

Appendices for Chapter 3

B.1 Estimation of Variance for $\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) - \widehat{\delta}_{\mathcal{C}_2}(\widehat{f})$

In a binary tree, consider a binary partition (e.g., based on $X_1 > c$ or $X_1 \leq c$) to partition n subjects with feature variables $\boldsymbol{X}_i$ into two subgroups $\mathcal{C}_1$ and $\mathcal{C}_2$. Recall that in an honest CITree, $\widehat{f}$ denotes the ITR estimated from an independent training sample and the estimated subgroup benefit for $\mathcal{C}_1$ is

$$\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) = \frac{\sum_{i=1}^n I\left\{\boldsymbol{X}_i \in \mathcal{C}_1, A_i = \text{sign}(\widehat{f}(\boldsymbol{X}_i))\right\} R_i/\pi_i}{\sum_{i=1}^n I(\boldsymbol{X}_i \in \mathcal{C}_1)} - \frac{\sum_{i=1}^n I\left\{\boldsymbol{X}_i \in \mathcal{C}_1, A_i \neq \text{sign}(\widehat{f}(\boldsymbol{X}_i))\right\} R_i/\pi_i}{\sum_{i=1}^n I((\boldsymbol{X}_i \in \mathcal{C}_1)},$$

where $\pi_i = \pi(A_i|\boldsymbol{X}_i)$.

To apply HTB test for $\mathcal{C}_1$ and $\mathcal{C}_2$, it is necessary to estimate the variance of $\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) - \widehat{\delta}_{\mathcal{C}_2}(\widehat{f})$, which is equivalent to

$$\text{Var}\left\{\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) - \widehat{\delta}_{\mathcal{C}_2}(\widehat{f})\right\} = \text{Var}\left\{\widehat{\delta}_{\mathcal{C}_1}(\widehat{f})\right\} + \text{Var}\left\{\widehat{\delta}_{\mathcal{C}_2}(\widehat{f})\right\} - 2\text{Cov}\left\{\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}),\ \widehat{\delta}_{\mathcal{C}_2}(\widehat{f})\right\}.$$

Denote

$$\begin{aligned}\mathbb{F}_n &= \frac{1}{n}\sum_{i=1}^{n}\left[I\left\{\boldsymbol{X}_i \in \mathcal{C}_1, A_i = \text{sign}(\widehat{f}(\boldsymbol{X}_i))\right\} - I\left\{\boldsymbol{X}_i \in \mathcal{C}_1, A_i \neq \text{sign}(\widehat{f}(\boldsymbol{X}_i))\right\}\right] R_i/\pi_i \\ &= \frac{1}{n}\sum_{i=1}^{n}\tilde{f}(X_i),\end{aligned}$$

and

$$\mathbb{G}_n = \frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{X}_i \in \mathcal{C}_1) = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i).$$

Define $\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) = \phi(\mathbb{F}_n, \mathbb{G}_n) = \frac{\mathbb{F}_n}{\mathbb{G}_n}$, and $\mathbb{P}_n = (\mathbb{F}_n, \mathbb{G}_n)$. By functional delta method of Van der Vaart (2000), we have

$$\sqrt{n}\left\{\phi(\mathbb{P}_n) - \phi(P)\right\} \approx \phi'_P\left\{\sqrt{n}(\mathbb{P}_n - P)\right\}.$$

Notice

$$\sqrt{n}\left\{\phi(\mathbb{P}_n) - \phi(P)\right\} = \sqrt{n}\left\{\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) - \delta_{\mathcal{C}_1}(f)\right\} \xrightarrow{d} N(0, V),$$

and

$$\phi(P) = \frac{E\mathbb{F}_n}{E\mathbb{G}_n} = \frac{E\tilde{f}(\boldsymbol{X})}{Eg(\boldsymbol{X})},$$

we obtain

$$\begin{aligned}V &= \text{Var}\left[\sqrt{n}\left\{\phi(\mathbb{P}_n) - \phi(P)\right\}\right] \\ &\approx \text{Var}\left[\frac{1}{E\mathbb{G}_n}\left\{\sqrt{n}(\mathbb{F}_n - E\mathbb{F}_n)\right\} - \frac{E\mathbb{F}_n}{(E\mathbb{G}_n)^2}\left\{\sqrt{n}(\mathbb{G}_n - E\mathbb{G}_n)\right\}\right] \\ &= \frac{1}{(E\mathbb{G}_n)^2}\text{Var}\left\{\sqrt{n}(\mathbb{F}_n - E\mathbb{F}_n)\right\} + \frac{(E\mathbb{F}_n)^2}{(E\mathbb{G}_n)^4}\text{Var}\left\{\sqrt{n}(\mathbb{G}_n - E\mathbb{G}_n)\right\} \\ &\quad -2\frac{E\mathbb{F}_n}{(E\mathbb{G}_n)^3}\text{Cov}\left\{\sqrt{n}(\mathbb{G}_n - E\mathbb{G}n),\ \sqrt{n}(\mathbb{F}_n - E\mathbb{F}_n)\right\} \\ &\to \frac{1}{E\{g(\boldsymbol{X})\}^2}\text{Var}\left\{\tilde{f}(\boldsymbol{X})\right\} + \frac{\{E\tilde{f}(\boldsymbol{X})\}^2}{\{Eg(\boldsymbol{X})\}^4}\text{Var}\left\{g(\boldsymbol{X})\right\} - 2\frac{E\tilde{f}(\boldsymbol{X})}{\{Eg(\boldsymbol{X})\}^3}\text{Cov}\left\{\tilde{f}(\boldsymbol{X}),\ g(\boldsymbol{X})\right\},\end{aligned}$$

where $Eg(\boldsymbol{X})$, $E\tilde{f}(\boldsymbol{X})$, $\text{Var}(\tilde{f}(\boldsymbol{X}))$, $\text{Var}(g(\boldsymbol{X}))$, and $\text{Cov}\left\{\tilde{f}(\boldsymbol{X}),\ g(\boldsymbol{X})\right\}$can be estimated by empirical mean, variance and covariance.

Similarly, denote

$$\begin{aligned}\mathbb{F}_n^* &= \sum_{i=1}^{n}\left[I\left\{\boldsymbol{X}_i \in \mathcal{C}_2, A_i = \text{sign}(\widehat{f}(\boldsymbol{X}_i))\right\} - I\left\{\boldsymbol{X}_i \in \mathcal{C}_2, A_i \neq \text{sign}(\widehat{f}(\boldsymbol{X}_i))\right\}\right] R_i/\pi_i \\ &= \sum_{i=1}^{n}\tilde{f}^*(\boldsymbol{X}_i),\end{aligned}$$

and

$$\mathbb{G}_n^* = \sum_{i=1}^{n} g^*(\boldsymbol{X}) = \sum_{i=1}^{n} I(\boldsymbol{X}_i \in \mathcal{C}_2).$$

Then, we obtain

$$\sqrt{n}\left\{\widehat{\delta}_{\mathcal{C}_2}(\widehat{f}) - \delta_{\mathcal{C}_2}(f)\right\} \approx \frac{1}{E\mathbb{G}_n^*}\left\{\sqrt{n}(\mathbb{F}_n^* - E\mathbb{F}_n^*)\right\} - \frac{E\mathbb{F}_n^*}{(E\mathbb{G}_n^*)^2}\left\{\sqrt{n}(\mathbb{G}_n^* - E\mathbb{G}_n^*)\right\}.$$

Therefore, we have

$$\begin{aligned}V^* &= n\text{Var}(\widehat{\delta}_{\mathcal{C}_2}(\widehat{f})) \\ &\to \frac{1}{E\{g^*(\boldsymbol{X})\}^2}\text{Var}\left\{\tilde{f}^*(\boldsymbol{X})\right\} + \frac{\{E\tilde{f}^*(\boldsymbol{X})\}^2}{\{Eg^*(\boldsymbol{X})\}^4}\text{Var}\left\{g^*(\boldsymbol{X})\right\} \\ &\quad -2\frac{E\tilde{f}^*(\boldsymbol{X})}{\{Eg^*(\boldsymbol{X})\}^3}\text{Cov}\left\{\tilde{f}^*(\boldsymbol{X}),\ g^*(\boldsymbol{X})\right\}.\end{aligned}$$

Next, we obtain the covariance of the two estimated subgroup benefits as

$$\begin{aligned}n\text{Cov}\left(\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}),\ \widehat{\delta}_{\mathcal{C}_2}(\widehat{f})\right) &\approx \frac{1}{E\mathbb{G}_n E\mathbb{G}_n^*}\text{Cov}\left\{\sqrt{n}(\mathbb{F}_n - E\mathbb{F}_n),\ \sqrt{n}(\mathbb{F}_n^* - E\mathbb{F}_n^*)\right\} \\ &- \frac{E\mathbb{F}_n^*}{E\mathbb{G}_n(E\mathbb{G}_n^*)^2}\text{Cov}\left\{\sqrt{n}(\mathbb{F}_n - E\mathbb{F}_n),\ \sqrt{n}(\mathbb{G}_n^* - E\mathbb{G}_n^*)\right\} \\ &- \frac{E\mathbb{F}_n}{(E\mathbb{G}_n)^2 E\mathbb{G}_n^*}\text{Cov}\left\{\sqrt{n}(\mathbb{G}_n - E\mathbb{G}_n),\ \sqrt{n}(\mathbb{F}_n^* - E\mathbb{F}_n^*)\right\} \\ &+ \frac{E\mathbb{F}_n E\mathbb{F}_n^*}{(E\mathbb{G}_n)^2(E\mathbb{G}_n^*)^2}\text{Cov}\left\{\sqrt{n}(\mathbb{G}_n - E\mathbb{G}_n),\ \sqrt{n}(\mathbb{G}_n^* - E\mathbb{G}_n^*)\right\} \\ &\to \frac{1}{Eg(\boldsymbol{X})Eg^*(\boldsymbol{X})}\text{Cov}\left\{\tilde{f}(\boldsymbol{X}),\ \tilde{f}^*(\boldsymbol{X})\right\} \\ &- \frac{E\tilde{f}^*(\boldsymbol{X})}{Eg(\boldsymbol{X})(E(g^*(\boldsymbol{X}))^2}\text{Cov}\left\{\tilde{f}(\boldsymbol{X}),\ g^*(\boldsymbol{X})\right\} \\ &- \frac{E\tilde{f}(\boldsymbol{X})}{(Eg(\boldsymbol{X}))^2 Eg^*(\boldsymbol{X})}\text{Cov}\left\{g(\boldsymbol{X}),\ \tilde{f}^*(\boldsymbol{X})\right\} \\ &+ \frac{E\tilde{f}(\boldsymbol{X})E\tilde{f}^*(\boldsymbol{X})}{(Eg(\boldsymbol{X})Eg^*(\boldsymbol{X}))^2}\text{Cov}\left\{g(\boldsymbol{X}),\ g^*(\boldsymbol{X})\right\}.\end{aligned}$$

Therefore, the estimated variance is

$$\mathrm{Var}\left\{\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}) - \widehat{\delta}_{\mathcal{C}_2}(\widehat{f})\right\} = V/n + V^*/n - 2\mathrm{Cov}\left(\widehat{\delta}_{\mathcal{C}_1}(\widehat{f}),\ \widehat{\delta}_{\mathcal{C}_2}(\widehat{f})\right).$$

Appendix C

Appendices for Chapter 4

C.1 Simulation Study for Multiple Trials with Different Treatment Effects

In this setting, the simulation models are slightly different across trials with different treatment effect sizes. The clinical outcome R for all the trials is generated by

$$R = \phi(\boldsymbol{X}) * A + 0.5 * X_1 + \epsilon, \quad \epsilon \sim N(0, 0.25)$$

$$\phi(\boldsymbol{X}) = X_1 - 0.5X_2 - \alpha X_3 - \beta X_4 + 0.3(X_5^2 > 0.64) - 0.3(X_6 > 0). \tag{C.1}$$

In the study with comprehensive feature variables (study 0), $X_1, \cdots, X_5$ are observed, only X_6 is an unobserved latent variable. The interaction effect of X_3 and X_4 are both $\alpha = \beta = 0.8$. However, in the study with low-resolution variables (study 1), only X_1, X_2 and X_3 were collected in the trial and can be used to estimate a coarsened ITR. The interaction effect of X_3 and X_4 were set to 0.1, meaning that both X_1 and X_4 don't contribute much to the selection of optimal treatment. All the feature variables $X_1, \cdots, X_6$, observed or unobserved, were generated independently and follow

standard normal distribution. Treatment A was randomly assigned from $\{-1, 1\}$ for each subject with equal probability.

In the validation data set, we are trying to simulate a more general population with treatment interaction effect between the data collected from study 0 and study 1. For validation setting 1 to 4, we set $\alpha = \beta$ as 0.3, 0.4, 0.5, and 0.6. Similar to Section 4.4.2, we applied integrative learning for high-resolution ITR (study 0) using coarsened ITR learned from study 1 and validated it on validation setting 1 to 4. Results from 100 repetitions are shown in Table C.1 and Figure C.1. In validation setting 1, ABLO-L performs better than ABLO-H with larger benefit and accuracy. It is also more efficient compared to all other methods with a smaller standard deviation in estimating benefit and value function. This is because the interaction effect of X_3 and X_4 is 0.3, which is relatively small and closer to the sample collected from study 1. Also, the coarsened ITR is simpler than high-resolution ITR and with fewer model assumptions, and therefore making the model less variable in estimating the benefit and value function. However, when the treatment effect gets larger in setting 2 to 4, Int-Ben always performs the best with the largest benefit and accuracy, also with the smallest standard deviation of benefit.

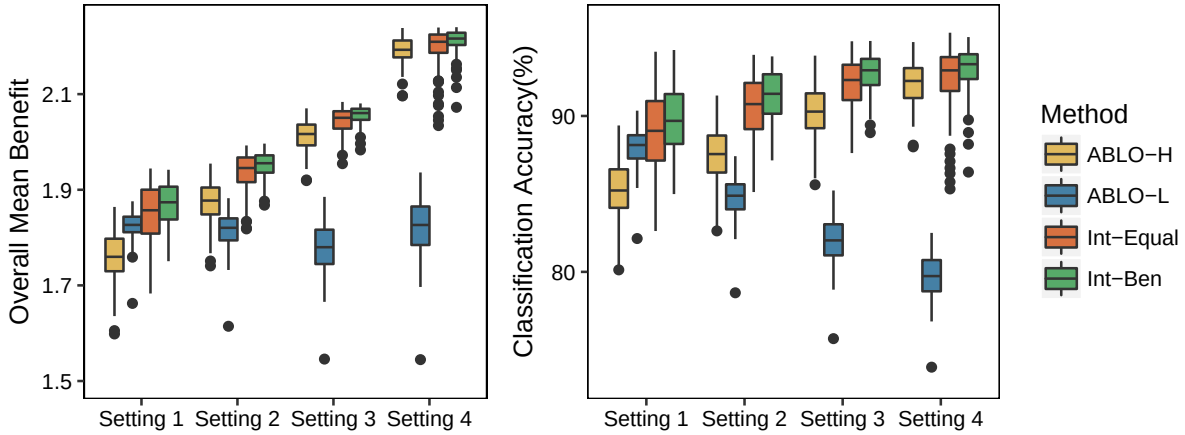
Next, we apply integrative learning for low-resolution ITR (study 1) using high-resolution ITR learned from study 0 and validated it on validation setting 1 to 4. Results are shown in Table C.2 and Figure C.2. In all validation settings, Int-Ben performs best in terms of a larger estimated benefit and smaller standard deviation by borrowing information from high-resolution ITR learned from study 0. Overall, the estimation of benefit became less efficient for all methods as the treatment effect increases since the validation data became more and more different to the study where low-resolution ITR was estimated.

Table C.1: Mean and standard deviation of overall benefit, value and accuracy rates for integrative learning of high-resolution ITR comparing to ITRs by ABLO on single studies

		ABLO-H	ABLO-L	Int-Equal	Int-Ben
Benefit	Setting 1	1.757(0.057)	1.824(0.028)	1.849(0.063)	1.868(0.047)
	Setting 2	1.874(0.044)	1.816(0.037)	1.935(0.041)	1.951(0.029)
	Setting 3	2.014(0.032)	1.779(0.053)	2.042(0.028)	2.056(0.019)
	Setting 4	2.190(0.029)	1.821(0.059)	2.195(0.044)	2.209(0.027)
Value	Setting 1	0.887(0.027)	0.913(0.014)	0.932(0.030)	0.941(0.022)
	Setting 2	0.938(0.022)	0.901(0.019)	0.967(0.020)	0.975(0.014)
	Setting 3	1.015(0.018)	0.901(0.027)	1.031(0.014)	1.038(0.010)
	Setting 4	1.089(0.016)	0.905(0.029)	1.092(0.023)	1.098(0.014)
Accuracy	Setting 1	0.852(0.020)	0.880(0.012)	0.890(0.029)	0.898(0.022)
	Setting 2	0.875(0.018)	0.848(0.013)	0.904(0.020)	0.913(0.017)
	Setting 3	0.903(0.017)	0.820(0.015)	0.919(0.016)	0.928(0.012)
	Setting 4	0.921(0.014)	0.797(0.013)	0.923(0.021)	0.930(0.015)

*: ABLO-H is estimated by ABLO using high-resolution data only;
ABLO-L is estimated by ABLO using low-resolution data only;
Int-Equal is estimated by integrative learning for high-resolution ITR with equal weights for all subjects;
Int-Ben is estimated by integrative learning for high-resolution ITR weighted by estimated benefit.

Figure C.1: Overall ITR benefit and optimal treatment allocation accuracy for the four methods



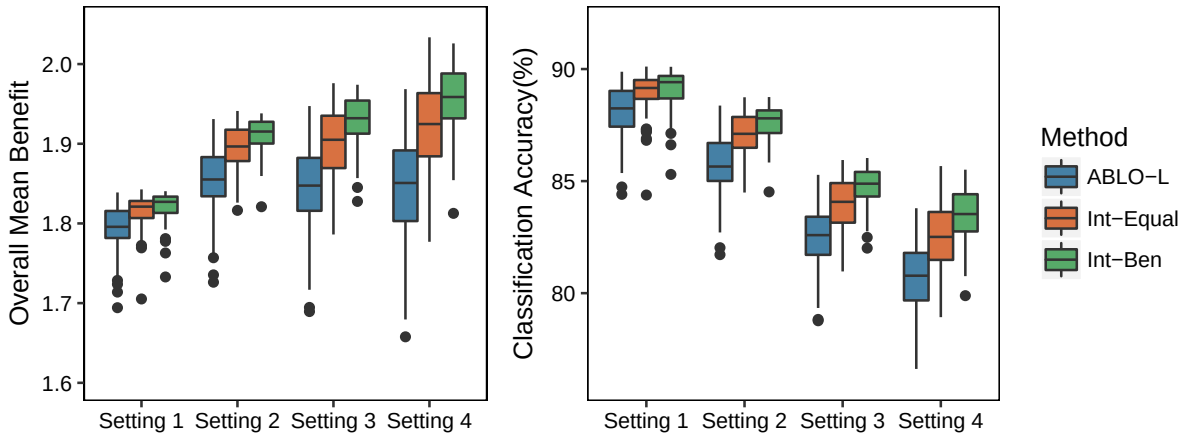
*: ABLO-H is estimated by ABLO using high-resolution data only;
ABLO-L is estimated by ABLO using low-resolution data only;
Int-Equal is estimated by integrative learning for high-resolution ITR with equal weights for all subjects;
Int-Ben is estimated by integrative learning for high-resolution ITR weighted by estimated benefit.

Table C.2: Mean and standard deviation of overall benefit, value and accuracy rates for integrative learning of low-resolution ITR comparing to ITRs by ABLO on single studies

		ABLO-L	Int-Equal	Int-Ben
Benefit	Setting 1	1.794(0.029)	1.816(0.020)	1.821(0.018)
	Setting 2	1.853(0.041)	1.895(0.029)	1.911(0.021)
	Setting 3	1.844(0.051)	1.903(0.042)	1.927(0.032)
	Setting 4	1.844(0.065)	1.922(0.056)	1.956(0.043)
Value	Setting 1	0.876(0.017)	0.890(0.011)	0.895(0.009)
	Setting 2	0.929(0.021)	0.950(0.014)	0.958(0.010)
	Setting 3	0.955(0.025)	0.983(0.021)	0.995(0.016)
	Setting 4	0.919(0.032)	0.958(0.028)	0.976(0.022)
Accuracy	Setting 1	0.881(0.012)	0.890(0.009)	0.891(0.008)
	Setting 2	0.857(0.014)	0.871(0.010)	0.876(0.008)
	Setting 3	0.825(0.013)	0.840(0.011)	0.847(0.009)
	Setting 4	0.806(0.015)	0.826(0.015)	0.835(0.012)

*: ABLO-L is estimated by ABLO using low-resolution data only;
Int-Equal is estimated by integrative learning for low-resolution ITR with equal weights for all subjects;
Int-Ben is estimated by integrative learning for low-resolution ITR weighted by estimated benefit.

Figure C.2: Overall ITR benefit and optimal treatment allocation accuracy for the three methods.



*: ABLO-L is estimated by ABLO using low-resolution data only;
Int-Equal is estimated by integrative learning for low-resolution ITR with equal weights for all subjects;
Int-Ben is estimated by integrative learning for low-resolution ITR weighted by estimated benefit.