Machine Learning Methods in Personalized Medicine Using Electronic Health Records

Peng Wu

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

2019

ABSTRACT

Machine Learning Methods in Personalized Medicine Using Electronic Health Records

Peng Wu

The theme of this dissertation focuses on methods for estimating personalized treatment using machine learning algorithms leveraging information from electronic health records (EHRs). Current guidelines for medical decision making largely rely on data from randomized controlled trials (RCTs) studying average treatment effects. However, RCTs are usually conducted under specific inclusion/exclusion criteria, they may be inadequate to make individualized treatment decisions in real-world settings. Large-scale EHR provides opportunities to fulfill the goals of personalized medicine and learn individualized treatment rules (ITRs) depending on patient-specific characteristics from real-world patient data. On the other hand, since patients' electronic health records (EHRs) document treatment prescriptions in the real world, transferring information in EHRs to RCTs, if done appropriately, could potentially improve the performance of ITRs, in terms of precision and generalizability. Furthermore, EHR data domain usually consists text notes or similar structures, thus topic modeling techniques can be adapted to engineer features.

In the first part of this work, we address challenges with EHRs and propose a machine learning approach based on matching techniques (referred as M-learning) to estimate optimal ITRs from EHRs. This new learning method performs matching method instead of inverse probability weighting as commonly used in many existing methods for estimating ITRs to more accurately assess

individuals' treatment responses to alternative treatments and alleviate confounding. Matching-based value functions are proposed to compare matched pairs under a unified framework, where various types of outcomes for measuring treatment response (including continuous, ordinal, and discrete outcomes) can easily be accommodated. We establish the Fisher consistency and convergence rate of M-learning. Through extensive simulation studies, we show that M-learning outperforms existing methods when propensity scores are misspecified or when unmeasured confounders are present in certain scenarios. In the end of this part, we apply M-learning to estimate optimal personalized second-line treatments for type 2 diabetes patients to achieve better glycemic control or reduce major complications using EHRs from New York Presbyterian Hospital (NYPH).

In the second part, we propose a new domain adaptation method to learn ITRs in by incorporating information from EHRs. Unless assuming no unmeasured confounding in EHRs, we cannot directly learn the optimal ITR from the combined EHR and RCT data. Instead, we first pre-train "super" features from EHRs that summarize physicians' treatment decisions and patients' observed benefits in the real world, which are likely to be informative of the optimal ITRs. We then augment the feature space of the RCT and learn the optimal ITRs stratifying by these features using RCT patients only. We adopt Q-learning and a modified matched-learning algorithm for estimation. We present theoretical justifications and conduct simulation studies to demonstrate the performance of our proposed method. Finally, we apply our method to transfer information learned from EHRs of type 2 diabetes (T2D) patients to improve learning individualized insulin therapies from an RCT.

In the last part of this work, we report M-learning proposed in the first part to learn ITRs using interpretable features extracted from EHR documentation of medications and ICD diagnoses codes. We use a latent Dirichlet allocation (LDA) model to extract latent topics and weights as features for learning ITRs. Our method achieves confounding reduction in observational studies through

matching treated and untreated individuals and improves treatment optimization by augmenting feature space with clinically meaningful LDA-based features. We apply the method to extract LDA-based features in EHR data collected at NYPH clinical data warehouse in studying optimal second-line treatment for T2D patients. We use cross validation to show that ITRs outperforms uniform treatment strategies (i.e., assigning insulin or another class of oral organic compounds to all individuals), and including topic modeling features leads to more reduction of post-treatment complications.

Table of Contents

List of Figures v

List of Tables viii

1 Background and Overview 1

1.1 Personalized Medicine and Individualized Treatment Rule 2

1.2 Q-learning and O-learning Methods 2

1.3 Methods for Observational Studies 3

1.4 Case Study: Electronic Health Records (EHR) 6

1.5 Topic Modeling in EHR data 8

2 Matched Learning for Personalized Treatment 10

2.1 Overview 10

2.2 Methodology 12

2.2.1 Individualized Treatment Rules (ITRs) 12

2.2.2 Matched Learning (M-learning) 13

2.2.3 Improved M-Learning 17

2.3 Theoretical Properties 19

2.3.1 Fisher Consistency . . . 19
2.3.2 Convergence Rate of M-Learning . . . 22
2.4 Simulation Studies . . . 26
2.5 Applications . . . 31
2.5.1 T2D Patient EHRs . . . 33
2.5.2 Handling Challenges of the Analyses of EHRs . . . 36
2.5.3 Analysis Results . . . 38
2.6 Discussion . . . 41

3 Domain Adaption Transfer Learning **46**
3.1 Overview . . . 46
3.2 Method to Improve ITRs by Borrowing Evidence from EHRs . . . 49
3.2.1 Learning the optimal ITR using RCT data . . . 49
3.2.2 Domain adaptation to improve learning ITRs . . . 50
3.2.3 Justification of domain adaptation learning . . . 53
3.3 Algorithms for Estimating ITRs . . . 57
3.4 Simulation Studies . . . 61
3.5 Applications . . . 65
3.5.1 Motivating Studies . . . 65
3.5.2 Analysis Results . . . 66
3.6 Discussion . . . 72

4 Learning Personalized Treatment Rule Using Topic Modeling Features **75**
4.1 Overview . . . 75

4.2 Methodology . . . 78
4.2.1 Review of Methods for Personalized Treatment . . . 78
4.2.2 EHR Data . . . 82
4.2.3 LDA-Based Features for Co-medication and Diagnosis . . . 84
4.3 Applications . . . 86
4.3.1 Cohort Identification . . . 87
4.3.2 LDA Feature Representation . . . 89
4.3.3 Learning Optimal ITR . . . 91
4.4 Discussion . . . 96

5 Conclusions and Future Direction **97**
5.1 Summary . . . 97
5.2 Limitations . . . 98
5.3 Extensions . . . 99

Bibliography **100**

A Appendices to Chapter 2 **112**
A.1 Additional Simulations Evaluating M-learning for Discrete Outcomes . . . 112
A.2 Additional Figures of Simulations and Real Data Analyses . . . 114

B Appendices to Chapter 3 **118**
B.1 Additional Simulation Results for Q-learning . . . 118
B.2 Additional Simulation Results for Unmeasured Confounder . . . 120

C Appendices to Chapter 4 **122**

C.1 EHR Data Preprocessing Flowchart 122

C.2 Measurement Pattern and Medication Figures 123

C.3 Learned Topics Visualization from LDA Model 128

List of Figures

2.1 Value comparison of four methods with propensity scores correctly specified (top panel) and misspecified (bottom panel). The numbers at the top of each subfigure are mean values. . . 29

2.2 Value comparison of four methods in the presence of unmeasured confounders. The numbers at the top of each subfigure are mean values. 31

2.3 Flowchart of EHR data processing procedures 32

2.4 T2D EHR Study Design 33

2.5 Heatmap of 17 extracted feature variables from EHRs of representative T2D patients in NYPH CDW. 35

2.6 Glucose values and measurement intensity (Time 0: time at first line treatment (MET) prescription) 37

2.7 Empirical value function of HbA1c in EHR data with 100 2-fold cross-validations (a low value is desirable) 42

2.8 Empirical value function of ICD diagnosis count in EHR data with 100 2-fold cross-validations (a low value is desirable) 42

3.1 Schematics of Proposed Domain Adaptation from EHR to RCT 54

3.2 Simulation Comparisons for M-learning (evaluated on independent testing sets generated from general or restricted population; scenario (i) has no latent tailoring variables while scenario (ii) has a latent tailoring variable not used in learning) . . . 64
3.3 t-SNE Plot for Features Extracted from CUMC EHRs and DURABLE trial 68
3.4 t-SNE Plot for Features of a DURABLE trial 70
3.5 Empirical Value Function of A1c Reduction in DURABLE Trial with 100 3-fold Cross-validations 71

4.1 Sunburst plot of treatment sequence for T2D patients* 88
4.2 Heatmap of LDA-Features Clustered by Patients 90
4.3 Association Network of Medication Prescriptions Based on Topics in LDA Model . . 92
4.4 Association Network of ICD9 Conditions Based on Topics in LDA Model 93
4.5 Empirical value function of ICD diagnosis count in EHR data with 100 2-fold cross-validations (a low value is desirable) 94
4.6 Feature Importance in M-learning with a Linear Kernel 96

A.1 HbA1c values and measurement intensity (Time 0: first stage treatment prescription) 115
A.2 Value comparison of four methods with propensity scores correctly specified (top panel) and misspecified (bottom panel). The numbers at the top of each subfigure are mean values. . . 116
A.3 Value comparison of four methods in the presence of unmeasured confounders. The numbers at the top of each subfigure are mean values. 117

B.1 Simulation Comparisons for Q-learning (evaluated on independent testing sets generated from general or restricted population; scenario (i) has no latent tailoring variables while scenario (ii) has a latent tailoring variable not used in learning) . . . 119

C.1 EHR Data Preprocessing Chart for Different Domains 122
C.2 Sample Patients Lab Tests Spikeplot 123
C.3 Sample Patients Medications Spikeplot 124
C.4 Sample Patients Glucose Spikeplot 125
C.5 Sample Patients HbA1c Spikeplot 125
C.6 Longitudinal Measurement Pattern (Gap Days) of HbA1c vs. Glucose 126
C.7 HbA1c Measurement Pattern vs. Medication (Gap Days) 127
C.8 Visualization of Co-medication Topic #3 and Their Most Relevant Terms 128
C.9 Visualization of Condition Topic #2 and Their Most Relevant Terms 129
C.10 Visualization of Condition Topic #5 and Their Most Relevant Terms 130

List of Tables

2.1 Cross-validated Empirical Value Function for HbA1c 43

2.2 Cross-validated Empirical Value Function for the Number of Major Complications . 43

3.1 HbA1c Reduction Comparing domain adaptation Learnings on DURABLE Trial (100 repetitions of 3-fold cross-validation) 71

4.1 Cross-validated Empirical Value Function for the Number of Major Complications Using Raw Count Data 94

4.2 Cross-validated Empirical Value Function for the Number of Major Complications Using LDA Features 95

B.1 Additional Results for Value Function and Success Rate 121

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my dissertation adviser, Dr. Yuanjia Wang, for her endless support and guidance over the last few years. Under her supervision, I learned how to define a research problem, work towards the solution with innovative thinking and collaborate with others. Her hardworking and passionate attitude inspired me to be a great statistician.

I would like to deeply thank Dr. Donglin Zeng, for his theoretical guidance and insightful support on the projects.

I would also like to thank the chair of my dissertation committee, Dr. Ken Cheung, and the other members of my committee: Dr. Christine Mauro, Dr. Chunhua Weng, and Dr. Scott Stroup for your thoughtful suggestions, which greatly improves my work.

To my family, thank you for your unconditional love and support. I especially want to thank my parents Liguo Wu and Yi Tang and my wife Dr. Shuling Liu, without your sacrifices and motivation, this would not have been possible.

To my parents, wife and daughter

Chapter 1

Background and Overview

Medical research is increasingly focusing on personalized treatment for an individual patient, which will gradually replace "one-size-fits-all" strategy in traditional medical practice. This transition is partially fueled by the recent advances in modern technologies that provide medical professionals and researchers with large-scale personal data (e.g. clinical measurements, biomarkers, pre-existing or developing conditions available in electronic health records). Selection of treatment strategy can be tailored towards each individual patient according to their person-specific characteristics, which is an important and beneficial move towards personalization of medical practice. The primary goal in this work is to merge statistical modeling techniques with machine learning algorithms to discover optimal individualized treatment rules (ITRs) applicable to various study designs and primary outcomes, including randomized experiments and observational studies with continuous or discrete outcomes. We will review how to achieve the goal of personalized medicine by assigning optimized ITRs to individual patients, existing methods for estimating ITRs, and introduce matching and propensity score based methods for observational studies. In addition, we will discuss critical issues of statistical analysis of electronic health records (EHRs), and demonstrate the utility of

our methods through case studies of a real world RCT data and EHRs extracted from New York Presbyterian Hospital (NYPH) and Columbia University Medical Center (CUMC). Lastly we will discuss topic modeling techniques in extracting latent topics in EHR data and engineering features.

1.1 Personalized Medicine and Individualized Treatment Rule

Personalized medicine calls for a paradigm shift from the universal strategy that assigns the same treatment to all patients affected by a disorder to selecting treatment strategies that optimize individual patient's health outcomes according to individual characteristics (Hamburg and Collins, 2010; Collins and Varmus, 2015). Improvements in technologies for collecting personal data, accompanied with developments of machine learning and statistical methods to analyze these data, hold promise to enable healthcare providers to prescribe the right therapy to the right patient at the right time (Collins and Varmus, 2015; Chakraborty and Moodie, 2013). By treating each patient with the optimal individualized treatment, patients can potentially gain enhanced clinical benefits, experience less side effects, and be more adherent to treatments (Chakraborty and Moodie, 2013). It is of interest to learn effective individualized treatment rules (ITRs) to reflect real world evidence using data collected in practical clinical settings (e.g., electronic health records, EHRs).

1.2 Q-learning and O-learning Methods

Over the past decade, a body of literature on how to accurately and efficiently estimate optimal individualized medical treatment using semiparametric methods and machine learning methods is growing in popularity (Murphy et al., 2007; Lavori and Dawson, 2004; Chakraborty and Moodie, 2013). Previously proposed machine learning approaches including Q-learning (Qian and Murphy,

2011), Outcome-weighted learning (O-learning) (Zhao et al., 2012) and augmented O-learning (Liu et al., 2018) provide valuable tools to effectively estimate simple and informative ITRs.

Q-learning was first proposed by Watkins (1989) as a form of reinforcement learning, in which "Q" represents "quality of action" (Watkins and Dayan, 1992; Watkins, 1989). Murphy, Qian et al. (2007, 2011) (Murphy et al., 2007; Qian and Murphy, 2011) implemented regression-based Q-learning to determine optimal ITR in each stage of a sequential multiple assignment randomized trials (SMARTs). The Q-function related to clinical outcomes is first approximated by a regression model and then maximized in each stage to select the optimal ITR. The ITR estimated using this regression-based approach may be sensitive to incorrect model assumptions in the estimation process, especially in high-dimensional feature space setting. Zhang et al. (2012) (Zhang et al., 2012) proposed an alternative approach referred as augmented inverse probability weighting (AIPW) to alleviate model misspecification and improve robustness. The method involves propensity score estimation which will be reviewed in the next section.

1.3 Methods for Observational Studies

Several methods discussed in the previous section mainly focus on randomized experiments. For observational studies, since treatments are not assigned randomly, without proper adjustments one cannot estimate an ITR with causal interpretation. Common practice to adjust for confounding due to observed covariates in the context of ITR estimation includes inverse probability weighting by propensity scores (i.e., the conditional probability of being assigned to a treatment given a subject's baseline covariates (Rosenbaum and Rubin, 1983)) or with additional augmentation in double robust estimation (Zhang et al., 2012, 2013). There are several potential limitations with

inverse probability weighted approaches. First, the approach requires a sophisticated model for estimating the propensity scores which often depends on some parametric and semiparametric models, and therefore subject to model misspecification; even if machine learning approach is used for this estimation such as classification and regression trees (CART), results can be sensitive to extreme weights and high variability of weights (Lee et al., 2010, 2011; Austin and Stuart, 2015). For example, the weights for observations with a very low propensity of being assigned to a particular observed treatment might be very unstable (Austin, 2011).Thus sensitivity analysis related to influential observations is recommended in the context of IPW approaches (Ellis et al., 2013). Second, it has been empirically observed that estimation of ITRs may be numerically unstable when treatment assignment is not balanced within some subgroups (Liu et al., 2018). Third, when the distribution of propensity scores have poor overlap between treatment arms, the inverse weighting approaches may perform inadequately due to imprecise propensity estimates and thus sensitive to misspecification (Crump et al., 2009).

Instead, a more robust and practically useful approach is based on matching. Matching has been shown to be a simple and effective adjustment in various empirical studies especially when model assumptions such as linearity is not be satisfied (Stuart and Green, 2008). Stuart (2010)(Stuart, 2010) provided a comprehensive review of matching methods for causal inference. In particular, they reviewed nearest neighborhood matching, subclassification, and full matching. Each of these methods has their own advantage in different practical applications. For example, ratio matching using nearest neighbor can ensure multiple good matching individuals from control group (or comparison treatment group) for an individual treated subject, especially when one treatment assignment is rare (Smith, 1997). Nearest neighbor matching with replacement is useful when treatment assignment is imbalanced and matching with replacement can reduce bias and avoid the order issue in

matching the treated units (Stuart, 2010; Dehejia and Wahba, 1999). Full matching, which can be considered as a complicated form of subclassification, has beneficial effects in optimizing average similarity between treated and control subjects within each matching set and makes good use of as many observations as possible (Stuart, 2010; Hansen, 2004; Stuart and Green, 2008).

In addition, propensity scores can also be incorporated to adjust for confounding in observational studies through matching or combination of weighting, matching and regression, and hence enable less model-dependent causal inference (Imai and Ratkovic, 2014; Ho et al., 2007). (Antonelli et al., 2018) proposed a doubly robust matching method by matching on both the propensity score and the prognostic score (Hansen, 2008) which is defined as the expected outcome for a subject under control treatment given his or her feature variables. They showed double robustness of this method, especially when dealing with confounding issues in high-dimensional feature space setting (Antonelli et al., 2018). Stuart et al. provided several guidance about matching methods for practical use: (1) one key issue is to determine whether a certain set of features included for matching is appropriate according to ignorability assumption; (2) similarity measure plays a crucial role in matching and one can choose similarity measure best suited for a particular application. For instance, Mahalanobis distance or exact matching within propensity score calipers could be utilized to achieve close balance on a small set of features (Stuart, 2010), and many data-driven similarity measures have been proposed to accommodate categorical or ordinal feature variables (Boriah et al., 2008).

Compared to IPW-based methods which rely on whether or not the propensity score is precisely and accurately estimated (Austin, 2011; Rubin, 2004), matching-based methods require less model specification and can be nonparametric (Ho et al., 2007). IPW-based methods ensure different treatment groups have similar distribution of confounders at the population level, while some matching methods ensure balanced distribution at subgroup level and provide more flexible tools to

control the quality of matching important confounders in subgroups or even on individual subjects. For example, covariates selection, distance metric and measure of covariates balance can be combined to optimize matching (Sekhon and Grieve, 2012). Since our goal is to achieve personalized medicine and identify optimal ITRs for subjects with a given set of feature variables, exploring matching methods to balance subgroup-level distributions is more desirable.

In conclusion, matching methods are important tools to achieve covariates balance and adjusting for confounding in observational studies and holds advantages compared to inverse probability weighting methods in various applications (Stuart, 2010; Rubin, 2004; Austin, 2011; Ho et al., 2007). For instance, one can identify matching subjects to guarantee numerical stability, especially when some subgroup rarely experience one particular treatment. All existing methods for estimating optimal ITR using observational studies involve various forms of inverse probability weighting, and none leverages the advantage of matching. In this dissertation, we will develop novel methods for ITR estimation that make use of the flexible and power general framework of matching applicable to both randomized experiment and observational studies. Matched learning may perform better especially when the treatment assignment is not balanced in some subgroups of RCT or in observational studies where researchers do not have control over treatment assignment mechanism.

1.4 Case Study: Electronic Health Records (EHR)

In recent years, the use of EHR is continuously growing among clinical researchers with access to large-scale clinical data warehouses and databases (Weiskopf and Weng, 2013; Weiskopf et al., 2013). EHR data resources contain massive information which may allow researchers to answer clinical questions at the subject-level, including the estimation of ITRs (Wang et al., 2016). Notwithstand-

ing EHR providing massive valuable healthcare information on large patient population, addressing research problems relying on EHR might face potential challenges since EHR data is not collected for research oriented studies in the first place, but primarily for billing or other clinical purposes (Haneuse and Daniels, 2016).

Hripcsak and Albers (2013) summarized several critical issues in clinical research using EHR data: completeness, accuracy, complexity, and bias (Hripcsak and Albers, 2013). In EHR, missing data is commonly seen and patients with sufficient complete clinical information may comprise only a small portion. The incompleteness in EHR data is typically related to selection bias in healthcare practice, which can be casted as a missing data problem. In order to properly address the issue, assumptions on the missing data mechanism or selection process should be taken careful consideration (Haneuse and Daniels, 2016; Little and Rubin, 2014). One simple method to handle missing data is to run "complete case analysis" by omitting the records with missing information, which requires strong assumption coined as missing completely at random (MCAR)(Little and Rubin, 2014) that often does not hold for EHR data. In EHR context, some variables might provide partial information on the missingness of the others, that is to say, using imputation to handle missingness is valid under the assumption of missing at random (MAR) (Wells et al., 2013; Rubin et al., 1995). Accuracy is another important issue for EHR research since EHR is almost never error free due to complexities in electronic-based data collection, integration and management process (Hogan and Wagner, 1997). Furthermore, complicated hierarchies of EHR data and temporal attributes carry different levels of uncertainty which might require both medical domain knowledge and statistical insights to better characterize patient information in EHR (Hripcsak and Albers, 2013; Tao et al., 2011). In contrast to selection bias, confounding bias has drawn extensive attention in literature (Pearl, 2009; Bareinboim and Pearl, 2016). The non-experimental feature of EHR data

collection brings major difficulties to making inferences and drawing valid conclusions (Wang et al., 2016; Hripcsak et al., 2011). As mentioned in previous section, methods on propensity scores, subclassification, and matching are emerging to deal with various biases in EHR (Haneuse and Daniels, 2016). In this dissertation, we will compare effectiveness of several alternative methods applied to EHR collected at NYPH and CUMC clinical data warehouse (CDW).

1.5 Topic Modeling in EHR data

EHR data contains rich patient level features to build effective ITRs. Some information in EHRs is documented in the form of text notes or similar structure that enable researchers to use topic modeling techniques to better extract the information. Topic modeling techniques including latent semantic analysis (LSA) (Landauer et al., 1998), probabilistic latent semantic analysis (pLSA) (Hofmann, 1999) have been widely used to represent features in a lower dimensional space. LSA is a well-known approach that maps count type data of documents to a reduced latent semantic space based on Singular Value Decomposition (SVD) while pLSA improves LSA in a statistical way that relies on a mixture decomposition (Hofmann, 1999). Furthermore, latent Dirichlet allocation (LDA) is proposed as another generative probabilistic model, which overcomes some disadvantages of pLSA such as generalizability to new documents and overfitting problems (Blei et al., 2003).

The rest of this dissertation is organized as follows. In Chapter 2, we introduce our proposed matching-based learning method (M-learning) to estimate ITR. We show the advantages of M-learning over two existing methods through simulation studies and apply it in a real world observational study. In Chapter 3, we propose a new framework in transferring information learned from observational study in ITR estimation to randomized clinical trial. Moreover, we extend M-learning

in Chapter 2 integrating with kernel method. In Chapter 4, we present a topic model based feature extraction approach in two domains in EHR data and apply M-learning for ITR estimation using the engineered features. We conclude this dissertation with limitations and future directions.

Chapter 2

Matched Learning for Personalized Treatment

2.1 Overview

Machine learning approaches provide valuable tools to estimate individualized treatment rules (ITRs) and dynamic treatment rules (DTRs) due to their powerful computing capabilities. Previously proposed machine learning approaches include Q-learning (Watkins and Dayan, 1992; Qian and Murphy, 2011), outcome weighted learning (O-learning) (Zhao et al., 2012), robust O-learning (AOL) (Kang et al., 2014; Liu et al., 2018) and subgroup identification methods(Fu et al., 2016). Most of these existing methods focus on analyzing randomized clinical trial (RCT) data. However, the ITRs estimated from RCTs may be inadequate to assist individualized treatment decision making in real-world settings due to stringent inclusion/exclusion criteria of RCTs, a lack of generalizability, and a lack of evidence for long-term outcomes.

Large-scale electronic health records (EHRs) provide new opportunities to learn ITRs using

real-world patient data. In recent years, access to clinical data warehouses and databases continues to grow and an increasing trend of using EHRs for scientific research is observed (Weiskopf and Weng, 2013; Hripcsak and Albers, 2013; Hripcsak et al., 2016). As exclusive evidence generated from clinical trials is inadequate due to a lack of external validity, EHRs can serve as an important complement to evidence-based research for personalized medicine. For instance, a broad range of real-world medication use patterns not captured by RCTs were observed in EHRs (Hripcsak et al., 2016). Furthermore, as compared to RCTs, using EHRs to learn ITRs has benefits such as containing information on a large population over relatively longer time frames that reflects patients' care management and disease course in more realistic settings.

However, EHRs are not collected for research purposes and conducting research with EHRs encounters great challenges. Critical issues including confounding bias and selection bias have been discussed: completeness, accuracy, complexity, and bias (Hripcsak and Albers, 2013; Haneuse, 2016). In the context of estimating ITRs, common practice to adjust for confounding is inverse probability weighting (IPW) of propensity scores with or without augmentation to achieve double robustness (Zhang et al., 2012, 2013). The IPW approach needs a sophisticated model to estimate propensity scores with high accuracy. Machine learning methods are thus proposed to predict propensity scores (Lee et al., 2010, 2011; Austin and Stuart, 2015), but they may result in extreme weights with high variability. In addition, the IPW approaches may not adequately balance covariate distributions between treatment groups, especially when the distribution of propensity scores has less overlap between treatment arms (Crump et al., 2009).

On the other hand, matching has been successfully used to estimate population average treatment effects, including ratio matching (Smith, 1997), nearest neighbor matching (Dehejia and Wahba, 1999), and full matching (Stuart, 2010; Hansen, 2004). However, to the best of our knowl-

edge, there is no method to leverage advantages of matching to estimate personalized treatment rules and apply to observational data such as EHRs.

In this chapter, we propose a machine learning approach, namely, Matched Learning (M-learning), to estimate ITRs through matching treated and untreated subjects with an application to EHRs. M-learning is a general framework that includes O-learning and AOL as special cases. M-learning introduces matching-based value functions to match individual treatment responses under alternative treatments and alleviate confounding. Various matching functions can be used to compare outcomes for matched pairs to accommodate different types of data (continuous, discrete, or ordinal) under a unified framework. The efficiency of M-learning can be improved by a denoise procedure and double robust matching. The implementation is based on a matched-pairs weighted support vector machine (SVM). We establish the Fisher consistency and convergence rate of M-learning and conduct extensive simulation studies. We show that M-learning outperforms existing methods when propensity scores are misspecified and in certain scenarios when unmeasured confounders are present. Lastly, we tackle challenges of EHRs, including confounding by indication, confounding bias, and selection bias, and apply M-learning to estimate the optimal second-line treatments for type 2 diabetes (T2D) patients to achieve better glycemic control or reduce major complications.

2.2 Methodology

2.2.1 Individualized Treatment Rules (ITRs)

Let H_i denote the pre-treatment covariates and let A_i denote the binary treatment assignment taking values from $\{-1, 1\}$. Let R_i denote the clinical outcome post treatment (reward), and

assume a larger R_i is more desirable (e.g., symptom reduction). An ITR is a decision rule, $\mathcal{D}(H_i)$, that maps the domain of H_i to the treatment choices in $\{-1, 1\}$. The value function associated with $\mathcal{D}$ used to evaluate an ITR is defined as the expected post-treatment outcome by following $\mathcal{D}$ to assign treatments, that is, $V(\mathcal{D}) = E^{\mathcal{D}}(R_i)$.

For RCTs, the assumption that the potential outcomes are independent of treatment assignment given covariates is satisfied, and the treatment assignment probability, denoted by $\pi(a, h) = \Pr(A_i = a|H_i = h)$, is known by design. O-learning proceeds by re-expressing the value function as $V(\mathcal{D}) = E\left[\frac{I(A_i = \mathcal{D}(H_i))R_i}{\pi(A_i, H_i)}\right]$, and then aims to maximize the empirical value function defined as

$$V_n(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i = \mathcal{D}(H_i))R_i}{\pi(A_i, H_i)}. \tag{2.1}$$

In an observational study, however, treatment propensities $\pi(A_i, H_i)$ are unknown and need to be estimated from data. Using the objective function (2.1) and IPW-based methods in observational studies suffer from instability and increased variance especially when weights are highly variable. In addition, IPW-based methods do not directly control the balance of covariate distributions between treatment groups.

2.2.2 Matched Learning (M-learning)

When comparing different treatment responses, matching methods can be designed to ensure balanced distribution at subgroup level and provide more flexible tools to control the matching quality of important confounders in subgroups or even on individual subjects. For example, covariates selection, distance metric and measure of covariates balance can be combined to optimize matching (Sekhon and Grieve, 2012) and identify matching subjects to guarantee numerical stability, especially when some subgroup of patients rarely receive one particular treatment. Denote the matched

set for subject i as $\mathcal{M}_i$, which consists of subjects with opposite treatments but similar covariates as subject i, where similarity is defined under a suitable distance metric. That is, we let

$$\mathcal{M}_i = \{j : A_j = -A_i, d(H_j, H_i) \leq \delta_i\},$$

where $d(\cdot, \cdot)$ is a metric defined in the covariate space and δ_i is a pre-specified positive threshold to determine the size of the matched set which may vary across subjects. For example, if we choose $\mathcal{M}_i$ to be the nearest neighbor, then δ_i is the minimal distance between subject i and any other subject with the opposite treatment. In some applications, subjects with empty matching sets may be excluded. In this chapter, we use nearest neighbor in the matching step of M-learning in the simulations and application, and study its theoretical properties.

M-learning is developed to maximize a matching-based value function defined in (3.1). The motivation of M-learning is that when two subjects are matched in confounders or propensity scores of treatments but are observed to receive opposite treatments, the subject with a larger clinical outcome should be more likely to have received the optimal treatment among two options. Based on this rationale, one expects that if $j \in \mathcal{M}_i$ and $R_j \geq R_i$, then the optimal ITR for subject i should more likely to be A_j, and vice versa. Furthermore, the likelihood is expected to be greater if the difference between R_j and R_i is larger. Specifically, for any given ITR $\mathcal{D}$, define the matching-based value function as

$$\begin{aligned} V_n(\mathcal{D}; g) &= n^{-1} \sum_{i=1}^{n} |\mathcal{M}_i|^{-1} \sum_{j \in \mathcal{M}_i} \Big\{ I(R_j \geq R_i, \mathcal{D}(H_i) = -A_i) \\ &\quad + I(R_j \leq R_i, \mathcal{D}(H_i) = A_i) \Big\} g(|R_j - R_i|), \end{aligned} \tag{2.2}$$

where $|\mathcal{M}_i|$ is the size of $\mathcal{M}_i$ and $g(\cdot)$ is a monotonically increasing function specified by users to weight different pairs of subjects. Typical choices of $g(\cdot)$ can be $g(x) = 1$ or $g(x) = x$. Furthermore,

let $\mathcal{D}(H) = \text{sign}(f(H))$ for some ITR decision function f, then the matching-based value function (3.1) is equivalent to

$$V_n(f; g) = n^{-1} \sum_{i=1}^{n} |\mathcal{M}_i|^{-1} \sum_{j \in \mathcal{M}_i} I(f(H_i) A_i \text{sign}(R_j - R_i) \leq 0) g(|R_j - R_i|).$$

M-learning maximizes $V_n(f; g)$, or equivalently, minimizes

$$n^{-1} \sum_{i=1}^{n} |\mathcal{M}_i|^{-1} \sum_{j \in \mathcal{M}_i} I(f(H_i) A_i \text{sign}(R_j - R_i) \geq 0) g(|R_j - R_i|), \tag{2.3}$$

in order to identify the optimal ITR.

The objective function (2.3) can be further expanded by allowing $\mathcal{M}_i = i$ (match subject i with himself/herself). If in addition we replace R_j in (2.3) by zero (when $R_j > 0$ for all subjects) or the smallest observed outcome when negative outcomes are present and choose $g(x) = x$, M-learning reduces to the original O-learning in (Zhao et al., 2012). Similarly, if we replace R_j by subject i's predicted outcome estimated from a parametric model including only the main effects of H_i, M-learning reduces to the single-stage AOL in (Liu et al., 2018). Thus, O-learning and single-stage AOL are special cases of M-learning, where they compare the observed outcome R_i with a constant or the predicted outcome given H_i averaged across treatments. In contrast, M-learning compares observed individual outcomes from two subjects in the matched set, where the treatment assignment is approximately "random" given H_i but the received treatments are opposite. Thus, M-learning is more informative in taking account of information on patient's outcome at the individual level (R_i and R_j), instead of comparing a patient's outcome with the predicted outcome averaged over treatments (as done in O-learning or AOL).

Minimizing the matching-based value function (2.3) is not feasible due to the discontinuity of the indicator function. Similar to O-learning, we replace the zero-one loss by other surrogate loss functions. In particular, when using the hinge-loss, the objective function to be optimized is the

loss function for the weighted support vector machine (SVM) with matched pairs:

$$V_{n,\phi}(f;g) = n^{-1}\sum_{i=1}^{n}|\mathcal{M}_i|^{-1} \sum_{j\in\mathcal{M}_i}\phi(-f(H_i)A_i\text{sign}(R_j-R_i))g(|R_j-R_i|)+\lambda_n\|f\|_{\mathcal{H}_K}, \quad (2.4)$$

where $\phi(x)=(1-x)_+$, λ_n is a tuning parameter and $\mathcal{H}_K$ is a reproducing kernel Hilbert space (RKHS) with kernel function $K(\cdot,\cdot)$. The solution to M-learning is obtained by minimizing $V_{n,\phi}(f;g)$. In terms of implementation, the dual problem of (4.5) is a quadratic problem which can be solved by any off-the-shelf quadratic programming packages.

Taking linear ITR decision rules as an example, we describe solution to the quadratic programming problem using Lagrange multipliers. Assume f in $V_{n,\phi}(f;g)$ is linear and $f(h)=\langle\beta,h\rangle+\beta_0$ where $\langle\cdot,\cdot\rangle$ denotes the inner product operator and $\|f\|_{\mathcal{H}_K}$ represents $\|f\|^2$ in Euclidean space. It is computationally convenient to re-write (4.5) in an equivalent form as

$$\min\frac{1}{2}\|\beta\|^2+C\sum_{i=1}^{n}\sum_{j\in\mathcal{M}_i}|\mathcal{M}_i|^{-1}g(|R_j-R_i|)\xi_{ij},$$

$$\text{subject to: } A_i\text{sign}(R_i-R_j)(\langle\beta,H_i\rangle+\beta_0)\geq(1-\xi_{ij}),\xi_{ij}\geq 0,\forall i \text{ and } j\in\mathcal{M}_i,$$

where ξ_{ij} is a slack variable that represents misclassification error for the jth subject in the matched set of the ith subject, C is a cost parameter, and $|\mathcal{M}_i|^{-1}g(|R_j-R_i|)$ is the individual-specific weight in a weighted SVM framework.

The Lagrange primal function follows as

$$\frac{1}{2}\|\beta\|^2+C\sum_{i=1}^{n}\sum_{j\in\mathcal{M}_i}|\mathcal{M}_i|^{-1}g(|R_j-R_i|)\xi_{ij}$$
$$-\sum_{i=1}^{n}\sum_{j\in\mathcal{M}_i}\alpha_{ij}\{A_i\text{sign}(R_i-R_j)(H_i^T\beta+\beta_0)-(1-\xi_{ij})\}-\sum_{i=1}^{n}\sum_{j\in\mathcal{M}_i}\mu_{ij}\xi_{ij},$$

where we minimize with respect to β, β_0 and ξ_{ij}. By taking the respective derivatives and setting them to zero to obtain,

$$\begin{cases} \beta = \sum_{i=1}^{n} \sum_{j \in \mathcal{M}_i} \alpha_{ij} A_i \text{sign}(R_i - R_j) H_i, \\ 0 = \sum_{i=1}^{n} \sum_{j \in \mathcal{M}_i} \alpha_{ij} A_i \text{sign}(R_i - R_j), \\ \alpha_{ij} = C|\mathcal{M}_i|^{-1} g(|R_j - R_i|) - \mu_{ij}, \forall i \text{ and } j \in \mathcal{M}_i. \end{cases}$$

By substituting above equations into Lagrangian dual function, we obtain

$$\max \sum_{i=1}^{n} \sum_{j \in \mathcal{M}_i} \alpha_{ij} - \frac{1}{2} \sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{j \in \mathcal{M}_i} \sum_{j' \in \mathcal{M}_{i'}} \alpha_{ij} \alpha_{i'j'} A_i A_{i'} \text{sign}(R_i - R_j) \text{sign}(R_{i'} - R_{j'}) \langle H_i, H_{i'} \rangle$$

subject to $0 \leq \alpha_{ij} \leq C|\mathcal{M}_i|^{-1} g(|R_j - R_i|)$ and $\sum_{i=1}^{n} \sum_{j \in \mathcal{M}_i} \alpha_{ij} A_i \text{sign}(R_i - R_j) = 0$. In addition, subject to Karush-Kuhn-Tucker conditions for $\forall i$ and $j \in \mathcal{M}_i$ (Zhao et al., 2012):

$$\begin{cases} \alpha_{ij}[A_i \text{sign}(R_i - R_j)(H_i^T \beta + \beta_0) - (1 - \xi_{ij})] = 0, \\ \mu_{ij} \xi_{ij} = 0, \\ A_i \text{sign}(R_i - R_j)(H_i^T \beta + \beta_0) - (1 - \xi_{ij}) \geq 0, \end{cases}$$

the solution to the primal and dual problem is optimal. It is straightforward to extend the algorithm to other kernels (e.g., Gaussian kernel) and obtain a nonparametric ITR based on kernel function $K(\cdot, \cdot)$ in the RKHS.

2.2.3 Improved M-Learning

To improve the performance of M-learning, we use a de-noise procedure first reported in (Liu et al., 2018). We replace R_i by a surrogate residualized outcome $\widetilde{R}_i = R_i - s(H_i)$ in $V_n(\mathcal{D}; g)$ for any measurable function of H_i, denoted as $s(H_i)$. These residualized outcomes remove the main effects of covariates, which improves efficiency of identifying tailoring variables exhibiting quantitative or

qualitative interaction with treatment. The residuals can be obtained through a regression model and the value function to be maximized becomes

$$\begin{aligned} V_n(\mathcal{D}; g) &= n^{-1} \sum_{i=1}^{n} |\mathcal{M}_i|^{-1} \sum_{j \in \mathcal{M}_i} \Big\{ I(\widetilde{R}_j \geq \widetilde{R}_i, \mathcal{D}(H_i) = -A_i) \\ &\quad + I(\widetilde{R}_j \leq \widetilde{R}_i, \mathcal{D}(H_i) = A_i) \Big\} g(|\widetilde{R}_j - \widetilde{R}_i|). \end{aligned}$$

As shown in (Liu et al., 2018), by removing the main effects of covariates, more stable weights are used in the weighted SVM to boost efficiency in estimating ITRs.

Furthermore, prognostic scores can be incorporated into M-learning under the framework of doubly robust matching estimator (DRME) proposed in (Antonelli et al., 2018). The DRME uses both propensity scores and prognostic scores to construct a matching set $\mathcal{M}(i, \theta)$, where $\theta = (\theta_1, \theta_2)^T$ denotes parameters for the propensity score and prognostic score models:

$$\pi(H) = P(A = 1|H) = u_1(H^T\theta_1), m(H) = E(R|A = -1, H) = u_2(H^T\theta_2). \tag{2.5}$$

(Antonelli et al., 2018) showed that only one of the two models in (2.5) is required to be correctly specified to ensure consistency of DRME, which achieves double robustness. Applying DRME to M-learning, both propensity scores and prognostic scores will be included in the matching step to create informative matched pairs. The doubly robust M-learning is consistent even if one of the propensity score model or prognostic model is misspecified, and it will be more efficient than regular M-learning if both models are correctly specified. Note that M-learning can be applied to RCT data where only prognostic scores need to be included in the matching step to improve efficiency.

2.3 Theoretical Properties

In this section, we establish the theoretical properties including Fisher consistency, different choices of $g(x)$ and convergence rate of of M-learning.

2.3.1 Fisher Consistency

Theorem 2.3.1 *Under regularity assumptions including* $\max_{i=1}^{n} \delta_i \to 0$*, and that the density of* H *and* $E[R|H, A = 1]$ *is continuously differentiable in the support of* H*, it holds that*

$$V_n(f, g) \to_{a.s} V(f, g),$$

where

$$V(f; g) = E\left\{\widetilde{E}\left[I(f(H)A\text{sign}(\widetilde{R} - R) \leq 0)g(|\widetilde{R} - R|)\Big|\widetilde{A} = -A, \widetilde{H} = H\right]\right\},$$

$\widetilde{E}$ *is the expectation with respect to* $(\widetilde{R}, \widetilde{H}, \widetilde{A})$*, an independent copy of* (R, H, A)*. In addition, define*

$$\Delta_g(r, h) = E\left[\frac{g(|R - r|)}{|R - r|}(R - r)|A = 1, H = h\right] - E\left[\frac{g(|R - r|)}{|R - r|}(R - r)|A = -1, H = h\right],$$

then for any h *in the support of* H*,*

$$sign(f^*(h)) = sign \int_r \Delta_g(r, h) dF(r|H = h),$$

where $F(r|H = h)$ *is the distribution of* $R = r$ *given* $H = h$ *and* f^* *is the optimal function minimizing* $V(f; g)$*.*

Proof of Theorem 2.3.1. After some algebra, we can show that the value function is equal to

$$E\left[I(f(H) > 0)\left\{\widetilde{E}[\frac{g(|\widetilde{R} - R|)}{|\tilde{R} - R|}(\tilde{R} - R)_+|\widetilde{A} = 1, \widetilde{H} = H]\right.\right.$$
$$\left.\left. + \widetilde{E}[\frac{g(|\widetilde{R} - R|)}{|\tilde{R} - R|}(\tilde{R} - R)_-|\widetilde{A} = -1, \widetilde{H} = H]\right\}\right]$$

$$+E\left[I(f(H)\le 0)\left\{\widetilde{E}[\frac{g(|\widetilde{R}-R|)}{|\tilde{R}-R|}(\tilde{R}-R)_+|\widetilde{A}=-1,\widetilde{H}=H]\right.\right.$$
$$\left.\left.+\widetilde{E}[\frac{g(|\widetilde{R}-R|)}{|\tilde{R}-R|}(\tilde{R}-R)_-|\widetilde{A}=1,\widetilde{H}=H]\right\}\right].$$

Hence, the optimal decision function, denoted by $f^*(H)$, should have the same sign as

$$\begin{aligned}
& E\left[\left\{\widetilde{E}[\frac{g(|\widetilde{R}-R|)}{|\tilde{R}-R|}(\tilde{R}-R)_+|\widetilde{A}=1,\widetilde{H}=H]\right.\right.\\
&\qquad\left.\left.+\widetilde{E}[\frac{g(|\widetilde{R}-R|)}{|\tilde{R}-R|}(\tilde{R}-R)_-|\widetilde{A}=-1,\widetilde{H}=H]\right\}|H\right]\\
&-E\left[\left\{\widetilde{E}[\frac{g(|\widetilde{R}-R|)}{|\tilde{R}-R|}(\tilde{R}-R)_+|\widetilde{A}=-1,\widetilde{H}=H]\right.\right.\\
&\qquad\left.\left.+\widetilde{E}[\frac{g(|\widetilde{R}-R|)}{|\tilde{R}-R|}(\tilde{R}-R)_-|\widetilde{A}=1,\widetilde{H}=H]\right\}|H\right]\\
=\;& E\left[\widetilde{E}[\frac{g(|\widetilde{R}-R|)}{|\tilde{R}-R|}(\tilde{R}-R)|\widetilde{A}=1,\widetilde{H}=H]\Big|H\right]\\
&-E\left[\widetilde{E}[\frac{g(|\widetilde{R}-R|)}{|\tilde{R}-R|}(\tilde{R}-R)|\widetilde{A}=-1,\widetilde{H}=H]\Big|H\right].
\end{aligned}$$

In other words, if we define

$$\Delta_g(r,h)=E\left[\frac{g(|R-r|)}{|R-r|}(R-r)|A=1,H=h\right]-E\left[\frac{g(|R-r|)}{|R-r|}(R-r)|A=-1,H=h\right],$$

then for any h in the support of H,

$$\text{sign}(f^*(h))=\text{sign}\int_r\Delta_g(r,h)dF(r|H=h),$$

where $F(r|H=h)$ is the distribution of $R=r$ given $H=h$. □

Here we make a few remarks.

Remark 1. When $g(x)=x$ and $r=0$, i.e. $\Delta_g(r,h)=E(R|A=1,H=h)-E(R|A=-1,H=h)$, Theorem 2.3.1 implies that the optimal treatment rule obtained from M-learning is the same as the optimal rule from O-learning, and thus M-learning is Fisher consistent for the usual optimal ITR.

Remark 2. When $g(x) = 1$, we obtain

$$\begin{aligned}
\Delta_g(r,h) &= E\left[\text{sign}(R-r)|A=1,H=h\right] - E\left[\text{sign}(R-r)|A=-1,H=h\right] \\
&= P(R>r|A=1,H=h) - P(R<r|A=1,H=h) \\
&\quad -[P(R>r|A=-1,H=h) - P(R<r|A=-1,H=h)] \\
&= 2[P(R>r|A=1,H=h) - P(R>r|A=-1,H=h)].
\end{aligned}$$

Remark 2 suggests that for subjects with $H = h$, the optimal rule chooses the treatment with a higher probability of having a greater outcome than the average outcome across treatments. Such choice of $g(x)$ ensures robustness against outliers of R. When R is an ordinal or binary random variable, this choice is especially suitable. For example, consider an ordinal outcome with three levels, then the optimal rule $f^*(h)$ has a desirable property

$$\text{sign}(f^*(h)) = \text{sign}[\text{AUC}_{13}(h) - \text{AUC}_{23}(h)], \tag{2.6}$$

where $\text{AUC}_{jk}(h)$ is the conditional AUC for comparing $R = j$ with $R = k$ for subjects with $H = h$. More generally, the function $\Delta_g(r,h)$ is similar to creating comparisons based on a reference level r of the outcome. Therefore, for a particular target value r (e.g., the value under a universal "one-size-fits-all" treatment assignment, or a clinically meaningful level for an ordinal outcome), one can construct $g(x)$ so that the weights concentrate on the difference from the reference value r.

Remark 3. Lastly, when applied to observational studies, the condition of no unmeasured confounders ensures that the optimal rule estimates the treatment with a higher potential outcome, since $\Delta_g(r,h) = E(R^{(1)}|H = h) - E(R^{(-1)}|H = h)$, where $R^{(k)}$ denotes the potential outcome under treatment k.

2.3.2 Convergence Rate of M-Learning

In this section, we establish the convergence rate of the risk bound for the estimated decision rule. We consider the nearest neighborhood matching, $\mathcal{H}_K$ is the RKHS based on a Gaussian kernel function with bandwidth σ_n, and assume R and H are bounded. Furthermore, we need the following assumptions:

(A.1) The density of $H = h$ with respect to the dominating measure and $E(R|A = a, H = h)$ are continuously differentiable in H's support for $a = -1$ and 1. Moreover, the density of H is bounded from below on the support of H, denoted by $\mathcal{X}_H$.

(A.2) The probability measure has a geometric noise exponent $\alpha > 0$ as in Definition 2.3 of (Steinwart and Scovel, 2007). That is, if let τ_H be the distance from any H to the decision boundary $\{h : f^*(h) = 0\}$, it holds

$$E[|f^*(H)| \exp\{-\tau_H^2/t\}] \leq ct^{\alpha d/2}, \quad t > 0.$$

(A.3) There exists $\gamma > 0$ and $r_0 > 0$ such that $|\mathcal{X}_H \cap B(h, r)| \geq \gamma |B(h, r)|$ for any $h \in \mathcal{X}_H$ and $0 < r < r_0$, where $B(h, r)$ is a ball centered at h with radius r, and $|A|$ denotes the volume of set A in $\mathcal{X}_H$.

Condition (A.1) is necessary to ensure the conistency of approximation in the nearest-neighbor based matching. Condition (A.2) is commonly assumed for SVMs and a similar condition has been considered for classification problem (c.f., (Steinwart and Scovel, 2007)) and establishing the learning rate for ITRs (Zhao et al., 2012). When the decision rule is completely separable, the exponent α can be as large as possible. The third condition (A.3) is used to obtain the convergence for the nearest-neighbor estimator (Devroye et al., 2013)

Theorem 2.3.2 *Under the above assumptions and letting* $\sigma_n = \lambda_n^{1/p(1+\alpha)}$, *it holds*

$$V(f^*;g) - V(\widehat{f};g) = O_p\left[\frac{1}{\sqrt{n}\lambda_n^{\beta_1}} + \frac{1}{\lambda_n^{\beta_2}}\left\{\left(\frac{m_n}{n}\right)^{1/p} + \sqrt{\frac{\log n}{m_n}}\right\} + \lambda_n^{\alpha/(1+\alpha)}\right],$$

where $\beta_1 = p/4 + (1/2 - p/8)d/[(1+\alpha)]$, $\beta_2 = 1/2p(1+\alpha) + 1/2$, *and* m_n *is the size of the nearest neighbor.*

Note that the convergence rate will depend on the dimension, the geometric noise exponent α and the choice of tuning parameter σ_n. Moreover, we observe that when $\lambda_n = n^{-\theta}$ with a constant θ and the size of nearest-neighbor equals to $n^{2/(p+2)}$, the polynomial convergence rate can be attained.

Proof of Theorem 2.3.2. For convenience of notation, we use $\|\cdot\|_n$ to denote the norm in the RKHS and omit g in the definition of the loss function, i.e., denote $L_n(f;g)$ as $L_n(f)$. We use c to denote a constant that is independent of n in the following proof. The M-learning algorithm estimates the decision function f as $\widehat{f}$ that minimizes (4.5), which can be rewritten as

$$L_n(f) \equiv \mathbf{P}_n\left[\frac{\int I(d(\widetilde{H},H)<\delta_n)\phi(-f(H)A\text{sign}(\widetilde{R}-R))g(|\widetilde{R}-R|)d\widetilde{\mathbf{P}}_n}{\int I(d(\widetilde{H},H)<\delta_n)d\widetilde{\mathbf{P}}_n}\right] + \lambda_n\|f\|_n^2.$$

Here, $\widetilde{\mathbf{P}}_n$ and $\widetilde{\mathbf{P}}$ to be used later refer to the measures with respect to an independent copy of random variables, $(\widetilde{R},\widetilde{A},\widetilde{H})$. We further define

$$Q_n(R,A,H;f) = \frac{\int I(d(\widetilde{H},H)<\delta_n)\phi(-f(H)A\text{sign}(\widetilde{R}-R))g(|\widetilde{R}-R|)d\widetilde{\mathbf{P}}_n}{\int I(d(\widetilde{H},H)<\delta_n)d\widetilde{\mathbf{P}}_n}$$

and

$$Q(R,A,H;f) = \widetilde{E}\left[\phi(-f(H)A\text{sign}(\widetilde{R}-R))g(|\widetilde{R}-R|)\Big|\widetilde{H}=H\right].$$

Clearly, $L_n(f) = \mathbf{P}_n Q_n(R,A,H;f) + \lambda_n\|f\|_n^2$.

Let $L_\phi(f) = E[Q(R, A, H; f)]$. From the general property of the weighted hinge-loss as shown in Theorem 3.2 of (Zhao et al., 2012), we have

$$V(f^*; g) - V(\widehat{f}; g) \leq c\left\{L_\phi(\widehat{f}) - L_\phi(f^*)\right\}.$$

Therefore, it is sufficient to obtain a bound for the right-hand side. First, since $L_{\phi n}(\widehat{f}) \leq L_{\phi n}(0)$, we obtain $\lambda_n \|\widehat{f}\|_n^2 \leq 1$. Let f_{0n} be the minimizer of $L_\phi(f) + \lambda_n \|f\|_n^2$ over $f \in \mathcal{H}_K$. Therefore,

$$\begin{aligned}
& L_\phi(\widehat{f}) - L_\phi(f^*) \\
\leq\ & E\left[Q(R, A, H; \widehat{f})\right] - E\left[Q(R, A, H; f_{0n})\right] + E\left[Q(R, A, H; f_{0n})\right] - V(f^*) \\
\leq\ & -(\mathbf{P}_n - \mathbf{P})\left[Q(R, A, H; \widehat{f}) - Q(R, A, H; f_{0n})\right] \\
& + \mathbf{P}_n\left[Q(R, A, H; \widehat{f})\right] - \mathbf{P}_n\left[Q(R, A, H; f_{0n})\right] \\
& + E\left[Q(R, A, H; f_{0n})\right] - V(f^*) \\
\leq\ & \sup_{f: \|f\|_n \leq \lambda_n^{-1/2}} \left|(\mathbf{P}_n - \mathbf{P}) Q(R, A, H; f)\right| \\
& + \mathbf{P}_n\left[Q(R, A, H; \widehat{f}) - Q_n(R, A, H; \widehat{f})\right] - \mathbf{P}_n\left[Q(R, A, H; f_{0n}) - Q_n(R, A, H; f_{0n})\right] \\
& + L_n(\widehat{f}) - \lambda_n \|\widehat{f}\|_n^2 - L_n(f_{0n}) \\
& + E\left[Q(R, A, H; f_{0n})\right] + \lambda_n \|f_{0n}\|_n^2 - V(f^*) \\
\leq\ & \sup_{f: \|f\|_n \leq \lambda_n^{-1/2}} \left|(\mathbf{P}_n - \mathbf{P}) Q(R, A, H; f)\right| && \text{(I)} \\
& + \sup_{R,A,H} \left|Q(R, A, H; \widehat{f}) - Q_n(R, A, H; \widehat{f})\right| && \text{(II)} \\
& + \sup_{R,A,H} \left|Q(R, A, H; f_{0n}) - Q_n(R, A, H; f_{0n})\right| && \text{(III)} \\
& + E\left[Q(R, A, H; f_{0n})\right] + \lambda_n \|f_{0n}\|_n^2 - V(f^*). && \text{(IV)}
\end{aligned}$$

We refer the terms in the right-hand side as (I), (II), (III) and (IV) in turn.

For term (I), we compute the bracket covering number of some finite balls in $\mathcal{H}_K$. First, from

Theorem 3.1 in (Steinwart and Scovel, 2007), the entropy number for the unit ball in $\mathcal{H}_K$, denoted by $\mathcal{O}_n$, satisfies

$$\log \mathcal{N}(\epsilon, \mathcal{O}_n, \|\cdot\|_\infty) \leq c\sigma_n^{-(1-p/4)d}\epsilon^{-p}$$

for a constant c depending on p and d, so it yields

$$\log \mathcal{N}_{[]}(\epsilon, \mathcal{O}_n, \|\cdot\|_{L^2(P)}) \leq c\sigma_n^{-(1-p/4)d}\epsilon^{-p}.$$

Thus, we obtain

$$\log \mathcal{N}_{[]}(\epsilon, \left\{f : f \in \mathcal{H}_{\sigma_n}, \|f\|_n \leq \lambda_n^{-1/2})\right\}, \|\cdot\|_{L^2(P)}) \leq c\sigma_n^{-(1-p/4)d}\epsilon^{-p}(1/\lambda_n)^{p/2}.$$

Note that $Q(R, A, H; f)$ is Lipschitz continuous with respect to f in the sense that

$$\left|Q(R, A, H; f_1) - Q(R, A, H; f_2)\right| \leq c|f_1(H) - f_2(H)|,$$

where c is a constant bounding $g(|R - \widetilde{R}|)$. Therefore, we obtain

$$\log \mathcal{N}_{[]}(\epsilon, \left\{Q(R, A, H; f) : \|f\|_n \leq \lambda_n^{-1/2}\right\}, \|\cdot\|_{L^2(P)}) \leq c\sigma_n^{-(1-p/4)d}\epsilon^{-p}/\lambda_n^{p/2}.$$

According to Theorem 2.14.2 in (Van Der Vaart and Wellner, 1996), we obtain that term (I) is bounded by

$$O_p(1)\left\{n^{-1/2}\int_0^c \sqrt{1 + \log \mathcal{N}_{[]}(\epsilon, \left\{Q(R, A, H; f) : \|f\|_n \leq \lambda_n^{-1/2}\right\}, \|\cdot\|_{L^2(P)})}d\epsilon\right\}$$
$$= O_p(1)n^{-1/2}\sigma_n^{-(1/2-p/8)d}/\lambda_n^{p/4}.$$

For term (II), since $\|\widehat{f}\|_n \leq \lambda_n^{-1/2}$, Theorem 4.48 in (Steinwart and Christmann, 2008), implies that $\widehat{f}$ is differentiable with derivative bounded by $c\sigma_n^{-1}\|\widehat{f}\|_n = c\sigma_n^{-1/2}\lambda_n^{-1/2}$. Using the uniform convergence rate result for nearest-neighbor estimators (Devroye et al., 2013; Jiang, 2017) and assumptions (A.1)-(A.3), we obtain that term (II) is bounded by

$$O_p(1)(\sigma_n\lambda_n)^{-1/2}\left[\left(\frac{m_n}{n}\right)^{1/p} + \sqrt{\frac{p\log n}{m_n}}\right].$$

The same bound holds for term (III). Finally, the last term is the approximation error as defined in (Steinwart and Christmann, 2008) but with a different definition of the loss function as $Q(R, A, H; f)$. We can follow exactly the same argument in Theorem 2.7 of (Steinwart and Christmann, 2008) to obtain its upper bound as $c(\sigma_n^{-p}\lambda_n + \sigma_n^{\alpha p})$ for any positive α.

In conclusion, we have shown

$$L_\phi(\widehat{f}) - L_\phi(f^*)$$
$$\leq O_p(1)\left[\frac{n^{-1/2}}{\sigma_n^{(1/2-p/8)d}\lambda_n^{p/4}} + (\sigma_n\lambda_n)^{-1/2}\left\{\left(\frac{m_n}{n}\right)^{1/p} + \sqrt{\frac{\log n}{m_n}}\right\} + \sigma_n^{-p}\lambda_n + \sigma_n^{\alpha p}\right].$$

By choosing $\sigma_n = \lambda_n^{1/p(1+\alpha)}$, we obtain the result in Theorem 2.3.2.

As a remark, the tail probability, $P(|V(f^*; g) - V(\hat{f}; g)| \geq t)$ where $t > 0$, can also be obtained under similar arguments. Theorem 2.3.2 provides a stochastic bound for term (I). One can obtain the bound of the tail probability for this term using the tail bound for empirical processes (Chapter 2.14, (Van Der Vaart and Wellner, 1996)). Then the tail probability, $P(|V(f^*; g) - V(\hat{f}; g)| \geq t)$, will follow. □

2.4 Simulation Studies

We conducted extensive simulation studies to compare M-learning with Q-learning and single-stage AOL as improved O-learning (Liu et al., 2018). Data were simulated under an observational study design where treatment assignment depends on pre-treatment variables H. Simulation settings and analyses we considered include: (1) No unmeasured confounder and the propensity score model given H is correctly specified in the analyses; (2) No unmeasured confounder but the propensity score model is misspecified; and (3) Unmeasured confounders are present and some components of H are not observed and not included in the analyses.

In these simulations, one-to-one matching with replacement was used and features were matched using shortest Euclidean distance function (one nearest-neighbor). The tuning parameters for AOL and M-learning (including choice of kernel as linear or Gaussian, inverse radius, and cost C) were selected by three-fold cross validation. The value function corresponding to the estimated optimal rule was computed on a large independent testing set with a sample size of $10,000$ using empirical average. Q-learning was fit with a linear model including feature variables and their interaction with treatment as covariates. We varied sample size of training data from 100 to $1,000$ and repeated the simulations 100 times.

We first considered continuous responses in two settings:

$$S_1 : R = 2H_3 - H_4 + A(H_1 - H_2) + 6\text{sign}(H_1) + N(0,1)$$

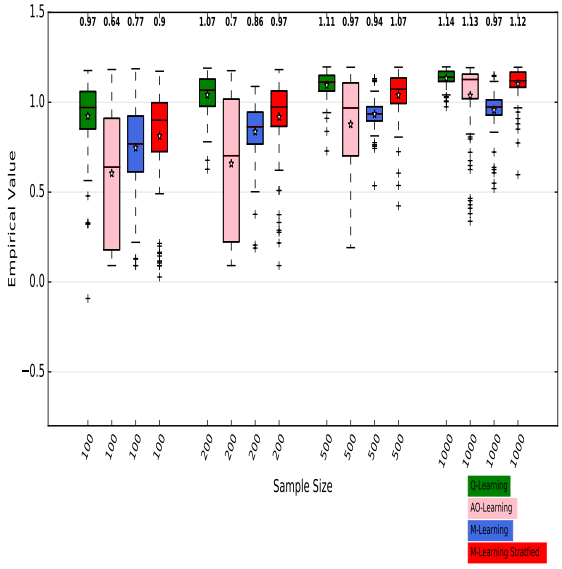
and

$$S_2 : R = 1 + 2H_1 + H_2 + 0.5H_3 + A(H_2 + H_1^2 - 1) + 6\text{sign}(H_1) + N(0,1).$$

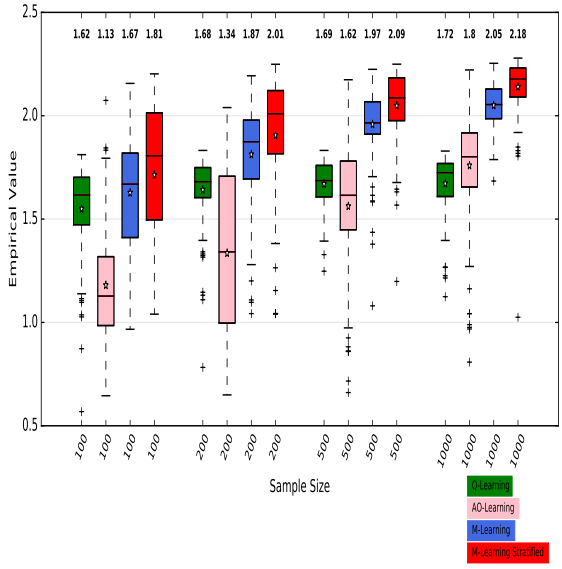
Uncorrelated feature variables H_k with standard normal distributions were simulated. Since heterogeneity and clustering effects are observed in the real-world patient population (e.g., Figure 2.5 of NYPH EHRs in Appendix A), we considered the distribution of reward outcomes to be clustered in strata depending on the first feature variable H_1. The true optimal treatment decision boundary is linear in setting S_1, and nonlinear in setting S_2. The true optimal value is 1.20 in S_1 and 2.29 in S_2. In the continuous response scenario, $g(x) = x$ was used for M-learning. In setting S_1 and S_2, M-learning and doubly robust M-learning by stratifying on prognostic scores (referred to as "M-learning Stratified" in Figure 2.1 and 2.2) were considered. For the latter, prognostic scores were obtained using random forest. Prognostic factors used in the matching step were created by dichotomizing the prognostic scores based on the median split.

In the first set of simulations, distribution of A depends on H and no unmeasured confounder is present. Clinical response outcomes were simulated under setting S_1 and S_2, and the true propensity model was specified as $P(A = 1|H) = \text{expit}(1 + 2H_1 + H_2)$. In this case, H_1 and H_2 are observed confounders. The propensity scores were estimated through a logistic regression model with treatment as binary outcome and features H_1, H_2 as linear predictors. On average, 64% of subjects received an active treatment and 36% received a control treatment. Simulation results are presented in the top panel of Figure 2.1. For setting S_1, Q-learning has the best performance since the linear function is the true optimal treatment separation boundary. Doubly robust M-learning performs similarly as Q-learning with larger sample size. It is clear that doubly robust M-learning improves efficiency. For S_2 with a nonlinear boundary, both M-learning and doubly robust M-learning achieve a higher empirical value than AOL and Q-learning. In this case Q-learning and AOL lose efficiency because they do not capture the information in prognostic scores, even though the propensity scores were consistently estimated.

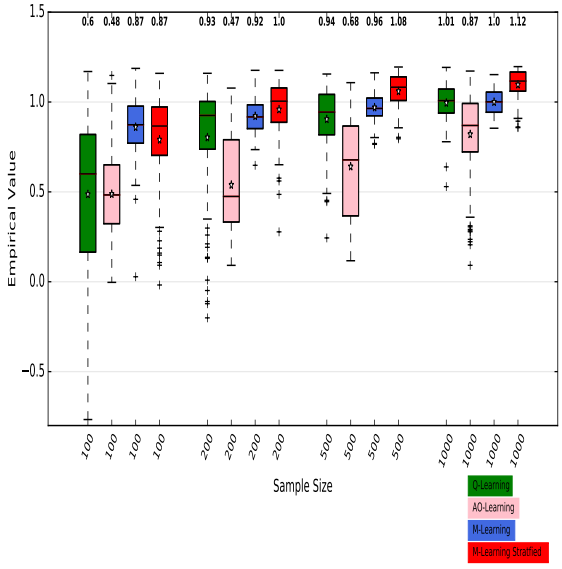
In the second set of simulations, the true propensity score model was specified as $P(A = 1|H) = \text{expit}(1 + \exp(H_2))$. The propensity scores were estimated through a logistic regression model with linear predictors, and thus the model was misspecified. On average, 88% of subjects received one treatment and 12% received the other. Simulation results are presented in the bottom panel of Figure 2.1. In both setting S_1 and S_2, the results suggest that M-learning is more robust to misspecified propensity model compared to Q-learning and O-learning. The best performance is achieved by the doubly robust M-learning, where the estimated value function is very close to the true optimal value with a large sample size. Matching using prognostic scores in doubly robust M-learning has protected against deteriorated performance when the propensity score model is misspecified.



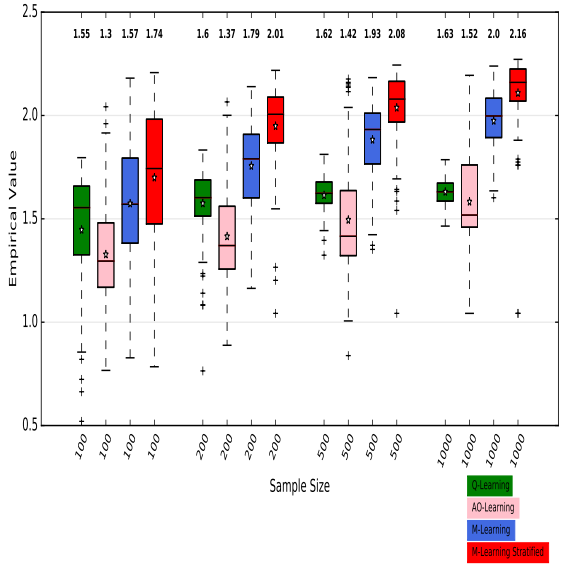
(a) Setting S1, propensity score model correctly specified



(b) Setting S2, propensity score model correctly specified



(c) Setting S1, propensity score model misspecified



(d) Setting S2, propensity score model misspecified

Figure 2.1: Value comparison of four methods with propensity scores correctly specified (top panel) and misspecified (bottom panel). The numbers at the top of each subfigure are mean values.

In the third set of simulations, we considered presence of unmeasured confounders. The clinical outcomes were simulated as

$$S_3 : R = 2H_3 - H_4 + A(H_1 - H_2 + X) + 6\text{sign}(H_1) + N(0,1)$$

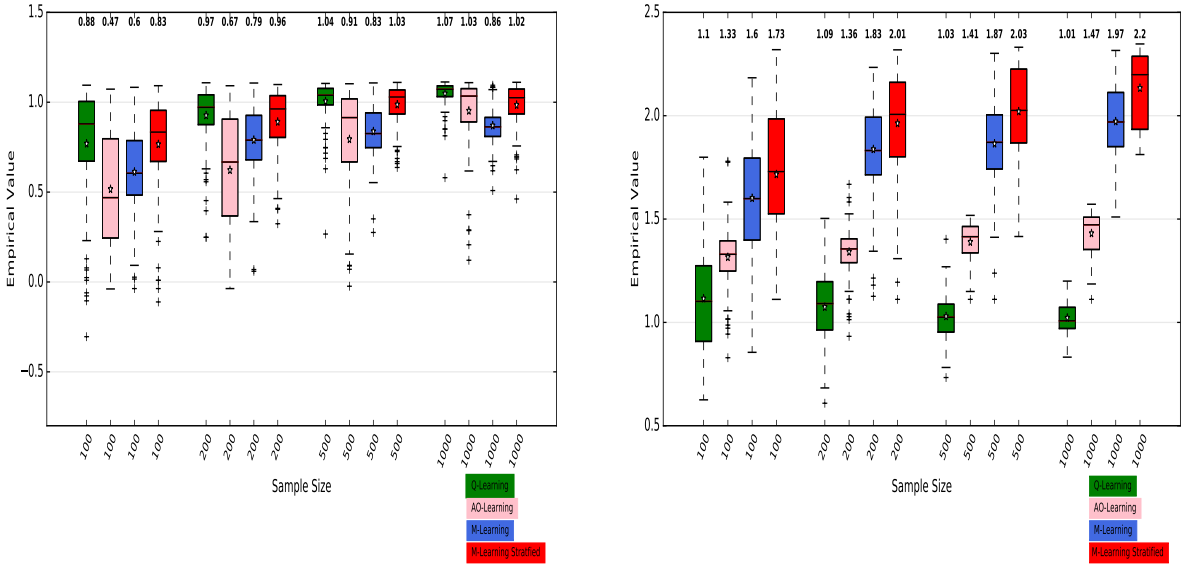
and

$$S_4 : R = 1 + 2H_1 + H_2 + 0.5H_3 + A(H_2 + H_1^2 + X - 1) + 6\text{sign}(H_1) + N(0,1)$$

where $P(A = 1|H, X) = \text{expit}(1 + R^{(-1)} - R^{(1)} + 2X + H_1)$ and X is an unmeasured confounder (not included in any analysis in any method) and $R^{(-1)}, R^{(1)}$ are potential outcomes under each treatment.

After introducing unmeasured confounding, the true optimal value function is 1.37 in S_3 and 2.61 in S_4. From Figure 2.2, we see that in S_3 with a linear decision boundary, Q-learning performs the best. Doubly robust M-learning has a higher mean value than M-learning. Matching-based methods have an advantage over AOL. Specifically, the value function of ITR estimated by AOL has a large variability, especially when the sample size is small. In S_4 with nonlinear decision boundary, two M-learnings much outperform AOL and Q-learning. In this case, the unmeasured confounder has a greater impact on AOL and Q-learning than M-learning.

We also examine M-learning with ordinal outcomes and report results in Appendix A.1. For linear decision boundary, since ordinal outcomes were generated by discretizing a continuous outcome, M-learning does not give an advantage over Q-learning and AOL. For nonlinear boundary, M-learning using matching function $g(x) = 1$ and $g(x) = x$ both achieves a higher value than Q-learning and AOL.



(a) Setting S3: unmeasured confounders present

(b) Setting S4: unmeasured confounders present

Figure 2.2: Value comparison of four methods in the presence of unmeasured confounders. The numbers at the top of each subfigure are mean values.

2.5 Applications

We apply various methods to a large clinical data warehouse (CDW) at New York Presbyterian Hospital (NYPH). NYPH CDW is one of the earliest pioneer CDWs in the United States developed 25 years ago, long before the wide adoption of EHRs and informatics methods. The database encompasses about 4.5 million patients in the New York City population, making it a useful data source for research and supports new research initiatives including eMERGE (Gottesman et al., 2013) and precision medicine initiative. The details of the informatics technology of NYPH CDW is described in Figure 2.3.

Our research goal is to optimize treatment sequence for T2D patients based on their person-specific characteristics. Current treatment guideline recommends metformin (MET) as the first line treatment for T2D patients (Diabetes Control and Complications Trial Research Group, 1993).

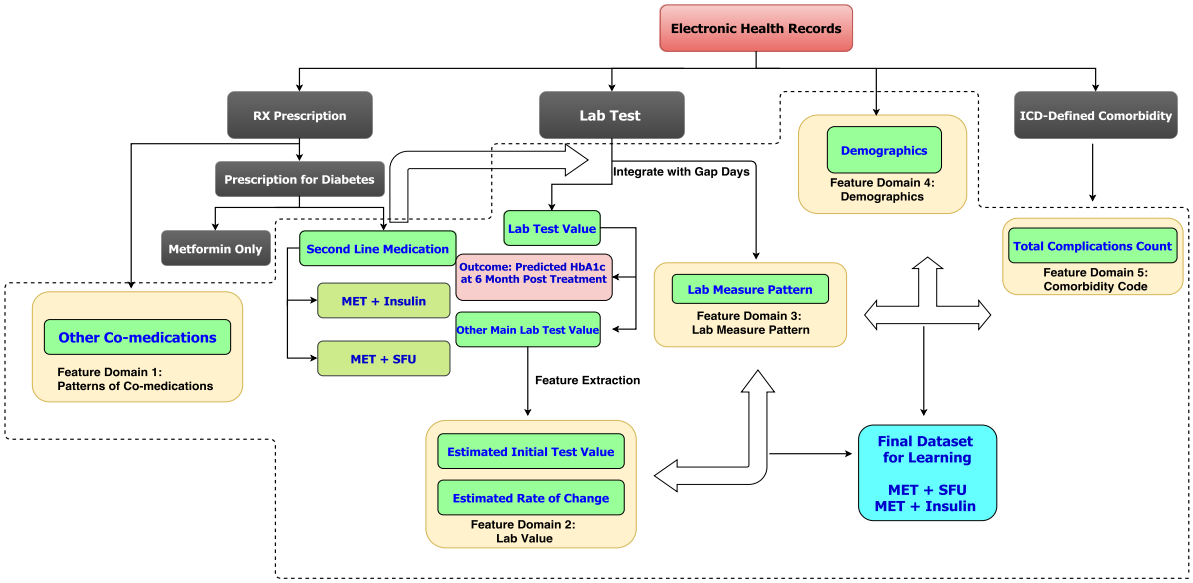


Figure 2.3: Flowchart of EHR data processing procedures

Literature reveals barriers of timely insulin initiation in clinical practice when patients do not achieve adequate glycemic control by using metformin alone, and the optimal sequence of treatments for insulin therapy versus second-line oral hypoglycemic agents (OHA) largely remains unknown (American Diabetes Association, 2014). In this work, we aim to estimate the optimal second-line treatment for T2D patients who received MET as the first-line treatment using real-world EHRs. Targeting the second-line treatments (metformin + insulin versus metformin + SFU, where SFU refers to oral agent sulfonylureas that includes glyburide and glipizide) partially reduces confounding by indication, where treatment uncertainty is present in real-world practice.

We excluded subjects with extreme baseline HbA1c values (greater than 10%), and used a new-user cohort design (Ray, 2003). Such design is often used in other studies of EHRs to properly capture time-varying confounding and early treatment responses. Specifically, the study design is illustrated in Figure 2.4. Subjects who started a second-line treatment (new users) are anchored at the treatment initiation (index date), and information before and after index date will be analyzed.

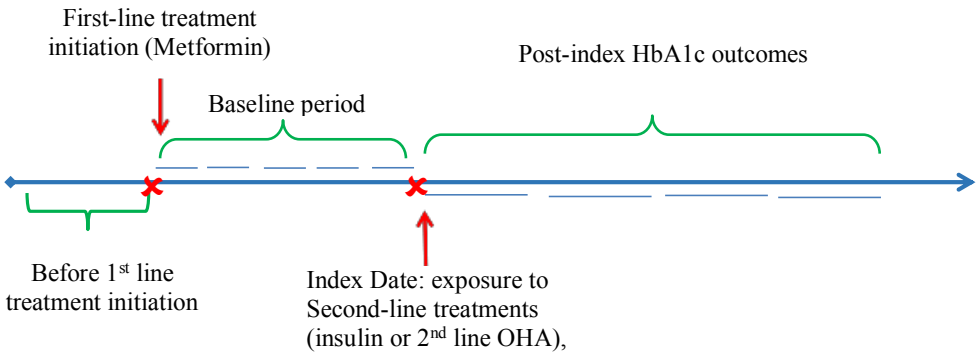


Figure 2.4: T2D EHR Study Design

Subjects were included in the analyses if they had MET as the first-line treatment, had insulin or SFU as the second-line treatment, and had at least one observation post index date. The median baseline period was around one year and the median follow up time post second-line treatment was about 18 months.

In the following sections 2.5.1 and 2.5.2, we describe details of patient records extraction and feature extraction.

2.5.1 T2D Patient EHRs

Quality of the information collected in the EHRs is essential for learning treatment strategies from EHRs. The current design of NYPH CDW uses the cutting-edge informatics technologies and utilizes a quality control process that helps accuracy of feature extraction/curation (e.g., all data have been normalized using the medical entities dictionary to facilitate semantic information retrieval). Several studies were launched to investigate data quality including completeness, correctness, concordance, plausibility, and currency (Weiskopf and Weng, 2013; Fort et al., 2014), where recent EHRs in NYPH CDW were deemed to possess better completeness and accuracy (Weiskopf and Weng, 2013; Hripcsak et al., 2016). The main domains of the information in EHRs include demographics, in-patient and out-patient medication prescriptions, ICD-9 diagnosis codes,

and laboratory tests, which were longitudinally documented in the CDW. In order to be more compatible with other observational databases as part of the Observational Health Data Sciences and Informatics (OHDSI, https://ohdsi.org/), clinical data warehouse at NYPH has been recently converted to the OMOP Common Data Model version 5.0 to support real-world evidence generation using scalable observational data. The quality of these EHR data was found to be suitable for studying treatment pathways of common diseases including T2D (Hripcsak et al., 2016).

Patients aged 18 or older who had at least one T2D diagnosis between 1/1/2008 to 12/31/2012 and were prescribed with metformin were included in the analysis. There were 1,279 patients who had a second-line treatment in the CDW EHRs during the 5 year window. Medication prescriptions, laboratory tests, demographics, and ICD-9 diagnosis codes from over $50,000$ patients were extracted. The details of data pre-processing procedures and intermediate datasets are in the flowchart in Figure 2.3. Figure 2.5 displays a heatmap of representative patients with 17 feature variables extracted from the CDW EHRs, among which ICD diagnosis and co-medication counts were derived from over $8,000$ unique variables. To explore patient heterogeneity, features were standardized and a hierarchical clustering analysis with Euclidean distance clustered patients into 3 groups: (1) a moderately ill group with a slow rate of change in glycemic measures (HbA1c, glucose) and a regular, frequent documentation pattern; (2) a moderately ill group with slow progression but a less frequent measurement pattern, less co-morbidity and non-diabetic medications; and (3) a fast progression group with less measurements. The cluster membership was included as a feature variable to construct the ITR.

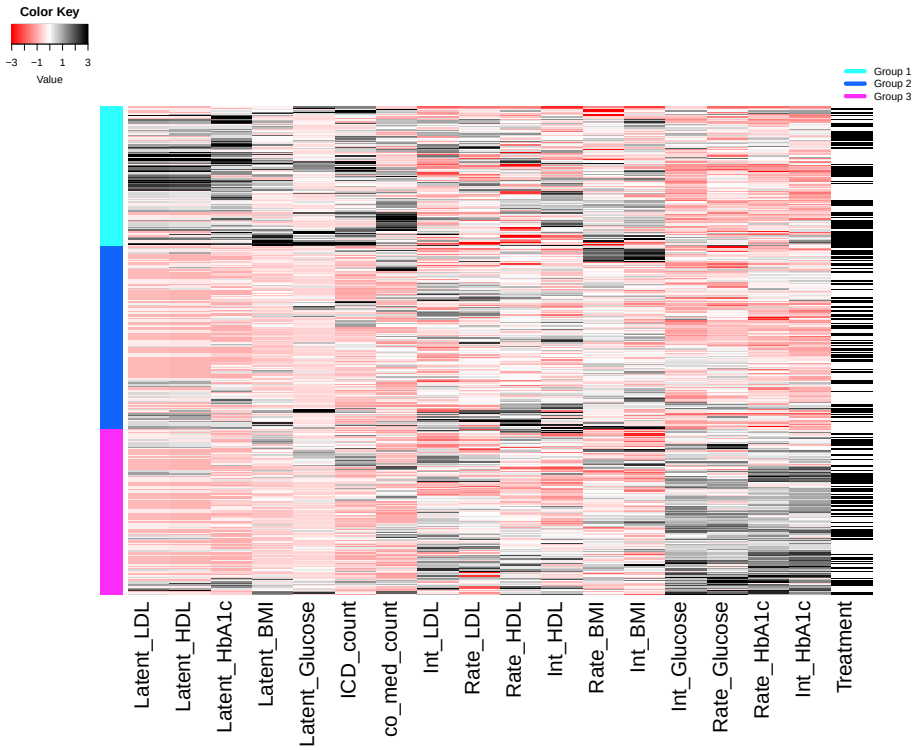


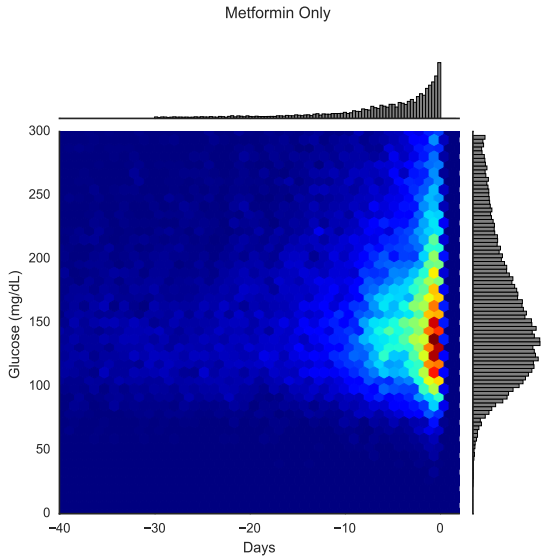
Figure 2.5: Heatmap of 17 extracted feature variables from EHRs of representative T2D patients in NYPH CDW.

2.5.2 Handling Challenges of the Analyses of EHRs

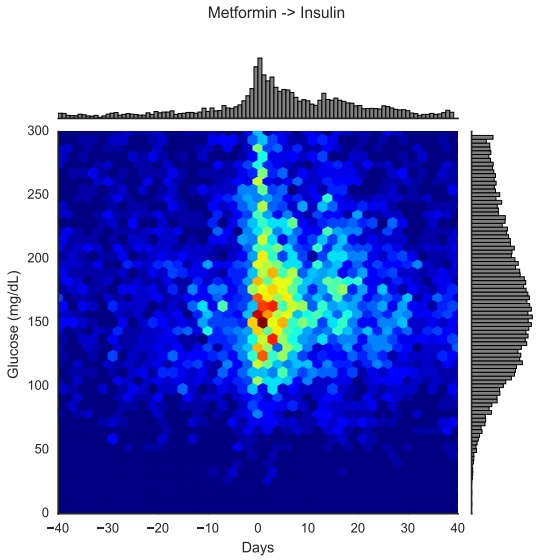
Three major challenges need to be addressed when conducting clinical research using EHR data (Haneuse, 2016). In our context, these issues are: confounding bias due to factors associated with treatment selection (i.e., insulin versus SFU) and outcome; selection bias due to missing post second-line treatment outcome (some patients had no follow up HbA1c test records and thus were not included in the analyses), and presence of missing feature variables. EHRs provide information that not only reflects patient's health status but also the healthcare process, shedding lights on how data are recorded (Hripcsak and Albers, 2013).

To extract such information, descriptive analysis in Figure 2.6 indicates that distinct measurement patterns of glucose tests are present for different treatments. For example, the measurement process of glucose tests shows a high discriminant power (y-axis, histogram on the side) between patients who received insulin as a second-line treatment (MET+insulin) and those who received glyburide or glipizide (MET+SFU); this information is thus useful for creating propensity scores and matching. The temporal process of measurement time intensity patterns is also highly discriminant (x-axis, histogram on the top) and useful for computing treatment propensities. Figure A.1 shows a similar discriminant power for HbA1c tests.

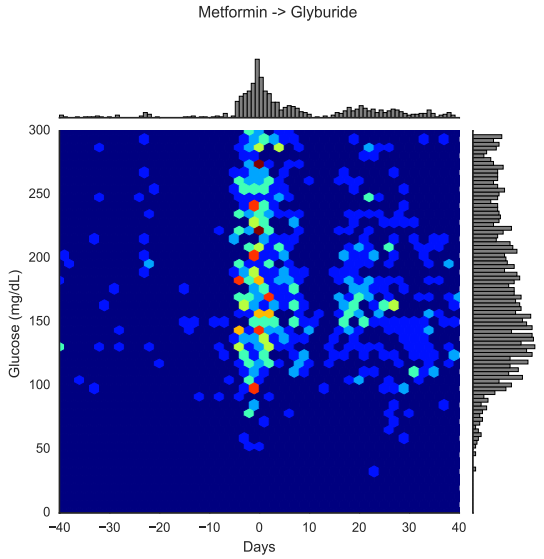
We constructed feature variables of lab measurement patterns by discretizing a two-dimensional feature space of lab test values versus gap time between two consecutive measurements in Figure 2.6, A.1. Indicators of subgroup membership based on low, medium, and high quartiles were created. For example, the first subgroup indicates subject with a "low glucose value and short measurement gap". In addition, we constructed features informative of treatment response such as initial lab test values and rate of change of lab values before treatment initiation from longitudinal records of four



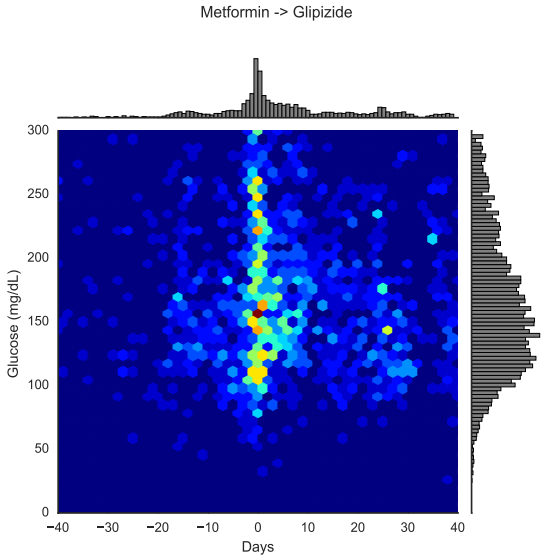
(a) Metformin Only



(b) Metformin + Insulin



(c) Metformin + Glyburide



(d) Metformin + Glipizide

Figure 2.6: Glucose values and measurement intensity (Time 0: time at first line treatment (MET) prescription)

major tests: HbA1c, glucose, high density lipoprotein (HDL) and low density lipoprotein (LDL). The initial test values and rates of change were estimated from a linear mixed effects model. These features are also potential prognostic factors associated with post index date HbA1c levels, and are used to create prognostic scores.

We constructed patterns of laboratory measurements to handle challenges in the analyses of EHRs (e.g., confounding bias and selection bias). Extracted features encompass information from five domains (Figure 2.3): demographics, medication prescription, ICD diagnosis codes, laboratory test values, and lab test measurement patterns. Propensity scores were estimated using two distinct logistic regression models for lab measurement pattern features and demographics covariates. The matching step in M-learning was performed using extracted features from lab test values and rates of change (HbA1c, glucose, HDL, LDL, BMI), ICD counts, and two propensity scores. In addition, to improve efficiency and perform doubly robust matching, we also included a prognostic score estimated from a linear regression model in the matching step. Mahalanobis distance was the matching similarity measure and one nearest-neighbor was used to select matched pairs. To address selection bias in missing post-treatment outcomes, we used the IPW method and constructed a logistic regression model predicting whether a subject had any post treatment lab measure to compute the weights. To handle incompleteness in features, imputation with chained equations was used (Buuren and Groothuis-Oudshoorn, 2011).

2.5.3 Analysis Results

Our final EHR data for learning optimal ITR consists of 740 patients, among whom 292 (39%) received insulin as the second-line treatment while 448 (61%) received SFU. The outcome is the HbA1c level (%) at 6 month post second-line treatment initiation estimated from a linear mixed

effect model with subject-specific random intercepts and random slopes. Feature variables for learning optimal ITR include initial lab test values (HbA1c, glucose, HDL, LDL, BMI) and rate of change of measurements before index date, demographic variables, the cluster membership estimated from a subset of features (Figure 2.5), counts of other non-glycemic medications and counts of positive ICD diagnosis codes. Two-fold cross validation was used to estimate the value function of fitted ITRs.

We divided our cohort to two groups according to the initial HbA1c level (high baseline HbA1c group: $>= 8.5$ and low baseline HbA1c group: < 8.5) and analyzed the groups separately to further reduce patient heterogeneity. We compared the cross-validated value function of doubly robust M-learning to non-personalized universal rules, Q-learning, and AOL. In the rest of this section, we refer doubly robust M-learning as M-learning and AOL as O-learning for simplicity. The results are displayed in Figure 2.7 and Table 2.1. In the low baseline group, there were 380 patients in total (240 received SFU, 140 received insulin). For universal rules, the IPW-adjusted mean HbA1c level is 7.99 for those treated by SFU and is 8.05 for insulin. M-learning achieves the best glycemic control among all methods (lowest post-treatment HbA1c at 6 month) with a median and mean of 7.85 that is much lower than both universal rules. Q-learning does not provide much improvement compared to universal rules and its estimated post-treatment HbA1c is slightly smaller than assigning SFU to all. In the high baseline group, there were 152 patients who received insulin and 208 received SFU. The universal rules for HbA1c level in SFU group is 8.90 and in insulin group is 9.21. O-learning and M-learning have very similar performance and both reduce the average post-treatment HbA1c level to 8.57, again much lower than universal rules.

By examining M-learning in all patients using a linear kernel in the low baseline group, we identified several features that are most informative in determining the optimal treatment: pre-

treatment rate of change of BMI, initial value of glucose and LDL at the index date, co-medication count, patient cluster membership and race. These feature variables can be considered by healthcare practitioner when recommending second-line treatment for T2D patients. There were 263 (69%) of the 380 patients predicted to have "MET + SFU" as the optimal choice and 117 (31%) with "MET + Insulin" as the optimal choice. Of the 240 patients who were prescribed SFU as the second-line treatment, majority of times (66%) medication was also the predicted optimal treatment in terms of lowering HbA1c level. In contrast, among the 140 patients who were prescribed insulin, only 36 (26%) were optimal. In the high baseline group, the important features we identified are initial value of HDL, age, sex and patient cluster membership. 294 (82%) of the 360 patients were recommended to "MET + SFU". Of the 208 patients who were prescribed SFU, 168 (81%) also had as the predicted optimal treatment. Among the 152 patients who received insulin treatment, only 26 (17%) were optimal. These results seem to suggest that some patients who received insulin as the second-line treatment might be better treated with SFU.

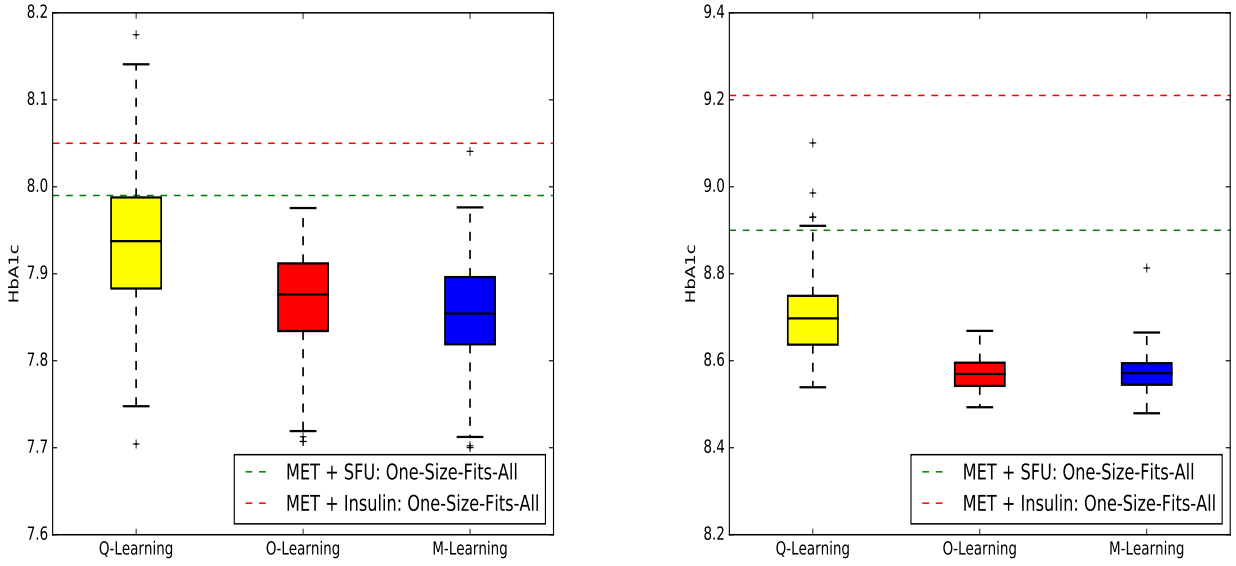
However, (Bianchi and Del Prato, 2011) suggested that tight glycemic control need to be studied carefully in different group of T2D patients to determine the balance of its negative and positive effect and treatment personalization should be recommended considering multiple factors such as risk of complications (e.g. cardiovascular events). Given a low rate of insulin predicted to be optimal among patients who were treated with insulin, we explored whether insulin could be prescribed based on other considerations such as risk of complications in addition to achieving glycemic control. We estimated the optimal ITR that reduces major complications of T2D measured by three ICD diagnosis counts including essential hypertension, hyperlipidemia and hypercholesterolemia as ordinal outcomes (0, 1, 2, 3). M-learning was implemented with $g(x) = x$. The results are displayed in Figure 2.8 and Table 2.2. In the low baseline group, O-learning is moderately better than

M-learning with an average count of 0.72. Based on M-learning, SFU was predicted to be optimal for 274 (72%) patients. Among patients who indeed received SFU, 175 (73%) were predicted to be optimal with regarding to reducing complications while 41 (29%) of the patients who received insulin were predicted to be optimal. In the high baseline group, M-learning performs the best with an average value of 0.84. Further investigation shows that insulin was predicted to be the optimal choice for 234 (65%) patients. In this group, among 152 patients who indeed received insulin, 106 (70%) were predicted to be optimal with regard to reducing complications, while only 80 (38%) of the patients who received SFU were predicted to be optimal.

In conclusion, the optimal ITRs outperform universal rules in all groups for both outcomes. M-learning performs better than Q-learning in all cases and better than O-learning in most cases. In addition, the proportion of patients treated by insulin and with insulin predicted to be optimal is higher when considering reducing complications as the outcome as compared to controlling for HbA1c (from 17% to 70% in the high baseline group). This result suggests that the rationale to prescribe insulin might be also based on concerns of complications especially when the baseline HbA1c is high (greater than 8.5%).

2.6 Discussion

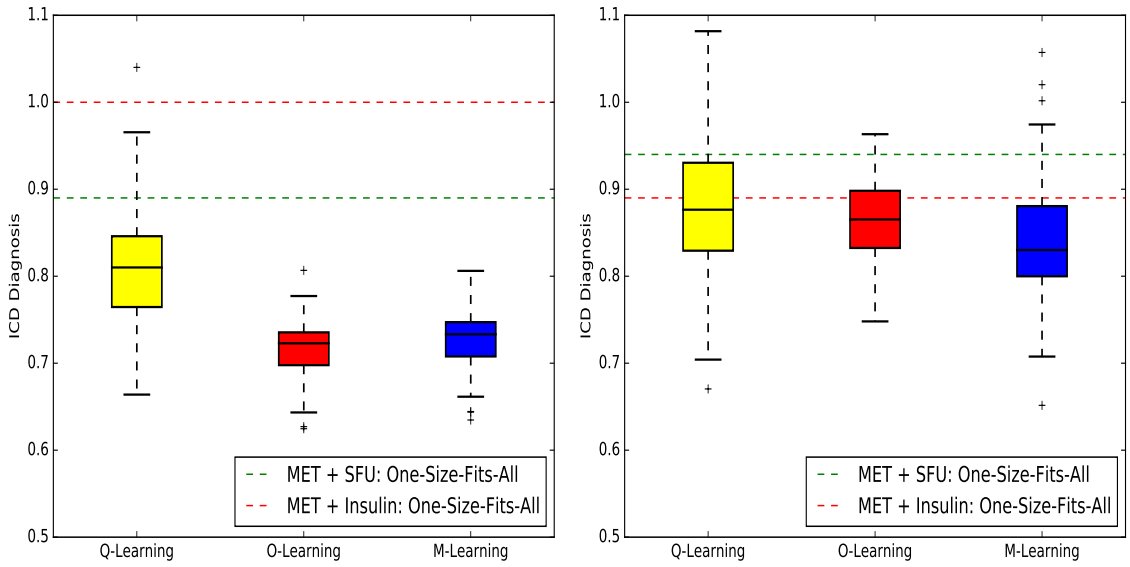
We have proposed a machine learning approach based on matching, M-learning, to estimate the optimal ITR from observational data. We show that M-learning is a general approach that includes O-learning and some of its derivatives as special case and it satisfies Fisher consistency. A general matching function is proposed to analyze continuous or discrete outcomes where in some cases the objective function maximizes a certain function of AUC. The choice of $g(\cdot)$ function provides



(a) Low Baseline Group

(b) High Baseline Group

Figure 2.7: Empirical value function of HbA1c in EHR data with 100 2-fold cross-validations (a low value is desirable)



(a) Low Baseline Group

(b) High Baseline Group

Figure 2.8: Empirical value function of ICD diagnosis count in EHR data with 100 2-fold cross-validations (a low value is desirable)

Table 2.1: Cross-validated Empirical Value Function for HbA1c

High Baseline Group Universal rules: MET + SFU: 8.90, MET + Insulin: 9.21		
ITR Method	Mean (Std)	Median (Q1, Q3)
Q-Learning	8.72 (0.124)	8.70 (8.64, 8.75)
O-learning	8.57 (0.038)	8.57 (8.54, 8.60)
M-Learning	8.57 (0.045)	8.57 (8.55, 8.59)
Low Baseline Group **Universal rules: MET + SFU: 7.99, MET + Insulin: 8.05**		
ITR Method	Mean (Std)	Median (Q1, Q3)
Q-Learning	7.94 (0.083)	7.94 (7.88, 7.99)
O-learning	7.87 (0.061)	7.88 (7.83, 7.91)
M-Learning	7.85 (0.068)	7.85 (7.82, 7.90)

Table 2.2: Cross-validated Empirical Value Function for the Number of Major Complications

High Baseline Group Universal rules: MET + SFU: 0.94, MET + Insulin: 0.89		
ITR Method	Mean (Std)	Median (Q1, Q3)
Q-Learning	0.88 (0.078)	0.88 (0.83, 0.93)
O-Learning	0.86 (0.050)	0.87 (0.83, 0.90)
M-Learning	0.84 (0.068)	0.83 (0.80, 0.88)
Low Baseline Group **Universal rules: MET + SFU: 0.89, MET + Insulin: 1.00**		
ITR Method	Mean (Std)	Median (Q1, Q3)
Q-Learning	0.81 (0.063)	0.81 (0.76, 0.85)
O-Learning	0.72 (0.033)	0.72 (0.70, 0.74)
M-Learning	0.73 (0.032)	0.73 (0.71, 0.75)

a flexible tool to weight outcome measures: $g(x) = 1$ gives the most robust estimation which only concerns with the ranking of outcomes; while other robust choices can prevent sensitivity to outliers of R_i's. Moreover, multivariate outcomes can be incorporated in the M-learning framework by creating suitable g function. The matching function $g(x)$ can be selected from a pool of non-decreasing functions to estimate the optimal ITRs in a data-driven way, which may lead to a better post treatment response.

M-learning has a few advantages over O-learning or other IPW-based methods. It does not rely on the validity of propensity score models and no inverse weighting is involved. Thus instability can be avoided when there are extremely small weights. The choice of $\mathcal{M}_i$ in M-learning is flexible and can include a large suite of matching tools including nearest neighbor, metrics defined on a dimension-reduced space determined by propensity scores or prognostic scores, yielding double robustness. For example, methods based on greedy matching or optimal matching algorithm are available to be implemented in M-learning. Different calipers can also be specified for individual subject and hence allow more "personalization". This strategy will introduce more flexibility but at the price of some computational complexity.

The choice of matching variables is important in M-learning. The performance of M-learning may be affected by the presence of high-dimensional features in the matching step. We suggest a dimension reduction approach to match on a lower dimensional space consisting of propensity score, prognostic score, and/or cluster membership of patients. We also included some key covariates as part of the matching criteria. A more general practical guide during the matching step is: first, choose major confounders according to domain knowledge or preliminary studies to achieve covariates balance; second, construct several propensity scores to reduce the dimensionality of the space of matching covariates; and third, include prognostic scores in order to improve robustness

and efficiency. In the EHR analysis here, we considered this general guideline and constructed domain-wise propensity scores as well as prognostic scores, and matching was performed based on these scores. Other variable selection techniques can be considered, for example, to estimate propensity and prognostic scores by penalized regression.

Single-stage M-learning can be generalized to multi-stage setting by changing the value function $V(\mathcal{D})$ to a corresponding matching-based value function involving multiple stages and applying the backward learning methods (Liu et al., 2018). In each stage, M-learning will have the flexibility to choose different matching function and matched features. Furthermore, an extension to handle efficacy and safety outcomes (e.g., glycemic control and risk of complications) simultaneously when learning ITR is desirable. Here we only considered choosing between two treatment options. M-learning is ready to be generalized to more than two treatments by, for example, adopting one-versus-one or one-versus-all strategies for multicategory learning (Allwein et al., 2001). Lastly, our analyses were restricted to EHRs from those who had at least one second-line T2D treatment documented at a single academic medical center. It would be of interest to examine the performance of our methods on other EHR databases.

Chapter 3

Domain Adaption Transfer Learning

3.1 Overview

Personalized medicine based on individualized treatment rules or recommendations (ITRs) has been proposed as an alternative to population based strategy (Hamburg and Collins, 2010; Collins and Varmus, 2015). Personalized treatment decision making becomes closer to reality than before, due to recent advances in modern technologies that produce a large amount of patient-specific data. In particular, patients' electronic health records (EHRs), which contain medical history, laboratory measures, and disease diagnosis for a large number of patients over years, provide rich information about each patient's comorbiditiy, treatment history and outcomes in a real-world setting. How to incorporate such real-world evidence to learn ITRs remains an important and challenging research question in the modern era of personalized medicine.

Recent development of statistical and machine learning methods for optimizing ITRs aims to assist healthcare providers to prescribe the right therapy to the right patient at the point of care (Chakraborty and Moodie, 2013; Collins and Varmus, 2015). Patients can have improved clin-

ical benefits and adherence to the treatment with the prescribed optimal individualized therapy (Chakraborty and Moodie, 2013). For healthcare providers and industry, it is of interests to revolutionize personalized treatment by estimating individualized treatment rules (ITRs). These methods include regression-based approaches that estimate the conditional mean of treatment outcomes or their contrasts given treatments and patient's feature variables, to name a few, Q-learning (Qian and Murphy, 2011; Chakraborty and Moodie, 2013), A-learning (Murphy, 2003) and G-computation (Lavori and Dawson, 2004; Moodie et al., 2007). Alternatively, one can directly optimize a value function or its equivalence to learn optimal ITRs. These methods include outcome-weighted learning (O-learning, (Zhao et al., 2012)), a doubly-robust version of O-learning (Liu et al., 2018) and more recently, matching-based M-learning (Wu et al., 2019).

Many existing methods are developed in the context of randomized controlled trials (RCTs), for which the estimated ITRs are known to be consistent. However, although RCTs have high internal validity, they are usually conducted under specific inclusion/exclusion criteria, which potentially limits the generalizability of the resulting ITRs to a broader real-world patient population. In fact, some findings of treatment efficacy in clinical trials may not be directly translated to a general patient population in the real world clinical setting with the same condition (Haynes, 1999). On the other hand, real-world data such as EHRs document medical practices in a real-world setting. Therefore, in order to improve the generalizability of the ITRs, it is desirable to incorporate the EHR evidence when learning the ITRs in clinical trials. However, due to the presence of unmeasured confounding in the EHR data and the potential differences between the trial and EHR patients, how to effectively incorporate the EHR data poses a challenging issue.

In this chapter, we propose a novel method to improve learning ITRs from RCTs borrowing evidence from EHRs. First, we pre-train two "super" feature mappings from the EHR data: one is

a probability feature that estimates the propensity of physician's choice of a given treatment being the same as the optimal treatment for an EHR patient; the other is a benefit feature that reflects the observed benefit in the real world under the optimal treatment. Since the prescribed treatments in the EHR data are likely to be at least beneficial (e.g., better than random assignment) (Wallace et al., 2016), the "super" features learned in the EHRs can be potentially informative of the optimal ITRs for real world patients as well as RCT patients. Next, we augment the feature space of the clinical trial data by these two features when learning final ITRs using only trial patients. To enhance the signal from super features, we propose stratified learning to estimate the optimal ITRs separately within each stratum defined by the super features. Particularly, we apply Q-learning and a modified M-learning to estimate the optimal ITRs in each stratum.

We provide theoretical justification of several advantages of the proposed method. First, since the final optimal ITRs are estimated using the RCT data, they remain valid due to the virtue of randomization regardless of whether EHR super features are used. Second, since the super features are informative of the treatment benefit and optimal ITRs, our learning method, which is based on stratifying by super features, should yield more precise estimation of ITRs compared to the methods without these features. Finally, the included super features are learned using the EHR data so they are very likely to be correlated with the actual optimal treatments for the EHR population. Thus, even if the trial population may be different from the EHR population, the optimal ITRs from our method, which are partially directed by the super features, are potentially more generalizable to the EHR population. As a note, the proposed method falls into a general framework of transfer learning in machine learning community, which refers to domain adaptation by allowing different assumptions in different domains with a single task in hand (Pan and Yang, 2010). Reweighting or data transformation methods are commonly used techniques to handle

challenges in domain adaptation (Zhang et al., 2013). However, our method does not rely on any weighting or transformation, but instead, extracts most informative features to improve ITR estimation in RCTs.

We conduct simulation studies to demonstrate performance of the domain adaptation learning compared to the methods without using information from the EHRs. Lastly, we apply our method to derive super features from EHRs of type 2 diabetes (T2D) patients to improve learning individualized insulin therapies from a T2D randomized trial, DURABLE study (Fahrbach et al., 2008). We show that directly applying ITRs learned from one domain to the other performs even worse than universal treatment strategy, while proposed domain adaptation leads to improvement in value function. We conclude the chapter with discussions and future extensions.

3.2 Method to Improve ITRs by Borrowing Evidence from EHRs

3.2.1 Learning the optimal ITR using RCT data

Let X denote the features collected prior to treatment and let A denote the binary treatment assignment coded as $\{-1, 1\}$. Let R_i denote the clinical outcome after treatment. Assume a larger R corresponds to better treatment outcome (e.g., reduction in symptoms). An ITR is a decision rule, denoted as $\mathcal{D}(X)$, which maps the feature space to the treatment decision space. The value function associated with $\mathcal{D}$ used to evaluate an ITR is defined as the expected post-treatment outcome by following $\mathcal{D}$ to assign treatments, that is, $V(\mathcal{D}) = E^{\mathcal{D}}(R)$, where $E^{\mathcal{D}}$ refers to the expectation under a probability distribution with $A = \mathcal{D}(X)$. When the treatment assignment mechanism is known (e.g., for a RCT), this expectation can be equivalently expressed as

$$V(\mathcal{D}) = E\left[\frac{R}{\pi(A, X)} I\{A = \mathcal{D}(X)\}\right],$$

where $\pi(a, x)$ is the randomization probability for $A = a$ given $X = x$. Thus, the goal of learning the optimal ITR is to find $\mathcal{D}(\cdot)$ that maximizes $V(\mathcal{D})$. The corresponding empirical value function using n i.i.d. observations collected from a RCT can be expressed as

$$V_n(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} \frac{I\{A_i = \mathcal{D}(X_i)\}R_i}{\pi(A_i, X_i)}. \tag{3.1}$$

There are many statistical and machine learning methods developed to estimate the optimal ITR. They can all be unified into maximizing some surrogates of $V_n(\mathcal{D})$, in which different weights and loss functions are used to replace R and $I(A = \mathcal{D}(X))$, respectively. For example, in terms of the surrogate weight for R, Q-learning replaces R by the estimated treatment benefit based on a regression model (Qian and Murphy, 2011), doubly-robust O-learning replaces R by a doubly-robust augmented residual of $R - E(R|X)$ (Liu et al., 2018), and M-learning uses R subtract the averaged outcomes from matched subjects who receive opposite treatment as the weight (Wu et al., 2019). For the surrogate loss function, both doubly-robust O-learning and M-learning use the hinge-loss to replace the zero-one loss in $V_n(\mathcal{D})$.

3.2.2 Domain adaptation to improve learning ITRs

Now suppose that in addition to the RCT, we also observe the data from patients in the EHRs, which include feature variables, received treatments, and outcomes. Our goal is to extract information from the EHR data to be included in learning the optimal ITR from the trial data. We refer this information extraction as "domain adaptation" from EHR to the RCT and this framework is illustrated in Figure 3.1.

Due to the presence of unmeasured confounding in EHRs, we cannot directly learn the optimal ITR from the combined EHR and RCT data. Instead, we pre-train useful feature mappings from

EHRs to augment the feature space of RCT. Since physicians' treatment decisions documented in the EHRs are likely to carry clinical insights and deemed to be beneficial to a patient among available options, they are likely closer to being optimal than random assignments (as in RCT). Thus, features that can summarize physicians' prescription patterns are informative of the optimal treatment for a patient. Furthermore, because physicians may prescribe a treatment based on many considerations including efficacy, risk of complications and cost, their prescription patterns may not be sufficient when the goal is to learn an ITR to maximize efficacy. The observed benefit under the optimal treatment in the EHRs captured by an efficacy outcome similar to RCT is also useful.

Therefore, in the first step of domain adaptation, we capture additional information available in the EHRs but not in the RCTs (i.e., a physician's judgment of beneficial treatment and patient's observed benefits). We create feature mappings predictive of optimal ITRs and observed benefits referred to as "super" features. Specifically, we pre-train an ITR to determine each EHR patient's optimal treatment by machine learning algorithms described in Section 3.3. The first super feature is an optimal treatment propensity measure denoted as H_1, defined as the probability that a physician prescribes a particular treatment a_0 as the most beneficial choice estimated from all EHR patients who have received optimal treatments predicted by an algorithm (either predicted optimal is a_0 or $-a_0$). This variable captures the concordance between clinician-based optimal treatment assignment documented in the EHRs and algorithm-based optimal treatment computed by the ITRs. To construct this feature, we learn an ITR using a common set of features in EHR and RCT, say, X_c and methods in Section 3.3. Next, we fit a classification model (e.g., by random forest classifier or logistic regression) to estimate the probability of a_0 being optimal given X_c among the subset of EHR patients who received optimal treatments as predicted by the ITRs. For information

transfer, we apply the fitted model to RCT patients and obtain their predicted optimal treatment propensity measure as the first super feature.

The second super feature is a benefit feature (denoted as H_2), which measures EHR patient's observed gain or loss on an outcome under the optimal treatment. That is, H_2 is the expected difference in outcome when a subject receives the optimal treatment compared to non-optimal treatment. To obtain H_2, we first fit a random forest regression model for outcomes under optimal treatment to estimate $E(R|A = a^*, X_c) = m_1(X_c)$, using the subset of EHR patients who received the optimal treatments a^*. Similarly, we fit a model for outcomes under non-optimal treatment, $E(R|A = -a^*, X_c) = m_2(X_c)$, using subjects who received non-optimal treatments. The second super feature is the predicted benefit $H_2 = \widehat{m}_1(X_c) - \widehat{m}_2(X_c)$.

In the second step of domain adaptation learning, we estimate the final optimal ITRs from RCT patients in the augmented feature space (original RCT features and EHR super features H). One approach is to learn optimal ITRs using the combined features from (H_1, H_2) and X. However, directly including super features into the feature set and using the same tuning parameter in the ITR learning may not distinguish their importance from the original RCT features so may weaken their effects. To amplify the signals of the super features, we can treat them separately as important predictive variables through stratification. More specifically, using one of the super features or both to stratify patients into multiple strata, we learn optimal ITRs separately within each stratum.

The procedure of our method can be summarized in Algorithm 1.

The algorithms (Q-learning and M-learning) to learning the ITRs in RCT and EHR are described in Section 3.3.

Algorithm 1 The Algorithm for Domain Adaptation Learning

Step 0. Use machine learning methods (Q-learning or M-learning) in Section 3.3 to pre-train ITRs from EHR patients to predict optimal treatment with common features X_c available in RCT and EHR.

Step 1. Construct H_1 and H_2 from EHR patients by:

1a. Learn $H_1 = P(A = a_0|X_c)$ by random forest classification among EHR patients who received optimal treatments predicted by ITRs obtained in Step 0, where a_0 is a pre-specified treatment.

1b. Learn $H_2 = E(R|A = a^*, X_c) - E(R|A = -a^*, X_c)$ using random forest regression separately for EHR patients who received optimal treatments and non-optimal treatments, where a^* denotes the optimal treatment predicted by ITRs in Step 0.

1c. Predict H_1, H_2 on RCT patients.

Step 2. Learn final ITRs from RCT using algorithms in Section 3.3 by:

1. using all RCT subjects and features (X, H_1, H_2); OR
2. stratify RCT subjects by H_1 and/or H_2 and learn separate ITR in each strata.

3.2.3 Justification of domain adaptation learning

Essentially, the proposed domain adaptation learning obtains the optimal ITR as

$$\mathcal{D}^* = \max_{\mathcal{D}} E\left[\frac{\widetilde{R}I\{A = \mathcal{D}(X)\}}{\pi(A, X)}\Big|H\right],$$

where $\widetilde{R}$ is the surrogate outcome for R depending on which learning method is used. Note that $\widetilde{R} = E[R|A = 1, X, H] - E[R|A = -1, X, H]$ in Q-learning, $\widetilde{R} = R$ in the original O-learning, $\widetilde{R} = R - E[R|X, H]$ in an augmented doubly-robust O-learning, and it is $R(a) - E[R|A = -a, X, H]$ in M-learning where a is the treatment actually received. Correspondingly, let $\widehat{\mathcal{D}}^*$ be the estimated ITR using empirical data as given in the chapter. Let $\mathcal{D}^0$ be the optimal ITR that maximizes $E[\check{R}I(A = \mathcal{D}(X))/\pi(A, X)]$ without super features, where $\check{R}$ is similar to $\widetilde{R}$ except that the former is calculated without stratifying by H but the latter with, and let $\widehat{\mathcal{D}}^0$ be its estimator.

Case I. When there is no structural assumption on $\mathcal{D}(x)$, since H is a function of X, it is clear

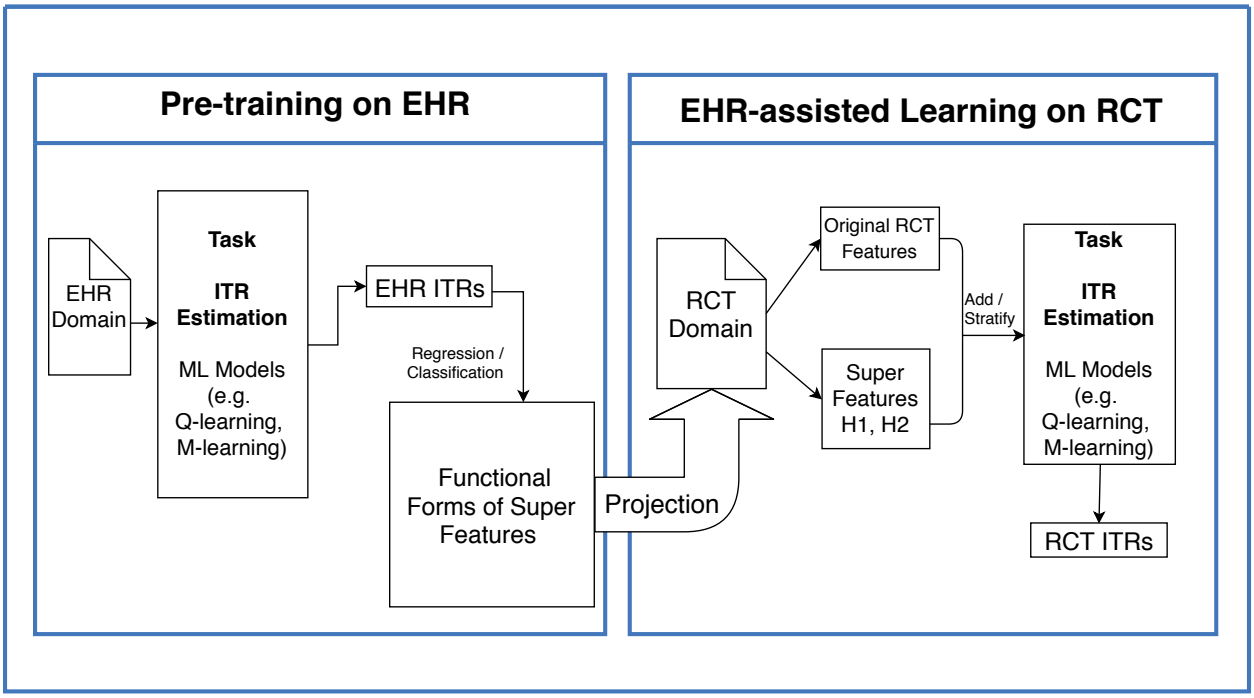


Figure 3.1: Schematics of Proposed Domain Adaptation from EHR to RCT

that $\mathcal{D}^*(x)$ maximizes $V(\mathcal{D})$. Thus, both $\mathcal{D}^*(x)$ and $\mathcal{D}^0$ yield the same optimal rule. Furthermore, following the same derivation as in (Liu et al., 2018), we have

$$V(\widehat{\mathcal{D}}^*) - V(\mathcal{D}^*) \leq (E\widetilde{R}^2)^{\alpha} a_n + b_n$$

and

$$V(\widehat{\mathcal{D}}^0) - V(\mathcal{D}^*) \leq (E\check{R}^2)^{\alpha} a_n + b_n,$$

where α is a constant depending on the underlying distribution and the dimension of X, and a_n and b_n are constants depending on n and the geometric noise index in the underlying distribution. Since $\widetilde{R}$ is obtained and centered in each stratum but $\check{R}$ is not, we expect

$$E\widetilde{R}^2 \leq E\check{R}^2,$$

and we conclude that the domain adaptation ITR, $\widehat{\mathcal{D}}^*$ leads to a more efficient value than the ITR

without H. In addition, when H is more predictive of R, which is likely to hold since H contains the predictive benefit feature H_2 using the EHR data, more efficiency is expected when using the domain adaptation ITR.

Case II. Suppose that there is a structural assumption on $\mathcal{D}(x)$, for instance $\mathcal{D}(x)$ is linear when learning ITRs. Since the domain adaptation ITR is learned in each stratum of H, the class of ITRs considered in the domain adaptation learning is a stratified (piece-wise) linear rule which is broader than the non-stratified rules. We thus conclude $V(\mathcal{D}^*) \geq V(\mathcal{D}^0)$. Furthermore, if the treatment benefit is more heterogeneous across H, which likely holds because H contains H_2, then the gain of the value from the domain adaptation ITR is even more significant.

Case III. Consider that some structural assumption is placed on the ITRs, for example, linear in feature variables. We notice that $\mathcal{D}^*$ maximizes

$$\max_{\mathcal{D}} E\left[\left\{R^{(1)}(X) - R^{(-1)}(X)\right\} I\{\mathcal{D}(X) = 1\}\Big| H\right],$$

where $R^{(1)}$ and $R^{(-1)}$ denote the potential outcomes for treatment 1 and -1 in the trial population, respectively. In the EHR population, the treatments may act similarly in terms of qualitative effects (sign of the treatment effect), but the magnitude of the treatment effects (benefits) may not be as large as what is seen in the trial population which is well managed and performed under ideal conditions. As such, it is reasonable to assume that the potential outcomes for the EHR population, denoted by $\widetilde{R}^{(a)}, a = 1, -1$, satisfy

$$R^{(1)}(X) - R^{(-1)}(X) = \left(\widetilde{R}^{(1)} - \widetilde{R}^{(-1)}\right) g(X), \tag{3.2}$$

where $g(X) > 0$ and it reflects the heterogeneous ratios of the treatment effects across subgroups between the two population. Hence, the domain adaptation optimal rule maximizes

$$\max_{\mathcal{D}} E\left[g(X)\left\{\widetilde{R}^{(1)} - \widetilde{R}^{(-1)}\right\} I(\mathcal{D}(X) = 1)\Big| H\right].$$

On the other hand, the EHR super features, H, are highly associated with the treatment effects in the EHR population. That is, $g(X)$ is highly correlated with H. We conclude that the domain adaptation optimal rule approximately maximizes

$$\max_{\mathcal{D}} E\left[\left\{\widetilde{R}^{(1)} - \widetilde{R}^{(-1)}\right\} I(\mathcal{D}(X) = 1)\Big| H\right].$$

In other words, the domain adaptation rule leads to an approximately best linear rule maximizing the value as if treatment outcomes were obtained from the EHR population. In contrast, without using EHR features, $\mathcal{D}^0$ maximizes

$$E\left[g(X)\left\{\widetilde{R}^{(1)} - \widetilde{R}^{(-1)}\right\} I\{\mathcal{D}(X) = 1\}\right].$$

We note that such a rule is not only driven by the treatment effects in the EHR population, but also depends on the magnitude of $g(X)$. For example, if some group has a large $g(X)$, the optimal rule $\mathcal{D}^0$ is likely to be the optimal linear rule for this particular group but not generalizable to others. We conclude that domain adaptation ITRs are more generalizable to the EHR population than ITR from RCT features alone. In summary, we obtain the following conclusions:

(1) If ITRs are learned nonparametrically, the domain adaptation rule leads to the optimal treatment rule with a more efficient value estimation as compared to the optimal rule without the EHR super features H.

(2) If ITRs are learned in some restrictive class (e.g., linear rules), the domain adaptation rule always leads to a higher value than the one without H; moreover, it can be more generalizable to the EHR population when this population has different magnitudes of treatment effects compared to the trial population.

Remarks: The above conclusion (1) shows that since H are constructed from common features in RCT and EHR, if the ITRs are learned nonparametrically from RCTs, we may not expect to

obtain a higher value function than not using H (especially with a large sample). However, by carefully constructing H to be predictive of optimal rules and observed outcomes from an external source (EHRs), the precision of value estimation can be improved. Moreover, from conclusions (2) and (3), when estimating parametric ITRs, including H will lead to a higher value and precision, and a greater generalizability (under assumptions).

3.3 Algorithms for Estimating ITRs

In Algorithm 1, the pre-training step 0 and the final stratified learning step 2 require some method to estimate the optimal ITRs. In this section, we describe two learning algorithms that are implemented in our numeric studies.

The first learning algorithm is Q-learning, one of the popular methods for estimating ITRs. We first fit a predictive model (e.g., random forest regression) with R as output and A and X as inputs. Next, the optimal treatment is selected as $a^* = \arg\max_{\{a=1,-1\}} \widehat{f}(X, a)$ where $\widehat{f}$ is the predicted mean of R.

The second learning algorithm is M-learning (Wu et al., 2019) with an improvement we develop in this work. This choice of algorithm is based on the fact that it is a general method including O-learning as a special case. In this method, a matching function is introduced to estimate individual treatment responses under alternative treatment assignment. Furthermore, doubly-robust matching through using prognostic scores in M-learning can further improve efficiency of ITR estimation and it is expected to perform better especially when the treatment assignment is imbalanced as often for observational studies. In the original work (Wu et al., 2019), the matched set can be empty; to avoid this situation, we modify their approach by introducing a soft-matching method, where all

possible pairs are used but weighted differently according to their feature similarities or prognostic scores. More specifically, in matched learning (Wu et al., 2019), the value function is re-expressed in a matching-based alternative form as

$$\begin{aligned} V_n(\mathcal{D};u) &= n^{-1}\sum_{i=1}^{n}|\mathcal{M}_i|^{-1}\sum_{j\in\mathcal{M}_i}\Big[I\{R_j\geq R_i,\mathcal{D}(X_i)=-A_i\} \\ &\quad +I\{R_j\leq R_i,\mathcal{D}(X_i)=A_i\}\Big]u(|R_j-R_i|), \end{aligned} \tag{3.3}$$

where $\mathcal{M}_i$ is a matching set for subject i that consists of subjects with similar covariates defined under a suitable distance metric but opposite treatments. The function $u(\cdot)$ is a monotonically increasing function to weight different subjects' outcomes and in our implementation, we choose $u(x)=x$. The rationale of (3.3) is that for two subjects who are matched in confounders or propensity scores of treatments but are observed to receive different treatments, the subject with a higher outcome should be more likely to have received the optimal treatment. Equivalently, the objective function can be expressed as

$$n^{-1}\sum_{i=1}^{n}|\mathcal{M}_i|^{-1}\sum_{j\in\mathcal{M}_i}I(f(X_i)A_i\text{sign}(R_j-R_i)\leq 0)|R_j-R_i|.$$

A disadvantage of the above matching function is that only a limited number of pairs of subjects are used. To improve this limitation while accounting for dissimilarity in confounding variables of subject pairs measured by a suitable distance (e.g., Euclidean or Mahalanobis distance in the feature space), a kernel weighted objective function can be defined as

$$\begin{aligned} V_n(f) &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}I(A_i\neq A_j)k_{a_n}\{s(X_i,X_j)\} \\ &\quad I\{f(X_i)A_i\text{sign}(R_j-R_i)\geq 0\}u(|R_j-R_i|), \end{aligned} \tag{3.4}$$

where $k_{a_n}(\cdot)$ is a kernel function with bandwidth a_n. For example, with Gaussian kernel $k_{a_n}\{s(X_i,X_j)\}=$

$\exp(-a_n\|X_i, X_j\|^2)$, where $\|\cdot\|$ denotes some suitable distance function. Note that in (3.4), subject pairs with different treatments and more similar feature variables or confounding variables and larger difference in clinical outcomes will receive higher weights. Kernel M-learning thus extracts information from all possible pairs but adjusts their contribution to the objective function based on their similarity in confounding variables and differences in clinical outcomes. To solve the optimization problem based on $V_n(f)$ in (3.4), one can replace the zero-one loss by other surrogate loss function. Specifically, when replacing by the hinge-loss, the optimization problem is transformed to a weighted support vector machine (SVM) for kernel-weighted pairs:

$$\begin{aligned} L_n(f) &= n^{-2}\sum_{i=1}^{n}\sum_{\{j:A_i\neq A_j\}}\phi\{-f(X_i)A_i\text{sign}(R_j-R_i)\} \\ &\quad k_{a_n}\{s(X_i,X_j)\}u(|R_j-R_i|)+\lambda_n\|f\|_{\mathcal{X}_K}, \end{aligned} \tag{3.5}$$

where $\phi(x) = (1-x)_+$, λ_n is a tuning parameter and $\mathcal{X}_K$ is a reproducing kernel Hilbert space (RKHS) with kernel function $K(\cdot,\cdot)$. The dual problem of (3.5) is a quadratic programming problem which can be solved by quadratic programming packages (e.g., through a weighted SVM with matched pairs) similar to (Wu et al., 2019).

We describe solution to the quadratic problem of weighted SVM for kernel-weighted pairs in (3.5) using Lagrange multipliers and take linear ITR decision rules as an example. Assume f is linear in $V_n(f)$ and $f(x) = \langle\beta, x\rangle + \beta_0$ where $\langle\cdot,\cdot\rangle$ denotes the inner product operator and $\|f\|_{\mathcal{X}_K}$ represents $\|f\|^2$ in Euclidean space. Equivalently, it is convenient to re-write (5) as

$$\min \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{n}\sum_{\{j:A_i\neq A_j\}} k_{a_n}\{s(X_i,X_j)\}u(|R_j-R_i|)\xi_{ij},$$

$$\text{subject to: } A_i\text{sign}(R_i-R_j)(\langle\beta,X_i\rangle+\beta_0)\geq(1-\xi_{ij}),\xi_{ij}\geq 0,\forall i \text{ and } j\in\{j:A_i\neq A_j\},$$

where C is a "cost" parameter, ξ_{ij} is a slack variable that denotes a small portion of misclassification error for the jth subject which is matched with the ith subject and $k_{a_n}\{s(X_i, X_j)\}u(|R_j - R_i|)$ is the sample weight in a SVM framework (Wu et al., 2019).

The Lagrange primal function is

$$\frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{n}\sum_{\{j:A_i\neq A_j\}} k_{a_n}\{s(X_i, X_j)\}u(|R_j - R_i|)\xi_{ij}$$
$$-\sum_{i=1}^{n}\sum_{\{j:A_i\neq A_j\}} \alpha_{ij}\{A_i\text{sign}(R_i - R_j)(X_i^T\beta + \beta_0) - (1 - \xi_{ij})\} - \sum_{i=1}^{n}\sum_{\{j:A_i\neq A_j\}} \mu_{ij}\xi_{ij}$$

where we minimize with respect to β, β_0 and ξ_{ij}. By taking the respective derivatives and setting them to zero to obtain,

$$\begin{cases} \beta = \sum_{i=1}^{n}\sum_{\{j:A_i\neq A_j\}} \alpha_{ij}A_i\text{sign}(R_i - R_j)X_i, \\ 0 = \sum_{i=1}^{n}\sum_{\{j:A_i\neq A_j\}} \alpha_{ij}A_i\text{sign}(R_i - R_j), \\ \alpha_{ij} = Ck_{a_n}\{s(X_i, X_j)\}u(|R_j - R_i|) - \mu_{ij}, \forall i \text{ and } j \in \{j : A_i \neq A_j\}. \end{cases}$$

By substituting above equations into Lagrangian dual function, we obtain

$$\max \sum_{i=1}^{n}\sum_{\{j:A_i\neq A_j\}} \alpha_{ij} - \frac{1}{2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{\{j:A_i\neq A_j\}}\sum_{\{j':A_{i'}\neq A_{j'}\}} \alpha_{ij}\alpha_{i'j'}A_iA_{i'}\text{sign}(R_i - R_j)\text{sign}(R_{i'} - R_{j'})\langle X_i, X_{i'}\rangle$$

subject to $0 \leq \alpha_{ij} \leq Ck_{a_n}\{s(X_i, X_j)\}u(|R_j - R_i|)$ and $\sum_{i=1}^{n}\sum_{\{j:A_i\neq A_j\}} \alpha_{ij}A_i\text{sign}(R_i - R_j) = 0$. In addition, subject to Karush-Kuhn-Tucker conditions for $\forall i$ and $j \in \{j : A_i \neq A_j\}$ (Zhao et al., 2012):

$$\begin{cases} \alpha_{ij}[A_i\text{sign}(R_i - R_j)(X_i^T\beta + \beta_0) - (1 - \xi_{ij})] = 0, \\ \mu_{ij}\xi_{ij} = 0, \\ A_i\text{sign}(R_i - R_j)(X_i^T\beta + \beta_0) - (1 - \xi_{ij}) \geq 0, \end{cases}$$

the solution to the primal and dual problem is optimal. It is not difficult to extend the solution using other kernels (e.g., Gaussian kernel) and obtain a nonparametric ITR decision rule based on kernel function $K(\cdot,\cdot)$ in the RKHS.

3.4 Simulation Studies

In this section, we generated two separate domains of data source to mimic a scenario involving data from both an RCT and EHR. To accommodate treatment benefit heterogeneity across patient populations observed in real-world data, we considered the underlying true optimal ITR with a piece-wise linear tree structure. Specifically, outcome data were simulated as

$$C1 : R = \eta(X) + \phi(X) * A + \epsilon, \epsilon \sim N(0, 0.25),$$

where $\phi(X) = 0.5I(X_1 + X_2 > 0) - 0.5I(X_1 + X_2 <= 0)[1 + I(X_2 <= -0.5)] + X_3^2 - X_2^2$, feature variables X_k were generated from a standard normal distribution, and $\eta(X) = X_1 - 0.5X_2$. Here, $\eta(X)$ is the main effect, and the sign of $\phi(X)$ defines the true optimal ITR. The true treatment propensity model in the observational study was specified as $P(A = 1|X) = \text{expit}(1 + 2X_1 + X_2)$ and the treatment assignment probability for RCT was 0.5.

To accommodate heterogeneity between RCT and EHR patients and unobserved tailoring variable, we considered two scenarios:

(i) All features are observed but X_1 has a different distribution in RCT where X_1 is restricted to $[-0.4, 0.4]$;

(ii) Same as (i) but with an additional feature variable X_3 as an unobserved tailoring.

Scenarios (i) and (ii) mimic randomized trials where subjects are recruited from subpopulations under certain restrictive inclusion criteria, while the EHR data represent the more general real-world

patient population.

We compared four strategies of learning ITRs, one using RCT information alone and three domain adaptation learning:

(S1) Using RCT data and RCT features only;

(S2) Augment the RCT feature set by EHR super features H_1, H_2 (section 2.2);

(S3) Include H_1 in the feature set and stratify by H_2;

(S4) Include H_2 in the feature set and stratify by H_1.

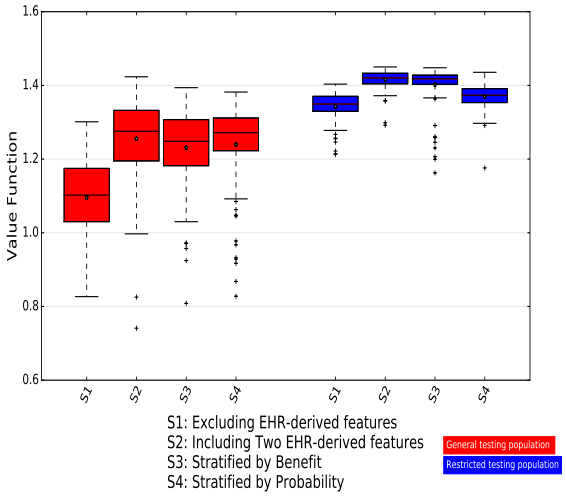
The super feature H_1 was estimated from a random forest classification model and H_2 from a random forest regression model. To speed up computation, the tuning parameter a_n in kernel M-learning was chosen so that the matched pairs with distance less than a_n was fixed to be a proportion of pairs (e.g., 25%) (Liu et al., 2017). In S3, we stratified the training dataset and testing set based on a median split of the average predicted benefit H_2 and included H_1 as an extra feature variable in learning ITR. Similarly, in S4, we stratified the data by the predicted optimal probability H_1 (median split) and included H_2 as an additional feature. We considered stratifying by both H_1 and H_2 in a simulation in Appendix B.

Q-learning and kernel M-learning were performed under each strategy. In Q-learning, linear regression was used to fit the linear rule and random forest regression was used to fit the nonparametric rule. In M-learning, weighted SVM with a linear kernel or Gaussian kernel was used. The tuning parameters for the latter (e.g., cost parameter, bandwidth for Gaussian kernel) were selected by two-fold cross validation. The value function of the estimated optimal rule was computed on a large independent testing set (sample size of $10,000$) generated from the general population (EHR) without restricting X_1 or restricted population (RCT) while other procedures remain the same. We repeated the simulations 100 times.

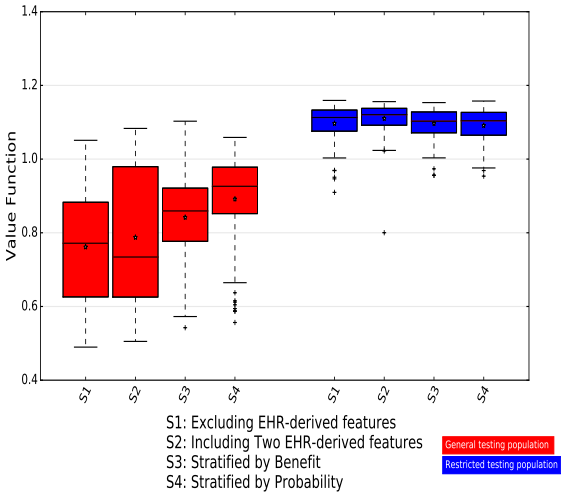
Simulation results for M-learning with fitted nonparametric ITRs evaluated on both general (red) and restricted testing population (blue) are summarized in Figure 3.2. The optimal empirical value function is 1.48. When testing on the general population, adding super features directly reduces variability in scenario (i) and also improves value to 1.26 from the original value of 1.10. Domain adaptation by stratification also performs better than without considering super features in both (i) and (ii), which is consistent with our theoretical justification given in Section 3.2.3 (Case I). When testing on the restricted distribution (RCT), adding super features directly achieves the highest mean value in both scenarios (i) and (ii) (Figure 2a and 2b). The value functions evaluated on the general population are lower than the restricted population, demonstrating that optimal rule fitted from the restricted RCT population may not achieve the same effect in the general population. However, the difference between populations is smaller for the domain adaptation rules, showing that they may be more generalizable.

In another set of analyses, we fitted linear rules to scenarios (i) and (ii), which are misspecified in these settings since the true underlying optimal decision function has a piece-wise linear structure. The results are summarized in Figure 3.2c and 3.2d. Both stratification methods lead to a large improvement in scenario (i), demonstrating our theoretical justification in Section 3.2.3 (Case II). Furthermore, domain adaptation rules reach a similar value on the general testing (EHR) population and restricted (RCT) population. Thus, we show that even though domain adaptation rules are fitted from RCT population, they behave as if learned from the general population with a better generalizability. This is consistent with our theoretical justification that domain adaptation learning is more generalizable due to that super features are highly correlated with the optimal EHR rules (Section 3.2.3, Case III). The results for Q-learning show similar trends (Appendix B.1). Additional simulation results for kernel M-learning under presence of unmeasured confounders are

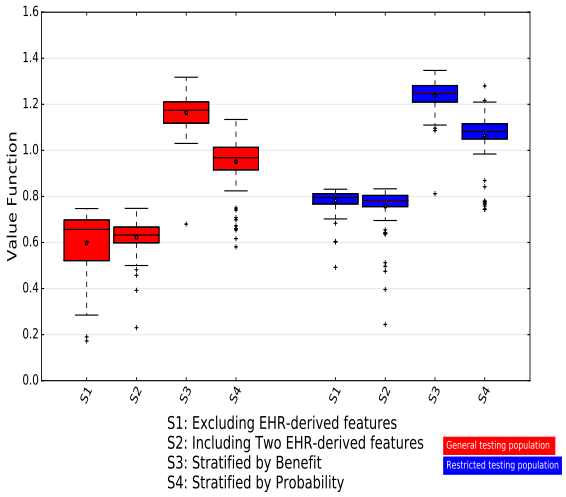
presented in Appendix B.2.



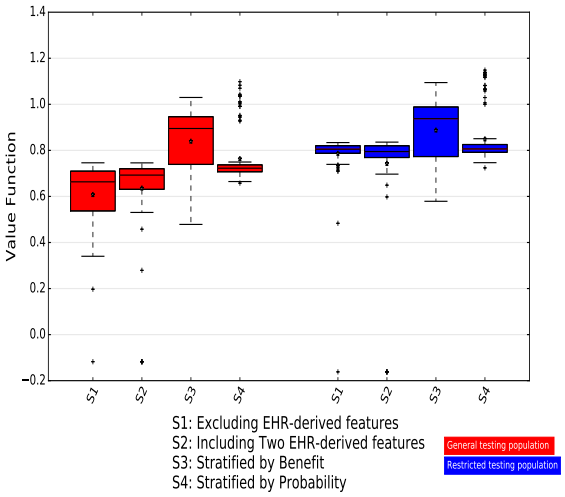
(a) Scenario (i), nonparametric rule



(b) Scenario (ii), nonparametric rule



(c) Scenario (i), linear rule



(d) Scenario (ii), linear rule

Figure 3.2: Simulation Comparisons for M-learning (evaluated on independent testing sets generated from general or restricted population; scenario (i) has no latent tailoring variables while scenario (ii) has a latent tailoring variable not used in learning)

To summarize, domain adaptation learning assisted by EHR super features (by directly including H or stratifying by H) improves performance of ITR estimation (a higher value or a smaller

variability). The improvement is observed when evaluated both on the general population and restricted population. The difference of value between populations is smaller with domain adaptation than without. The simulations suggest that the information gained from EHR may be transferred to ITR estimation on the RCT data.

3.5 Applications

3.5.1 Motivating Studies

Our research goal is to optimize insulin therapy for T2D patients based on their individual characteristics from RCT data assisted by real-world clinical practices documented in the EHRs. A randomized controlled trial, DURABLE, was conducted to compare insulin lispro mix 75/25 (fast-acting medication) versus insulin glargine (Fahrbach et al., 2008). Over 2,000 patients were enrolled in this study from 11 countries between 2005 and 2007. The study was designed to compare safety and efficacy of two insulin types with a 6-month initiation phase. There were 965 patients randomized to lispro mix and 980 patients to insulin glargine. The primary outcome was hemoglobin A1C (HbA1c) reduction at the end of study (1 year post treatment) and the enrolled patients had a median baseline HbA1C of 8.8% (Fahrbach et al., 2008).

In recent years, the use of EHRs is continuously growing among clinical researchers with access to large-scale clinical data warehouses and databases (Weiskopf and Weng, 2013). and the EHR warehouses contain massive information that can be used to assist researchers with medical decision making (Weiskopf and Weng, 2013). For example, EHRs were used to construct ITRs from real-world patients (Wang et al., 2016). In this chapter, we extracted EHRs from T2D patients at New York Presbyterian Hospital (NYPH) clinical data warehouse (CDW). The CDW has

implemented a well-defined quality control process and studies were launched to investigate data quality including completeness, correctness, concordance, plausibility, and currency (Weiskopf and Weng, 2013). Recent EHRs in NYPH CDW were used to learn treatment pathways of various common diseases including T2D (Hripcsak et al., 2016). The main information contained in the CDW includes demographics, in-patient and out-patient medication prescriptions, ICD diagnosis codes and laboratory tests, which are longitudinally documented (Wu et al., 2019).

Subjects were included in the EHR analysis if they had insulin aspart or insulin glargine, and had at least one observation during one year follow up period post insulin initiation. Literature reveals insulin aspart and insulin lispro are two comparable rapid-acting analogs with similar profile which have a shorter duration of action, while insulin glargine is a long-acting medication (Plank et al., 2002; Raja-Khan et al., 2007). In the EHR data, there were $1,741$ T2D patients on long-acting insulin glargine and $1,016$ on fast-acting insulin aspart. In domain adaptation, we aim to borrow information from super features on the optimal treatment strategies between two treatments (insulin aspart vs. insulin glargine) learned from EHR patients and apply to RCT patients to improve estimation of ITRs. The RCT data are the primary source for learning ITRs to maintain consistency of the optimal rule due to the virtue of randomization, while EHRs are auxiliary information to improve efficiency and accuracy of the ITR learning.

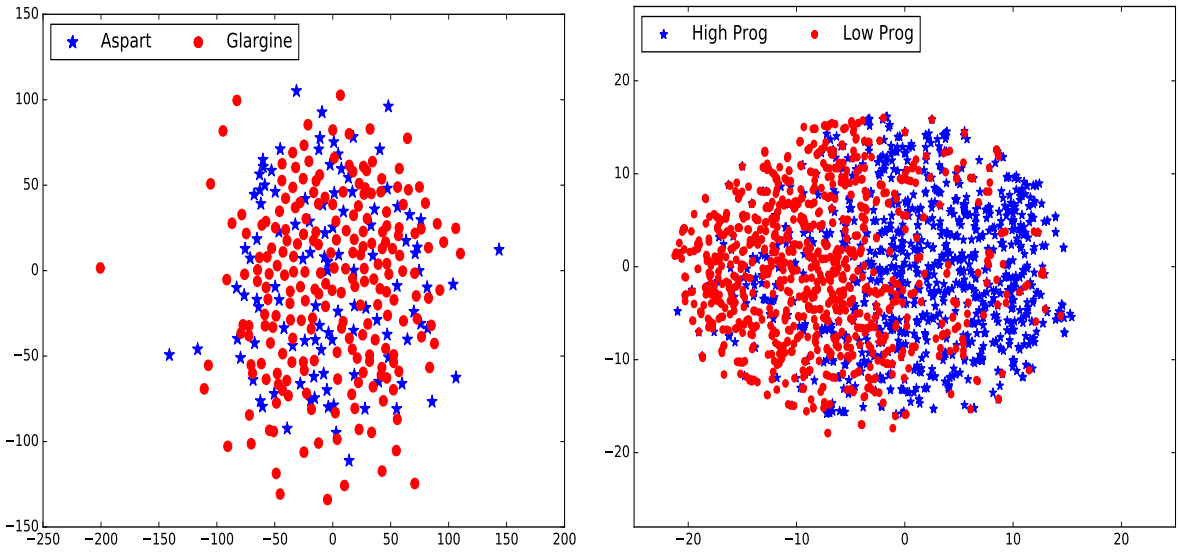
3.5.2 Analysis Results

We applied four strategies S1 - S4 as examined by simulations. Our goal was to estimate an optimal ITR to select the best second-line T2D treatment (lispro mix vs. insulin glargine). The outcome measure was reduction in HbA1c level 12 months post insulin initiation. Baseline features extracted from EHRs include age, gender, race, baseline value and rate of change of HbA1c, glucose,

SBP, DBP, and BMI estimated from a linear mixed effects model. When creating super features from the EHR, we used inverse probability weighting (IPW) to adjust for missing outcomes. The weights were obtained by a logistic regression predicting whether a subject had any post insulin treatment HbA1c measure. For kernel M-learning, baseline and rate of change of lab test values and demographics variables were used in creating the matching set.

To explore information available in features from two data sources, in Figure 3.3 we present the t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize the feature space. The left panel shows two dimensional features embedded from t-SNE for EHR patients. The embedded features show a large overlap between the individuals who received different treatments in both dimensions. Higher dimensional t-SNE figures suggest similar overlap. Therefore, most subjects in the EHR can find matched neighbors based on their feature variables for M-learning and few extreme subjects are present. There is small subgroup of patients who received glargine clustered in the bottom of the figure. Correctly learning ITRs for these subjects will be useful in information transfer. The right panel shows two-dimensional features embedded for RCT dataset, labeled by the median split of prognostic scores estimated from a linear regression model fitted under insulin glargine group. The red dots represent subjects in high prognostic score group (larger than median) while blue stars represent subjects in low prognostic score group. Although there is some overlap, two groups clearly separate into two centers. This suggests that RCT features have some power in predicting prognostic scores and matching on prognostic scores or include them in the learning step of ITR estimation could improve efficiency.

When pre-training super features on the EHRs, Q-learning was fitted by random forest regression while kernel M-learning was performed using weighted SVM and tuning parameters (including choice of kernel as linear or Gaussian) were chosen from the whole sample. In the next step, we



(a) EHR dataset

(b) RCT dataset

Figure 3.3: t-SNE Plot for Features Extracted from CUMC EHRs and DURABLE trial

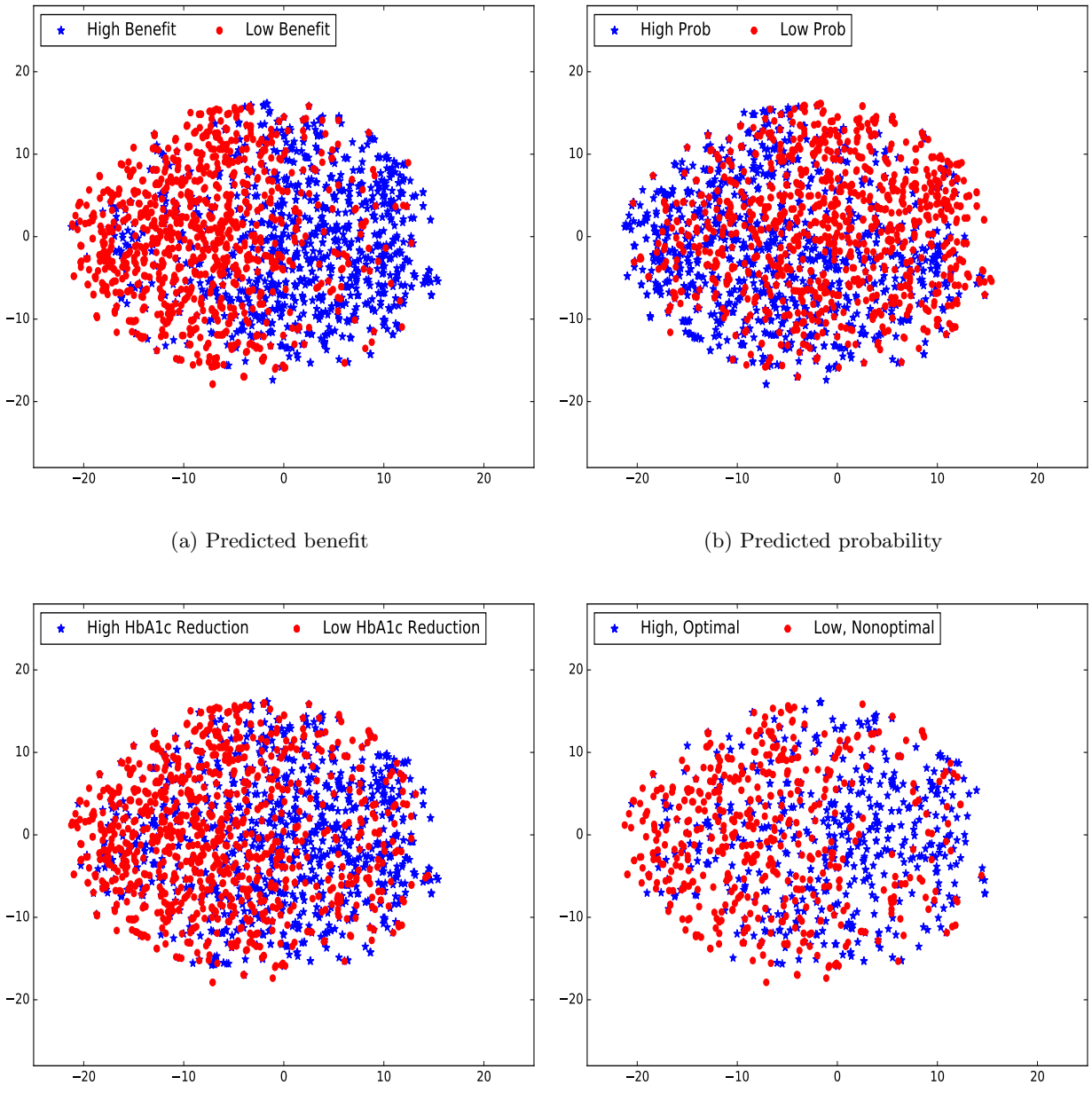
trained random forest models with inverse propensity weighting on the subgroup of patients who received the optimal or non-optimal treatment assignment according to the estimated ITRs. The features included in the models were common subset of baseline features in EHR and RCT cohort including baseline HbA1c, glucose, SBP, DBP and BMI. Furthermore, we used the trained models to predict two EHR-defined super features on each RCT subject: predicted probability of lispro mix being optimal for an individual (H_1) and predicted benefit under the optimal treatment (H_2).

To shed light on the available information from the EHR-derived super features, Figure 3.4 displays t-SNE embedding of RCT features labeled by dichotomized EHR super features. In the top panel, subjects cluster into two groups based on the dichotomized optimal benefit (H_2) and optimal probability (H_1) features, suggesting likely information gain in estimating ITRs if stratified by these features. In the lower left panel, embedded RCT features can slightly separate HbA1c reduction as the treatment outcome. On the lower right panel, adding the EHR-derived probability

feature (H_1), more subjects are further separated in terms of treatment outcome, which suggests optimal treatment probabilities are informative ("high, optimal" represents RCT patients who received the EHR-predicted optimal treatment and had high HbA1c reduction; "low, non-optimal" represents RCT patients who received the EHR-predicted non-optimal treatment and had low HbA1c reduction).

There were 18 baseline covariates in the RCT cohort along with EHR super features included to estimate ITRs from $1,945$ patients in the RCT. All covariates were standardized before fitting the model and the empirical value function of A1c reduction was estimated by 100 times of 3-fold cross-validation. In strategy S2, Q-learning included super features as additional features and M-learning included them in matching in addition to a prognostic score. In S3 and S4, the RCT sample was stratified by one of the two super features and the other was included in the learning or matching step.

The results are displayed in Table B.1 and Figure 3.5. For the non-personalized universal rules, HbA1c reduction is 1.827 in those assigned with lispro and 1.672 for glargine. Q-learning does not provide much improvement compared to the universal rule of lispro and directly incorporating super features barely helps. In S3 when stratifying by benefit feature, Q-learning tends to have a higher empirical value but also much increased variance. In S4 stratifying by probability, Q-learning has a higher value. Kernel M-learning achieves a more significant A1c reduction compared to Q-learning when including super features directly with a mean value function of 1.836. In M-learning, both stratification methods (S3 and S4) further improve the mean value function. In particular, stratifying by probability provides a large improvement with a mean value of 1.849. This suggests that incorporating EHR-derived super features transfer some useful qualitative information (optimal treatment probability) to improve performance of ITR estimation in the RCT data source.



(a) Predicted benefit

(b) Predicted probability

(c) HbA1c reduction

(d) Predicted optimal

Figure 3.4: t-SNE Plot for Features of a DURABLE trial

The increased variability of value function stratifying by probability is partially due to a higher within-group variability for subgroups defined by probability than by benefit for several important covariates (e.g. baseline HbA1c, baseline glucose).

Furthermore, we compared our results with applying ITRs learned from EHR directly on the RCT data. In Q-learning, this strategy is worse than the universal rule of assigning all to insulin glargine. In M-learning, the estimated value function was only around 1.7, which is between the means of two universal rules.

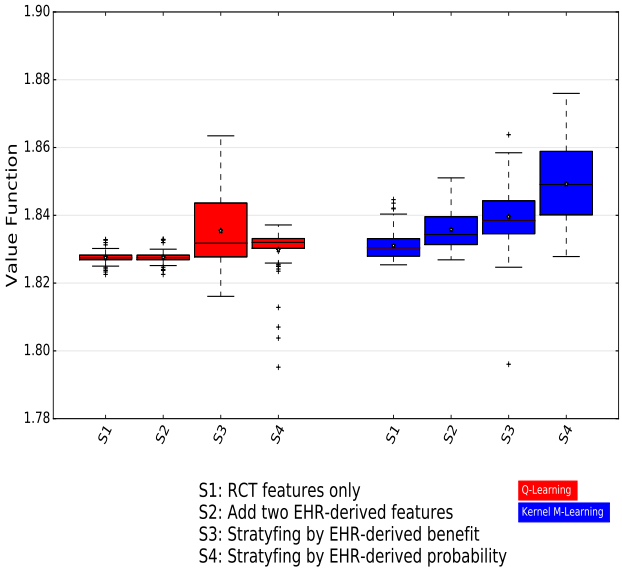


Figure 3.5: Empirical Value Function of A1c Reduction in DURABLE Trial with 100 3-fold Cross-validations

Table 3.1: HbA1c Reduction Comparing domain adaptation Learnings on DURABLE Trial (100 repetitions of 3-fold cross-validation)

One-Size-Fits-All: Lispro: 1.827, Glargine: 1.672				
	Q-Learning		Kernel M-Learning	
Strategy*	Mean (Std)	Median	Mean (Std)	Median
S1	1.828 (0.002)	1.827	1.831 (0.004)	1.830
S2	1.828 (0.002)	1.827	1.836 (0.006)	1.834
S3	1.835 (0.010)	1.832	1.840 (0.008)	1.838
S4	1.830 (0.009)	1.832	1.849 (0.011)	1.849

*: S1: RCT features only; S2: Augment RCT feature set by two EHR data-derived super features H_1, H_2; S3: Include

H_1 in the feature set and stratify by H_2; S4: Include H_2 in the feature set and stratify by H_1.

In conclusion, domain adaptation learning contributes to transfer informative feature variables extracted from EHR domain to RCT domain. Among two super features, stratifying by the qualitative probability feature and including the benefit feature in the covariate set improves performance more than stratifying by the benefit feature. In contrast, we demonstrated that direct applying fitted ITRs from EHR on RCT does not necessarily lead to a better value function.

3.6 Discussion

In this chapter, we propose domain adaptation learning to transfer information from observational studies to randomized experiments. In the framework, super features are pre-trained from EHR database to carry information and achieve domain adaptation to randomized trials. The probability feature is shown to be more robust than the benefit feature on the real-world EHRs, since it is more difficult to estimate benefit of the optimal treatment than assessing direction of treatments in subgroups. To obtain an efficient benefit feature, other approaches for estimating the contrast function can also be used (Tian et al., 2014). To further improve performance of domain adaptation, one may consider iteratively deriving informative features from RCT, validate on EHR and vice versa.

Our method provides performance gain when the super features learned from EHRs are informative of treatment responses in RCT. In real-world practice, valuable but unobserved tailoring variables informative of optimal treatment for real-world patients (e.g., clinician's insights and observations on patients) may be present and correlate with observed features in the EHRs. Thus, EHR super features fitted with observed variables may be informative of these latent factors and

also predictive of the optimal treatment. The higher this correlation, the more gain on efficiency and generalizability is expected for domain adaptation learning (e.g., Section 2.3). When the EHR treatments are not beneficial, our approach may not provide gain in efficiency but still remains consistent (converges to the true optimal treatment rule) because only RCT data are used for treatment rule learning. In this case, one may consider subgroups in EHRs for which the ITRs are beneficial.

The proposed pre-training in step 1 of the Algorithm 1 uses the same pre-treatment covariates in EHRs and RCT. Thus, one consideration of a good EHR/RCT pair to apply our method is the breadth of features captured in the EHR. Another consideration is the similar direction of treatment response in subgroups of two populations. While our work focuses on analysis of existing data, it can be conjectured to use EHR information when designing future RCTs, for example, recruiting patients from EHRs (Fraser et al., 2012).

Currently we only considered efficacy outcomes. It is desirable to better extract rich information on other types of outcomes available in the EHRs (e.g., adverse events and long term outcomes) to further improve ITRs fitted from the RCT. For example, clinicians prescribing treatments in real-world practice naturally take into account of both efficacy and risk of complications (Wang et al., 2018). Training ITRs to mimic this behavior would be beneficial. It is also worthwhile to examine advanced machine learning techniques to explore other features in the EHRs predictive of treatment prescriptions (e.g, text mining of physicians' notes), and consider dynamic treatment rules with longitudinal records.

The proposed domain adaptation framework is easy to implement in practice by following three steps in Algorithm 1. The main computational burden is on the analysis of the EHR using matching-based kernel weighting method, since one needs to create a large number of matched

pairs and fit weighted support vector machines with many matched pairs. When sample size is large, kernel weighting matching can be replaced by one-to-one matching (Wu et al., 2019). Lastly, one limitation of the current method is that when super features are less predictive of optimal treatment or benefit, stratification in domain adaptation learning may lose power. A more data adaptive framework to treat super features may be investigated (e.g., tuning the number of strata).

Chapter 4

Learning Personalized Treatment Rule Using Topic Modeling Features

4.1 Overview

Personalized treatment decision making has been proposed as a paradigm shift from the universal, "one-size-fits-all" strategy (Collins and Varmus, 2015) to address substantial heterogeneity between patients affected by chronic disorders such as type 2 diabetes (T2D) and mental disorders (Fried and Nesse, 2015). Recently, personalized medicine research is fast growing and greatly facilitated by large-scale data collection and technological advances in data storage and processing (Collins and Varmus, 2015). Personalized medicine allows healthcare providers to tailor treatment or treatment sequences to individual patient accounting for personalized information and patient heterogeneity (Chakraborty and Moodie, 2013).

One way to personalize medicine is to prescribe treatment using individualized treatment rules (ITRs) which are mappings from a patient's feature space (e.g., biomarkers of patient's health

status) to the space of potential treatment decisions. Randomized controlled trials (RCTs) are considered to be of high internal validity due to the virtue of randomizing treatments so that RCTs are not subject to unmeasured confounding. However, due to their stringent inclusion and exclusion criteria, RCTs lack generalizability or external validity to a broader and diversified population (Cole and Stuart, 2010). Studies comparing T2D patient population in EHRs and participants of RCTs for T2D found a large difference in their distributions (Weng et al., 2014). In contrast, EHRs provide a complementary resource for learning ITRs in a large real world patient population. Recently, large scale EHR data has been leveraged to characterize treatment pathways in medical decision making (Hripcsak et al., 2016). Due to the non-experimental characteristics in retrospectively documented observational data, it is difficult to infer causality or valid ITRs without employing confounding and selection bias reduction techniques (Rosenbaum, 2010; Wang et al., 2016). It is beneficial to integrate medical domain knowledge in feature engineering with valid machine learning and statistical methods in alleviating biases to better estimate optimal decision rule in observational studies.

Machine learning methods are commonly used to optimize treatment decisions by estimating ITRs according to patient level feature variables. Indirect methods such as Q-learning (Watkins and Dayan, 1992; Qian and Murphy, 2011) and A-learning (Murphy, 2003) fit a regression model to predict clinical outcome under alternative treatments and derive optimal ITRs by comparing predicted outcomes. Such methods are subject to model misspecification. Alternatively, outcome weighted learning (O-learning) and augmented O-learning (Zhao et al., 2012; Liu et al., 2018) are proposed to directly maximize clinical outcome after treatment. O-learning converts treatment optimization to a classification problem without requiring a model to predict the treatment responses. Tree-based methods use criteria related to testing for treatment by feature variable interaction effect

to partition patients (Dusseldorp and Van Mechelen, 2014; Laber and Zhao, 2015). Most of these methods are suitable for estimating ITRs from RCTs and cannot be directly applied to observational studies without adjusting for various sources of biases. Recently, a matching-based learning method referred as M-learning (Wu et al., 2019) is proposed to generalize O-learning methods under a unified framework and can be applicable to EHRs. However, M-learning method does not optimally engineer feature variables from EHRs.

Topic modeling techniques can be adapted to process text notes or similar text-like information in EHR data including latent semantic analysis (LSA) (Landauer et al., 1998), probabilistic latent semantic analysis (pLSA) (Hofmann, 1999). Both methods achieve a lower dimension representation for text information. In addition, latent Dirichlet allocation (LDA), as a generative statistical model for collections of discrete type of data, has also been adopted to reduce dimension and data representation (Blei et al., 2003).

In this chapter, we present analyses methods to estimate optimal individualized treatments from EHR data with two main goals. The first goal is to achieve more effective feature extraction for identifying optimal ITRs with EHRs. The extracted features should be predictive in selecting optimal treatment and interpretable to clinical practitioners. The second goal aims at improving algorithms for treatment optimization and reducing confounding bias in statistical estimation especially in real world setting such as EHRs. We will use LDA as a type of topic models to analyze medication and condition domain of EHR data to extract features and apply three learning methods (Q-learning, O-learning and M-learning) to estimate the optimal ITRs depending on a patient's personalized characteristics such as demographics and predictive markers augmented by LDA-based features. We compare the performance of the three learning algorithms with or without augmenting feature space by LDA-based topic features. Additionally, we investigate important

LDA-based features to provide the interpretation of ITRs using engineered topics in co-medication and diagnosis domains. The proposed method has the promise of discovering medical knowledge from observational data to facilitate individualized medical decision making.

4.2 Methodology

4.2.1 Review of Methods for Personalized Treatment

In this section, we review the basic concept and framework for individualized treatment rule and three methods in learning treatment rule including Q-learning, O-learning and M-learning methods. Since M-learning has been shown to be superior to the other two in certain cases in observational studies and our application in the chapter is in the area of electronic health records, we revisit the key idea and derivation of M-learning in Chapter 2.

4.2.1.1 Individualized Treatment Rule Framework

An individualized treatment rule (ITR) is a decision function that maps patients' features into the space of treatment options. In this chapter, we consider single stage decision rules and binary treatment options. Extension to multiple treatment options can be achieved using methods in (Zhou et al., 2018). Let H_i denote the feature variables measured prior to treatment and R_i denote the clinical outcome or reward post treatment and we can assume a larger R_i corresponds to more desirable treatment effect (e.g., symptom relief). Let T denotes the treatment assignment taking values of -1 or 1, and let $\mathcal{D}(H_i)$ denote an ITR. The value function used to evaluate an ITR associated with the decision rule $\mathcal{D}$ is defined as the expected post-treatment reward:

$$V(\mathcal{D}) = E^{\mathcal{D}}(R_i).$$

In particular, when the treatment assignment probability is known in an RCT, the value function in this expectation can be written as $V(\mathcal{D}) = E\left[\frac{R}{\pi(T,H)} I\{T = \mathcal{D}(H)\}\right]$, where $\pi(t,h)$ is the randomization probability for $T = t$ given $H = h$. In contrast, in an observational study (e.g. EHR data) treatment propensities $\pi(t,h)$ are usually unknown and need to be estimated from data using classification models such as logistic regression or random forest (Lee et al., 2010). The value function can be estimated empirically from data as

$$V_n(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} \frac{I(T_i = \mathcal{D}(H_i))R_i}{\pi(T_i, H_i)}. \tag{4.1}$$

Optimal ITRs are estimated from data by maximizing the above empirical value function (4.1).

4.2.1.2 Q-learning and O-learning Methods

There are many machine learning tools to estimate ITRs including Q-learning and O-learning methods. Q-learning is a reinforcement learning algorithm that plays an important role in estimating ITRs for multi-stage studies (Qian and Murphy, 2011). In Q-learning methods, regression-based approach is implemented to approximate R in $V(\mathcal{D})$ by the estimated treatment benefit through a model $R = \hat{f}(H,T)$ using H and interactions between H and treatment assignment T as predictors. The optimal ITR selected by Q-learning is expressed as $t^* = \max_{t\in\{1,-1\}} \widehat{f}(H,t)$ (Wang et al., 2016). Q-learning indirectly maximizes the value function and may be sensitive to incorrect model assumptions when feature space is high-dimensional (Zhao et al., 2012; Wang et al., 2016).

O-learning is a class of machine learning methods that directly maximize the expected clinical reward under a treatment assignment strategy. In the original O-learning, ITR estimation is transformed into a weighted classification problem by minimizing the empirical misclassification rate, $E\left[\frac{I(T_i \neq \mathcal{D}(H_i))R_i}{\pi(T_i, H_i)}\right]$ (Zhao et al., 2012). Augmented O-learning (AOL) is an improved version

of O-learning by replacing R_i with a doubly-robust residual of $R_i - E[R_i|H_i]$ (Liu et al., 2018). Efficiency gain can be guaranteed in the AOL by integrating O-learning with regression-based Q-learning in a doubly-robust manner (Liu et al., 2018). However, O-learning methods are based on inverse probability weighting (IPW) of estimated propensities $\pi(T_i, H_i)$ and may suffer from several issues such as instability of weights estimation and imbalance of covariates distribution across treatment groups (Wu et al., 2019) when applied to observational data.

4.2.1.3 M-learning Method

In this chapter, we focus on M-learning method. M-learning is a matching-based learning method that also directly maximizes the value function. It generalizes O-learning methods and integrates them with matching techniques. Specifically, M-learning performs matching to accurately measure individuals' treatment responses to alternative treatments, which requires less model specification and provides more flexibility in controlling confounders in the subgroups (Wu et al., 2019). In observational studies, $\pi(T_i, H_i)$ is unknown and confounding bias commonly complicates analyses. It is well known that IPW-based methods for adjusting for confounding suffer from high variability especially when weights $\pi(T_i, H_i)$ are small in subgroups of individuals and lack control over subgroup propensity score balance (Lee et al., 2011). Thus M-learning that avoids inverse weighting of $\pi(T_i, H_i)$ may be more advantageous over IPW-based approaches in the observational studies (Wu et al., 2019).

The key idea behind M-learning is that when two subjects are matched in terms of confounders or propensity scores of treatments but receive different treatments, the observed treatment leading to a larger clinical response should be more likely to be the optimal option. We describe M-learning method as follows: for any subject i, we first identify a matched set $\mathcal{M}_i$, which consists of subjects

with opposite treatments but similar feature variables as subject i, where similarity is defined under a suitable distance metric. That is, we let

$$\mathcal{M}_i = \{j : T_j = -T_i, d(H_j, H_i) \leq \delta_i\}.$$

For any given ITR, we define the matching-based preference value function to maximize as

$$\begin{aligned} V_n(\mathcal{D}; u) &= n^{-1} \sum_{i=1}^{n} |\mathcal{M}_i|^{-1} \sum_{j \in \mathcal{M}_i} u(|R_j - R_i|) \Big\{ I(R_j \geq R_i, \mathcal{D}(H_i) = -T_i) \\ &\quad + I(R_j \leq R_i, \mathcal{D}(H_i) = T_i) \Big\}. \end{aligned} \tag{4.2}$$

where $|\mathcal{M}_i|$ is the size of the matched set $\mathcal{M}_i$ and $u(\cdot)$ is proposed as a monotonically increasing matching function to weight individual subjects. Common choice of $u(\cdot)$ can be $u(x) = 1$ or $u(x) = x$ according to users' objective. Note that when $u(x) = x$ and we choose one-nearest-neighbor matching, the value function is weighted by the absolute difference between the clinical reward in the matched pair, i.e., $|R_j - R_i|$. Equivalently, value function (4.2) is further written as

$$V_n(f; u) = n^{-1} \sum_{i=1}^{n} |\mathcal{M}_i|^{-1} \sum_{j \in \mathcal{M}_i} u(|R_j - R_i|) I(f(H_i) T_i \text{sign}(R_j - R_i) \leq 0), \tag{4.3}$$

where $\mathcal{D}(H) = \text{sign}(f(H))$ for some decision function. To estimate the optimal ITR in (4.3), it is equivalent to minimize the loss function

$$L_n(f; u) = n^{-1} \sum_{i=1}^{n} |\mathcal{M}_i|^{-1} \sum_{j \in \mathcal{M}_i} u(|R_j - R_i|) I(f(H_i) T_i \text{sign}(R_j - R_i) \geq 0). \tag{4.4}$$

Due to the discontinuity of the indicator function in (4.4), we replace the zero-one loss by alternative loss function for optimization. In this chapter, we use hinge-loss and the optimization becomes a weighed support vector machine (SVM) problem for matched pairs:

$$V_\phi(f; u) = n^{-1} \sum_{i=1}^{n} |\mathcal{M}_i|^{-1} \sum_{j \in \mathcal{M}_i} u(|R_j - R_i|) \phi(-f(H_i) T_i \text{sign}(R_j - R_i)) + \lambda_n \|f\|_{\mathcal{H}_K}. \tag{4.5}$$

In $V_\phi(f;u)$, $\phi(x) = \max(1-x, 0)$, λ_n is a penalty parameter and $\mathcal{H}_K$ is a reproducing kernel Hilbert space (RKHS) with kernel function $K(\cdot,\cdot)$. The solution to the dual problem of (4.5) using any off-the-shelf quadratic programming packages leads to the minimization of M-learning.

Additionally, the use of idea in doubly robust matching estimator (DRME) (Antonelli et al., 2018) can improve M-learning. The DRME incorporates both propensity score and prognostic score in matching set $\mathcal{M}_i$. The propensity score model $\pi(H)$ and prognostic score model $m(H)$ are defined as

$$\pi(H) = P(T = 1|H) = g_1(H'\gamma_1)$$

and

$$m(H) = E(R|A = -1, H) = g_2(H'\gamma_2)$$

where γ_1, γ_2 denotes parameters for the two models. This extension in M-learning is doubly robust and is consistent even if one of the two models is misspecified (Wu et al., 2019).

It has been discussed in (Wu et al., 2019) that original O-learning and single-stage AOL are special cases of M-learning. Unlike O-learning methods which compare observed reward R with a constant or predicted outcome, M-learning directly compares observed individual rewards from the two subjects in the matched set $\mathcal{M}_i$. This demonstrates that M-learning methods have advantage over O-learning methods in ITR estimation leveraging individual information instead of predicted outcomes averaged across treatments (Wu et al., 2019).

4.2.2 EHR Data

Before introducing topic model as a method for feature extraction, we briefly review the EHR data that have been used in this work and the issues we have addressed for the analyses. The EHR data was collected and stored in a large clinical data warehouse (CDW) at New York Presbyterian

Hospital (NYPH) which contains over 20 years of health information for about 4.5 million patients (Johnson, 1996) . There are over 115,000 patients who had been diagnosed with type 2 diabetes (T2D) (Weng et al., 2014). The quality of the EHR data including completeness, correctness, and plausibility was investigated to ensure suitability for studying treatment sequence of common diseases (Weiskopf and Weng, 2013; Fort et al., 2014). Current T2D guidelines suggest metformin (MET) as the preferred first-line medication (Diabetes Control and Complications Trial Research Group, 1993), but in the real world setting, there is no recommendation of optimal sequence of treatments on the long-term outcomes in the literature (Bennett et al., 2011).

In this chapter we propose machine learning methods to estimate the optimal second-line treatment of T2D, specifically, we compare Met + insulin versus Met + sulfonylureas (SFU). To match this goal, we used a new-user design to select our cohort for analyses. Since our goal is to optimize second-line treatments, we included patients who had a first-line treatment and switched to a second-line treatment of SFU or insulin during 2008 and 2012. Each subject's observations are re-aligned at the time of switching (index time 0). A median baseline period of 12-months prior to the index time was established to extract baseline feature variables. The median follow-up period post index time to evaluate treatment response is 18-months. The primary outcome is the number of major complications of T2D after second-line treatment including essential hypertension, hyperlipidemia and hypercholesterolemia.

Unlike randomized experiment, EHR data are collected in an uncontrolled setting and pose challenges for inferring causal relationships in research studies. In our analyses, we address three major issues commonly confronting observational study research including confounding bias, selection bias and missing data. To alleviate confounding bias, we implement matching step in M-learning on potential confounders (e.g. lab test features), two propensity scores constructed by

two logistic regression models for lab measurement pattern features and demographics predictors and a doubly-robust prognostic score estimated from a linear regression model. To reduce selection bias due to missing post second-line treatment outcomes, IPW adjustment was implemented and a logistic regression model was used to predict the presence of at least one post-treatment measure during the follow up period. Lastly, missing features were imputed with chained equations.

4.2.3 LDA-Based Features for Co-medication and Diagnosis

In this chapter, we extracted clinically meaningful features from two EHR data domains (i.e. medications and ICD diagnosis codes) using latent Dirichlet allocation model (LDA) as a topic model based approach. LDA is a three-level hierarchical generative model suitable for analyzing collections of discrete data and has broad applications from information retrieval to medical informatics (Blei et al., 2003). When applied to EHRs, it can organize a large number of ICD diagnoses codes, procedure codes and medication prescription documentations into meaningful topics based on the correlation among variables. As a generative model, LDA is parameterized by distributions over topics, where each topic is characterized by a distribution over words in that topic. Thus, LDA models are highly interpretable.

In our EHR analyses, each patient can be viewed as a "document" in topic modeling applications. LDA model is adapted here to learn the hidden topics in the domain information for the collection of patients. In the medication prescription domain, the medication items in the generic name field except the T2D treatments are the "words" in a "patient's document". The latent topics learned in this domain are referred to as co-medication topics. Similarly, the ICD9 codes in the diagnosis domain are useful terms to learn topics of patient comorbidities. The number of topics is fixed for the two domains and the latent topics were learned separately.

Specifically, the key idea of using LDA to analyze medication prescriptions and ICD codes in EHRs is that the patient population (collection of "documents") can be represented as mixture of multiple latent subgroups or topics and each topic is defined to be a multinomial distribution over a collection of "vocabulary" of medications or conditions (i.e., terms) (Blei et al., 2003; Blei and Lafferty, 2009). LDA assumes a probabilistic generative process by which each patient is generated with K latent topics and each topic is assumed to follow a Dirichlet distribution over V medications or conditions in the "vocabulary". We describe the process as follows (Blei et al., 2002). Let $\mathcal{C}$ denote a corpus composed of M patients represented as $\mathcal{C} = \{c_1, c_2, \cdots, c_M\}$.

1. For a patient of N medications or conditions, first a distribution over topics is drawn from

$$\boldsymbol{\theta} \sim \text{Dir}(\boldsymbol{\alpha}),$$

where the probability density of a k-dimensional Dirichlet variable $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1},$$

$\Gamma(x)$ is the Gamma function, $0 < \alpha_i < 1, \forall i$ and $\sum_{i=1}^{k} \theta_i = 1$.

2. For each medication or condition of the patient, a topic $x \in \{1, \cdots, K\}$ is drawn from

$$x \sim \text{Mult}(\boldsymbol{\theta}),$$

where $\text{Mult}(\boldsymbol{\theta})$ is a multinomial distribution characterized by $\boldsymbol{\theta}$ and $p(x = i|\boldsymbol{\theta}) = \theta_i$.

3. Given the topic, medications or conditions of the patient are drawn from a conditional multinomial distribution $p(w|x; \boldsymbol{\phi})$, where $\boldsymbol{\phi}$ is the parameter of the per-topic distribution over medications or conditions.

Several Bayesian inference techniques can be used to approximate the posterior distribution

$$p(\boldsymbol{\theta}, x|\boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\phi}) = \frac{p(\boldsymbol{\theta}, x, \boldsymbol{w}|\boldsymbol{\alpha}, \boldsymbol{\phi})}{p(\boldsymbol{w}|\boldsymbol{\alpha}, \boldsymbol{\phi})}$$

in LDA including variational inference, expectation propagation and Gibbs sampling (Blei et al., 2003; Geman and Geman, 1984; Minka and Lafferty, 2002). We can obtain the marginal distribution of a patient as

$$p(\boldsymbol{w}|\boldsymbol{\alpha}, \boldsymbol{\phi}) = \int_{\boldsymbol{\theta}} (\prod_{n=1}^{N} \sum_{x_n=1}^{K} p(w_n|x_n; \boldsymbol{\phi}) p(x_n|\boldsymbol{\theta})) p(\boldsymbol{\theta}; \boldsymbol{\alpha}) d\boldsymbol{\theta}$$

and take the product of the marginal probabilities of patients to derive the probability of the corpus (Blei et al., 2003). In LDA, the Dirichlet is drawn for each patient, and within this patient the multinomial topic distribution over medication or conditions is drawn repeatedly (Blei et al., 2002).

The weights of posterior probabilities or proportions for different topics given patients' observed medications and conditions obtained from Bayesian inference are the feature representation in a reduced dimension (dimension equals to the number of topics). In our application, both clinical interpretability and topic coherence measure were taken into consideration and 5 topics were used. The LDA-based features were constructed based on these weights and they were augmented to the feature space in our analyses in learning the personalized treatments and confounding reduction. We compared using LDA-based features with using a simple summary feature of a total number of medications or ICD diagnoses.

4.3 Applications

In this section, we introduce our cohort identification and EHR preprocessing procedures. LDA model is used in representing the variables in two data domains in the EHR data in a lower dimension. The learned topics from the LDA model are visualized and association networks are

built to better interpret the topics. The features we extracted from the latent topics were augmented to the feature space in machine learning methods in estimating the ITRs.

4.3.1 Cohort Identification

Our cohort includes patients who were older than 18 years and had at least one T2D diagnosis between 1/1/2008 and 12/31/2012. The analysis was further restricted to those patients who were prescribed with Metformin as first-line medication (Wu et al., 2019). There were over 1,200 patients who were augmented with at least one second-line treatment. The sunburst plot in Figure 4.1 displays treatment sequence with 2-3 stages for T2D patients in the database. The most inner circle corresponds to the first-line treatment and the most commonly used first stage treatment is Metformin, which is consistent with the T2D treatment guideline. Our goal is to select optimal ITRs to reduce major complications. The ITRs were estimated in terms of lowering major complications of T2D measured by three ICD diagnosis counts as ordinal outcomes (0, 1, 2, 3).

The EHR data preprocessing flowchart is shown in Figure C.1. Most features were extracted from four domains: demographics, medication prescriptions, lab test and diagnosis. Several important lab tests (e.g. HbA1c, glucose, HDL, LDL, BMI) were summarized by estimated initial test values and estimated rate of change using linear mixed effect models. Initial test value is a measure of patient's illness condition at the baseline of study and rate of change of test value characterize patient's progression rate of disease. Lab measurement pattern features were constructed using the information in the lab test domain (Wu et al., 2019). Lab measurement patterns in EHRs were shown to reflect patient's underlying health status as well as EHR documentation biases and thus informative of adjusting for confounding and matching patients (Pivovarov et al., 2014). In Figure C.6, the measurement patterns of glucose test and HbA1c in terms of gap days in logarithm

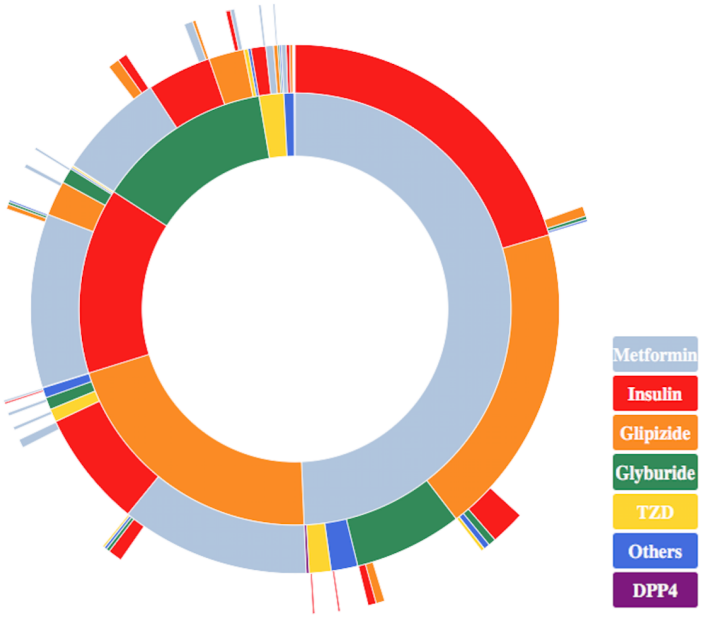


Figure 4.1: Sunburst plot of treatment sequence for T2D patients*

*: Layer corresponds to stage of treatment, patients with 2-3 stages were shown

scale is different for patients on different second-line treatments, suggesting it can distinguish between patients who started insulin (MET+insulin) and those who received glyburide or glipizide (MET+SFU). In Figure C.7, the measurement patterns of one of clinically important lab measures HbA1c and medications in terms of gap days in logarithm scale are displayed for different treatment strategies. The difference in these figures further demonstrates the usefulness of these measurement patterns in personalized treatment. Therefore the lab measurement pattern features were used as important predictors for computing a balancing propensity score (Wu et al., 2019) in M-learning. The features capturing temporal measurement patterns were created by discretizing a two-dimensional space of lab test values versus gap time between consecutive measurements according to high, medium, low quartiles (Wang et al., 2016). Additionally, age, race and gender were also included as demographic variables for matching patients. Spikeplots for important lab tests

such as glucose and HbA1c and medications among sample patients are shown in Appendix C.2. These figures capture patient heterogeneity in terms of measurement pattern of lab and medication.

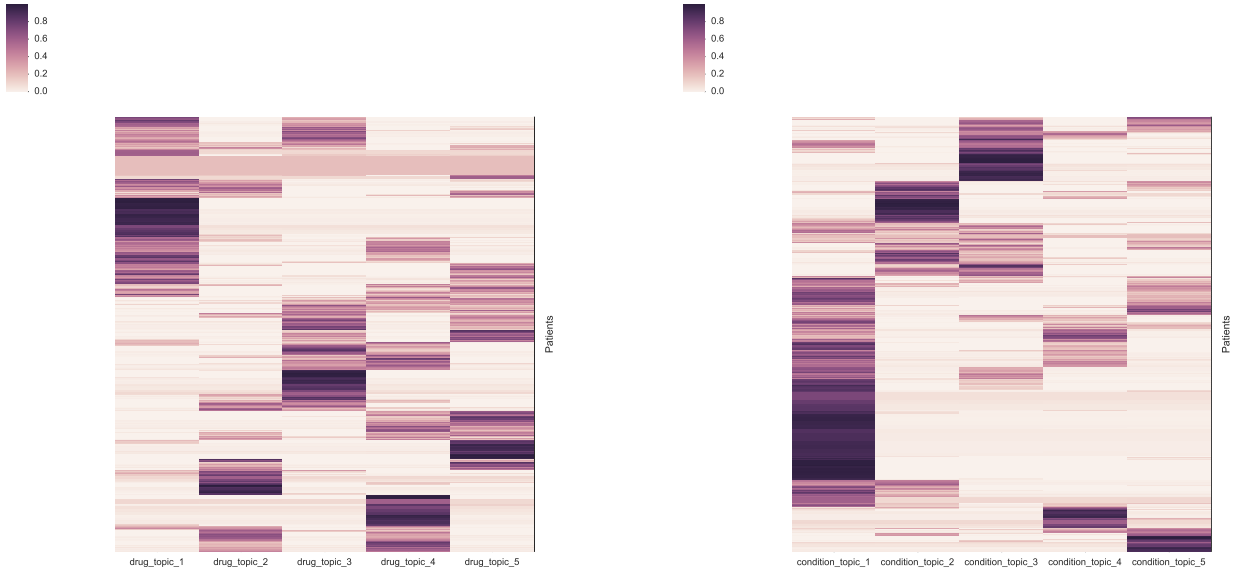
4.3.2 LDA Feature Representation

Topic model with LDA using bag of words is applied to the generic names in co-medications and condition ICD9 codes prior to second-line treatment initiation (during baseline period) with 5 latent topics respectively. There were over 700 generic medication names and over $2,800$ unique ICD9 codes in this T2D EHR data. In the co-medication domain, bigram and trigram were created for compound words in generic names. The number of topics is fixed in our application for interpretability purposes. LDA-based features were constructed as weights for each of the 5 topics per patient. We use heatmaps with hierarchical clustering to show topic feature weights in all patients. The similarity metric we used to cluster patients is Hellinger distance (Blei and Lafferty, 2009) defined as

$$\sum_{k=1}^{K}(\sqrt{\hat{\theta}_{d_1,k}} - \sqrt{\hat{\theta}_{d_2,k}})^2,$$

where K is the fixed number of topics and $\theta_{d_i,k}$ denotes weight for topic k in each patient. In Figure 4.2a, the heatmap for co-medication topic features suggests that most patients have higher weights in only one or two topics. Thus, co-medications for most patients comprise of one or two latent topics. Figure 4.2b shows patterns of ICD diagnosis condition, and a relatively high proportion of patients have larger weights in the first condition topic, which mainly consists of conditions such as hypertension, pure hypercholesterolemia, depressive disorder, hyperlipidemia and lumbago.

To understand the learned topics, we present visualization of some of the medication and condition topics in Figure C.3, C.3 and C.3. These topics were found to be informative of predicting optimal treatments in our subsequent analyses (next section). On the right panel of each sub-figure,



(a) Co-medication Topic Features

(b) Condition Topic Features

Figure 4.2: Heatmap of LDA-Features Clustered by Patients

the top 30 most relevant terms (co-medications or conditions) corresponding to the topics are presented. For example, topic #3 in the co-medication domain is a cluster of medications used to treat heart disease, high blood pressure and high cholesterol. Figure C.3 shows that this topic contains 5 most relevant prescriptions as aspirin, metoprolol, atorvastatin, enalapril and clopidogrel. In condition topic #2, the most relevant conditions include hypertension, hypercholesterolem, pain in limb, atrial fibrillation and hyperlipidemia, which are common co-morbidities of T2D. In condition topic #5, the top most relevant conditions are benign hypertension, hyperlipidemia, asthma and obesity.

Based on topics learned from LDA models, we constructed association network for medications and conditions separately in Figure 4.3 and 4.4. The medications or conditions (nodes in the network) in the same topic with a posterior probability of co-occurring higher than a threshold are linked by an edge and shown in the figures. The threshold is 1.5% for co-medications and 0.9%

for conditions for visualization. The edge widths are proportional to posterior probabilities. In the condition network, mental disorders including anxiety, depressive disorder and schizoaffective as well as some side effects such as cough, headache and backache are clustered in the rightmost group. Cardiovascular system related disorders conditions such as old myocardial infarction, mixed hyperlipidemia, atrial fibrillation and coronary atherosclerosis are connected together in the bottom cluster. Several topic clusters are connected through "bridge conditions/procedures" such as hypertension, unspecified hyperlipidemia, and flu vaccination which are associated with conditions in other clusters. In the co-medications network, several anti-diabetic or heart disease medications are clustered together including Rosuvastatin, Pioglitazone, Warfarin and Sitagliptin. Drugs used to treat co-morbidities are connected in a group such as Levothyroxine (treat thyroid disease), Esomeprazole (treat gastroesophageal reflux disease) and Atenolol (treat high blood pressure and chest pain). In this co-medications network for T2D, aspirin, as a blood thinner and anti-inflammatory drug, plays an important role in bridging medications from several clusters since aspirin is often co-prescribed with other medications.

4.3.3 Learning Optimal ITR

In our final analysis cohort, there were 740 subjects in total. The cohort was further divided into high baseline HbA1c level group and low baseline HbA1c level group according to median initial HbA1c level of 8.5% to account for patient heterogeneity. There were 380 patients in the low baseline group among whom 240 (63%) was prescribed SFU as second-line medication and 140 (37%) was prescribed insulin. In the high baseline group, the number of patients received SFU and insulin were 208 (58%) and 152 (42%) respectively. The outcome R_i used to minimize the value function (4.4) is count of post-treatment major complications (essential hypertension,

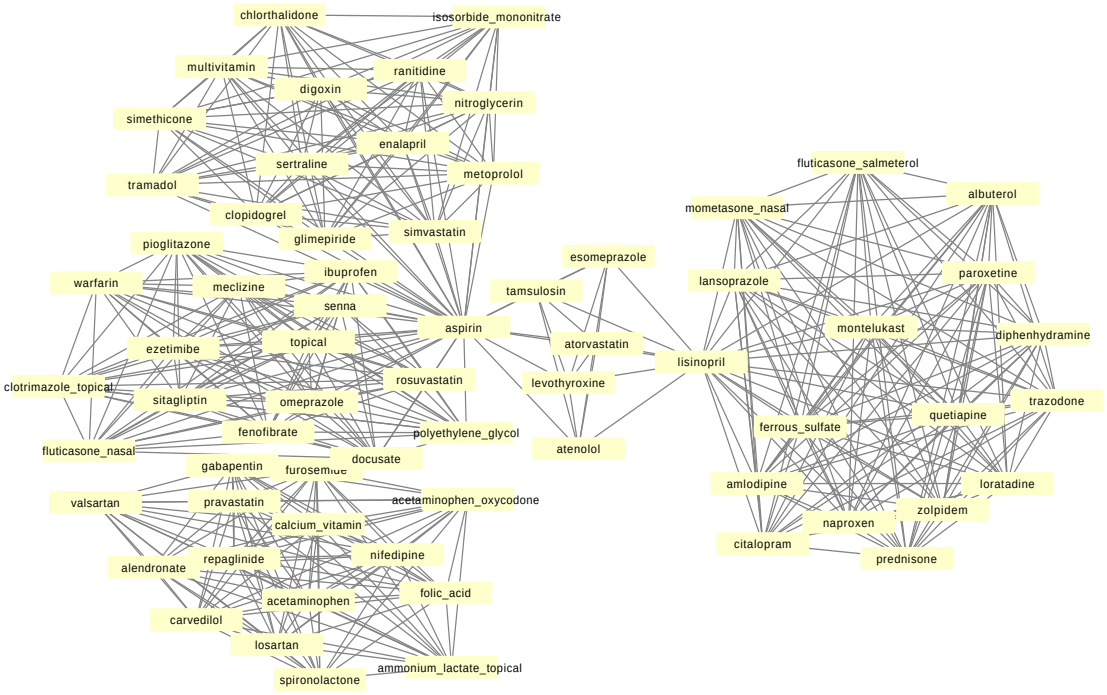


Figure 4.3: Association Network of Medication Prescriptions Based on Topics in LDA Model

hyperlipidemia and hypercholesterolemia) during the follow up period. We performed 2-fold cross-validation 100 times, and computed the cross-validated empirical value function under three learning methods (Q-learning, O-learning and M-learning).

In Table 4.1, we present the previous results in (Wu et al., 2019) with the raw counts from co-medication and diagnosis domain during the baseline period included as feature variables instead of using LDA-based features. In Table 4.2, we replaced the raw co-medication count and condition ICD9 codes count with our extracted LDA-based topic features. Comparing this to the results excluding LDA-based features, all the three learning methods in both groups improve with lower values of post-treatment major complications counts. In the two groups, most of the methods result in a smaller variability. In the high baseline group (see Figure 4.5a), the worst performance learning method using raw count data is Q-learning. It benefited most from including LDA-based

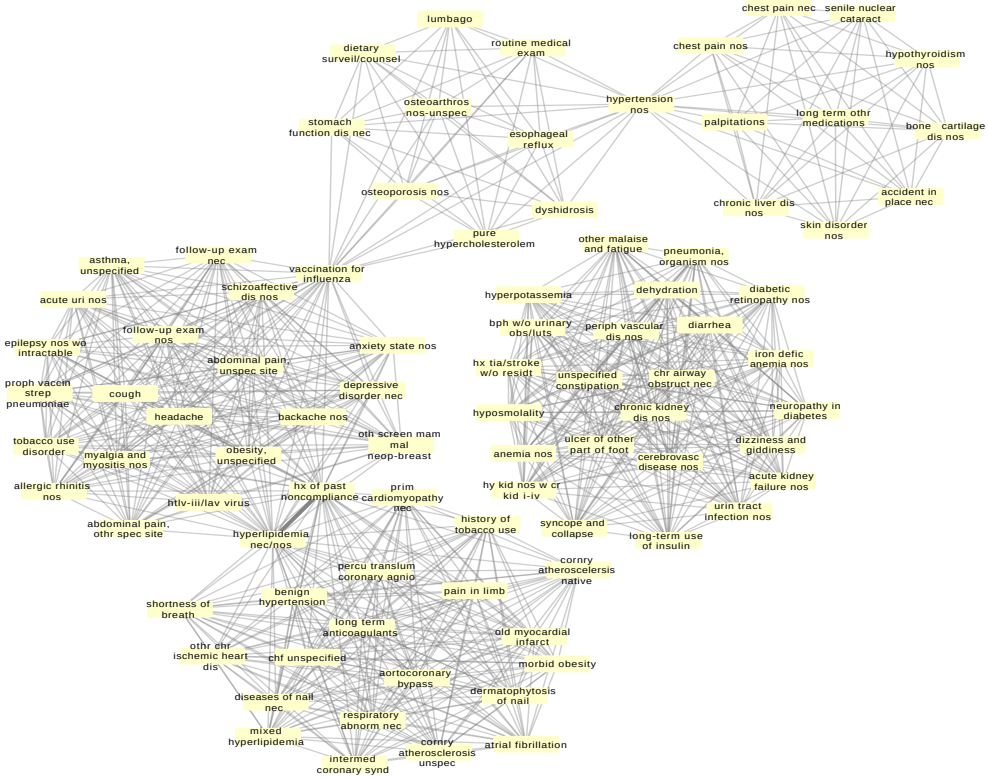
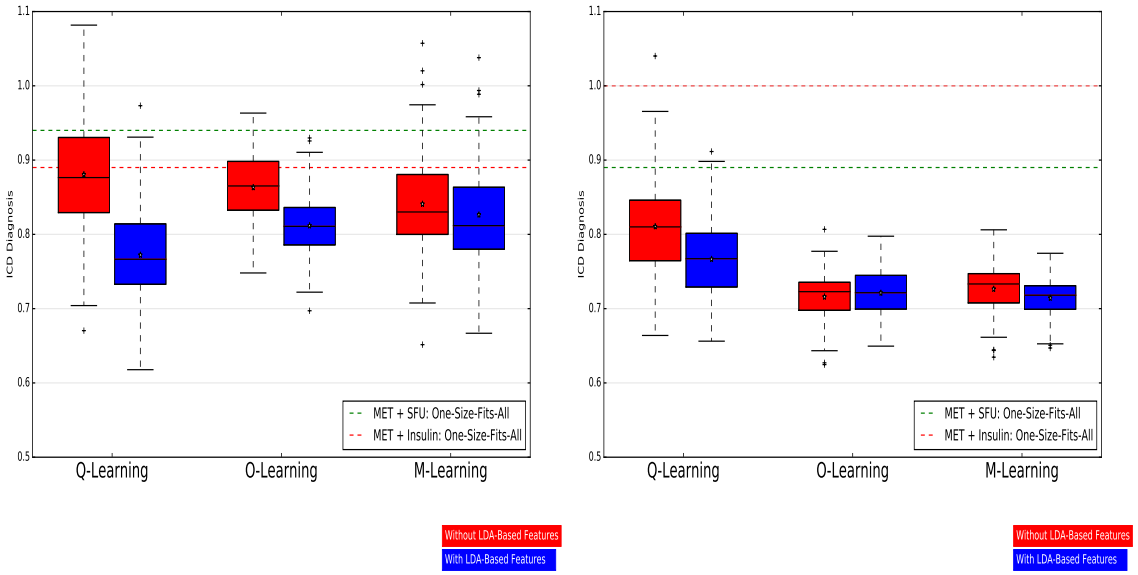


Figure 4.4: Association Network of ICD9 Conditions Based on Topics in LDA Model

features. In the low baseline group (see Figure 4.5b), M-learning using LDA-based features achieves the lowest value function with a mean of 0.71 and the standard deviation of 0.028. In Figure 4.5, we note that all three ITR learning methods outperform the "one-size-fits-all" strategies to assign insulin to all subjects or SFU to all. This analysis demonstrates that prescribing treatment based on each patient's characteristics reduces the number of major complications compared to universal treatment prescription.

We identify the important topic features for selecting the optimal treatment by implementing M-learning with linear kernel SVM including LDA-based features in the cohort of 740 patients. There were 427 (58%) patients predicted to have "MET + SFU" as the optimal second-line medication



(a) High Baseline Group

(b) Low Baseline Group

Figure 4.5: Empirical value function of ICD diagnosis count in EHR data with 100 2-fold cross-validations (a low value is desirable)

Table 4.1: Cross-validated Empirical Value Function for the Number of Major Complications Using Raw Count Data

High Baseline Group		
ITR Method	Mean (Std)	Median (Q1, Q3)
Q-Learning	0.88 (0.078)	0.88 (0.83, 0.93)
O-Learning	0.86 (0.050)	0.87 (0.83, 0.90)
M-Learning	0.84 (0.068)	0.83 (0.80, 0.88)
Low Baseline Group		
ITR Method	Mean (Std)	Median (Q1, Q3)
Q-Learning	0.81 (0.063)	0.81 (0.76, 0.85)
O-Learning	0.72 (0.033)	0.72 (0.70, 0.74)
M-Learning	0.73 (0.032)	0.73 (0.71, 0.75)

Table 4.2: Cross-validated Empirical Value Function for the Number of Major Complications Using LDA Features

High Baseline Group		
ITR Method	Mean (Std)	Median (Q1, Q3)
Q-Learning	0.77 (0.065)	0.77 (0.73, 0.81)
O-Learning	0.81 (0.045)	0.81 (0.79, 0.84)
M-Learning	0.83 (0.088)	0.81 (0.78, 0.86)
Low Baseline Group		
ITR Method	Mean (Std)	Median (Q1, Q3)
Q-Learning	0.77 (0.048)	0.77 (0.73, 0.80)
O-Learning	0.72 (0.029)	0.72 (0.70, 0.74)
M-Learning	0.71 (0.028)	0.72 (0.70, 0.73)

and 313 (42%) to have "MET + insulin". We show the absolute value of all the standardized coefficients in Figure 4.6. Besides race and cluster membership variable (based on hierarchical clustering analysis for patient heterogeneity using a subset of standardized features), co-medication topic #3, condition topic #5 and #2 rank as the most important topic features. The most relevant terms in these three topics are presented in Appendix C.3. Medication topic #3 involves past prescriptions of drugs used to treat heart diseases and its risk factors, while condition topic #5 and #2 includes major T2D co-morbidities. These results suggest that to recommend optimal second-line treatments, physicians may focus on past heart disease mediation prescriptions and past diagnostic history of hypertension, hypercholesterolem, and atrial fibrillation.

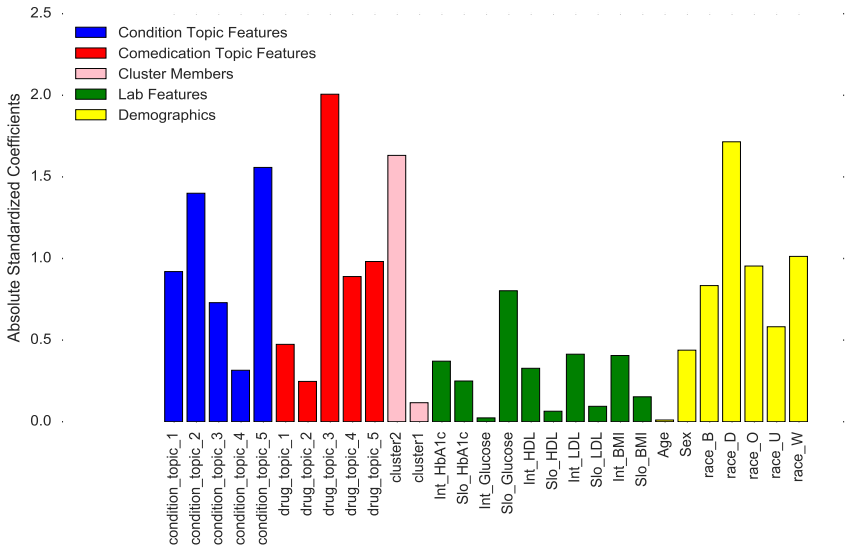


Figure 4.6: Feature Importance in M-learning with a Linear Kernel

4.4 Discussion

In this chapter, we investigate topic modeling with LDA as a feature extraction tool for EHR data and the use of a matching-based machine learning method for individualized treatment optimization. The learned features from LDA can be used in the ITR learning step to improve interpretability of the ITRs. Compared to treatment rules estimated excluding LDA-based features, our method achieves better post treatment clinical outcome and a smaller variability. In addition, LDA-based features summarize a large number of ICD codes and medications into a few understandable topics. By examining important topics predictive of optimal treatment, our method provides an intuitive summary of main useful characteristics a physician needs to focus on when making personalized medical decisions.

Chapter 5

Conclusions and Future Direction

5.1 Summary

The overall theme of this work is merging statistical modeling and feature engineering techniques with machine learning algorithms to assist in personalized medical decision making, leveraging the information from electronic health records.

Chapter 2 considers a matching-based learning method, referred to as M-learning, to learn personalized treatment rule from observational data. We show its advantage over existing methods in several scenarios through extensive simulation studies and a real data application using EHR from diabetes patients. In Chapter 3, we propose a novel framework using a few learning methods including an extended version of M-learning to transfer information learned from real-world observational data to RCT. Two super features are used in this framework. We show the efficacy of this framework through simulation studies and a data application using both EHR and RCT data of diabetes study. In Chapter 4, a topic model based method is considered to extract interpretable features from different domains of EHR data and use learning methods to estimate ITRs. We eval-

uate the benefit gained from this method by comparing to previous methods without augmenting the topic model based features using EHR from diabetes patients.

5.2 Limitations

In M-learning proposed in Chapter 2, one-nearest-neighbor matching might result in many repeated observations in the learning step and only a limited number of pairs of subjects are included. An extension of M-learning, which is proposed in Chapter 3, resolves this by integrating with kernel method using a kernel weighted objective function for optimizing. However, there is a tradeoff between accuracy and computational speed. The computation cost of implementing kernel M-learning is significantly higher without resort to more efficient algorithms or more powerful computational resources. Another limitation of the transfer learning in Chapter 3 is that when super features are less informative of optimal treatment or benefit, stratification in domain adaptation may not work. In Chapter 4, there is no consensus on optimal measure in selecting number of topics in LDA model and in our case, a fixed number is used based on clinical interpretability and topic coherence. The original topic model does not consider correlation between latent topics and the trend in topics change over time. Lastly, we use EHR data from NYPH clinical data warehouse and do not evaluate the performance of our proposed methods and framework in other databases. We believe that investigating performance using EHR data from other hospitals will be helpful, especially in transfer learning framework.

5.3 Extensions

In Chapter 2, we focus on single-state M-learning and it can be generalize to multi-stage applying the backward learning techniques. In addition, more general multi-arm treatment options can be considered. To improve the performance, data-driven oriented matching functions and suitable matching variables can be selected when applying our method to satisfy both clinical interpretability and precision of ITR estimation. In Chapter 3, our proposed super features and framework is not difficult to implement. However, there is certain computational burden in searching the optimal parameter space when running the algorithms. In the future direction, more computational efficient algorithms should be applied. In particular, a mini-batch version of weighted SVM can be implemented in our kernel M-learning. Additionally, instead of the original super features, other features predictive of treatment options in observational studies is worthwhile to be examined. In Chapter 4, several extensions can be considered. It is desirable to consider a data-driven method to select the best number of topics in LDA-based feature extraction. It is of interest to investigate a hierarchical, comprehensive LDA-based feature extraction combining different domains including medications, conditions and procedures altogether instead of modeling them separately. Moreover, correlated topic models can be considered to model the correlation between topics within each medical domain and dynamic topic models might be useful in capturing the temporal characteristics of the evolution of medical concepts. Hierarchical topic models can also be used to incorporate existing hierarchical structure of medications or diagnosis into learning topics. With these specific topic models, more meaningful feature representation can be achieved. Additionally, we can extend nearest-neighbor matching in M-learning to other optimal matching techniques such as integrating with kernel method. These extensions will provide more flexibility

to M-learning, especially for observational studies. Lastly, it is worthwhile to extend our methods and analyses to OHDSI network by following the format of OMOP Common Data Model (CDM) (https://www.ohdsi.org/data-standardization/the-common-data-model/).

Bibliography

Allwein, E. L., Schapire, R. E., and Singer, Y. (2001). Reducing multiclass to binary: A unifying approach for margin classifiers. *The Journal of Machine Learning Research* **1,** 113–141.

American Diabetes Association (2014). Standards of medical care in diabetes—2014. *Diabetes Care* **37,** S14–S80.

Antonelli, J., Cefalu, M., Palmer, N., and Agniel, D. (2018). Doubly robust matching estimators for high dimensional confounding adjustment. *Biometrics* **74,** 1171–1179.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research* **46,** 399–424.

Austin, P. C. and Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine* **34,** 3661–3679.

Bareinboim, E. and Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* **113,** 7345–7352.

Bennett, W. L., Maruthur, N. M., Singh, S., Segal, J. B., Wilson, L. M., Chatterjee, R., Marinopoulos, S. S., Puhan, M. A., Ranasinghe, P., Block, L., et al. (2011). Comparative effectiveness and

safety of medications for type 2 diabetes: an update including new drugs and 2-drug combinations. *Annals of internal medicine* **154,** 602–613.

Bianchi, C. and Del Prato, S. (2011). Metabolic memory and individual treatment aims in type 2 diabetes – outcome-lessons learned from large clinical trials. *The Review of Diabetic Studies* **8,** 432–440.

Blei, D. M. and Lafferty, J. D. (2009). *Topic Models*. CRC Press.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2002). Latent dirichlet allocation. In Dietterich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*, pages 601–608. MIT Press.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.* **3,** 993–1022.

Boriah, S., Chandola, V., and Kumar, V. (2008). Similarity measures for categorical data: A comparative evaluation. In *Proceedings of the 2008 SIAM International Conference on Data Mining*, pages 243–254. SIAM.

Buuren, S. and Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software* **45,**.

Chakraborty, B. and Moodie, E. E. (2013). *Statistical methods for dynamic treatment regimes*. New York: Springer-Verlag.

Cole, S. R. and Stuart, E. A. (2010). Generalizing evidence from randomized clinical trials to target populations: The actg 320 trial. *American journal of epidemiology* **172,** 107–115.

Collins, F. S. and Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine* **372,** 793–795.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96,** 187–199.

Dehejia, R. H. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association* **94,** 1053–1062.

Devroye, L., Györfi, L., and Lugosi, G. (2013). *A probabilistic theory of pattern recognition*, volume 31. New York: Springer-Verlag.

Diabetes Control and Complications Trial Research Group (1993). The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New England Journal of Medicine* **329,** 977–986.

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

Ellis, A. R., Dusetzina, S. B., Hansen, R. A., Gaynes, B. N., Farley, J. F., and Stürmer, T. (2013). Investigating differences in treatment effect estimates between propensity score matching and weighting: a demonstration using star* d trial data. *Pharmacoepidemiology and drug safety* **22,** 138–144.

Fahrbach, J., Jacober, S., Jiang, H., and Martin, S. (2008). The durable trial study design: Comparing the safety, efficacy, and durability of insulin glargine to insulin lispro mix 75/25 added to oral antihyperglycemic agents in patients with type 2 diabetes. *Journal of Diabetes Science and Technology* **2,** 831–838. PMID: 19885269.

Fort, D., Weng, C., Bakken, S., and Wilcox, A. B. (2014). Considerations for using research data to verify clinical data accuracy. *AMIA Summits on Translational Science Proceedings* pages 2014: 211–217.

Fraser, D., Christiansen, B. A., Adsit, R., Baker, T. B., and Fiore, M. C. (2012). Electronic health records as a tool for recruitment of participants' clinical effectiveness research: lessons learned from tobacco cessation. *Translational behavioral medicine* **3,** 244–252.

Fried, E. I. and Nesse, R. M. (2015). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the star* d study. *Journal of affective disorders* **172,** 96–102.

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,** 3285–3302.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6,** 721–741.

Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W. A., Li, R., Manolio, T. A., Sanderson, S. C., Kannry, J., Zinberg, R., Basford, M. A., and etc (2013). The electronic medical records and genomics (eMERGE) network: past, present, and future. *Genetics in Medicine* **15,** 761–771.

Hamburg, M. A. and Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine* **363,** 301–304.

Haneuse, S. (2016). Distinguishing selection bias and confounding bias in comparative effectiveness research. *Medical Care* **54,** e23–e29.

Haneuse, S. and Daniels, M. (2016). A general framework for considering selection bias in ehr-based studies: What data are observed and why? *eGEMs* **4,**.

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association* **99,** 609–618.

Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika* pages 481–488.

Haynes, B. (1999). Can it work? does it work? is it worth it?: The testing of healthcare interventions is evolving. *BMJ: British Medical Journal* **319,** 652.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis* **15,** 199–236.

Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, UAI'99, pages 289–296, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Hogan, W. R. and Wagner, M. M. (1997). Accuracy of data in computer-based patient records. *Journal of the American Medical Informatics Association* **4,** 342–355.

Hripcsak, G. and Albers, D. J. (2013). Next-generation phenotyping of electronic health records. *Journal of the American Medical Informatics Association* **20,** 117–121.

Hripcsak, G., Albers, D. J., and Perotte, A. (2011). Exploiting time in electronic health record correlations. *Journal of the American Medical Informatics Association* **18,** i109–i115.

Hripcsak, G., Ryan, P. B., Duke, J. D., Shah, N. H., Park, R. W., Huser, V., Suchard, M. A., Schuemie, M. J., DeFalco, F. J., Perotte, A., and etc (2016). Characterizing treatment pathways

at scale using the OHDSI network. *Proceedings of the National Academy of Sciences* **113,** 7329–7336.

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

Jiang, H. (2017). Rates of uniform consistency for k-NN regression. *arXiv preprint arXiv:1707.06261* .

Johnson, S. B. (1996). Generic data modeling for clinical rrepositories. *Journal of the American Medical Informatics Association* **3,** 328–339.

Kang, C., Janes, H., and Huang, Y. (2014). Combining biomarkers to optimize patient treatment recommendations. *Biometrics* **70,** 695–707.

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

Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes* **25,** 259–284.

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

Lee, B. K., Lessler, J., and Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine* **29,** 337–346.

Lee, B. K., Lessler, J., and Stuart, E. A. (2011). Weight trimming and propensity score weighting. *PloS One* **6,** e18174.

Little, R. J. and Rubin, D. B. (2014). *Statistical analysis with missing data.* John Wiley & Sons.

Liu, Y., Wang, Y., Huang, C., and Zeng, D. (2017). Estimating personalized diagnostic rules depending on individualized characteristics. *Statistics in Medicine* **36,** 1099–1117.

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

Minka, T. and Lafferty, J. (2002). Expectation-propagation for the generative aspect model. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, UAI'02, pages 352–359, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

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

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

Murphy, S. A., Oslin, D. W., Rush, A. J., and Zhu, J. (2007). Methodological challenges in constructing effective treatment sequences for chronic psychiatric disorders. *Neuropsychopharmacology* **32,** 257–262.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. on Knowl. and Data Eng.* **22,** 1345–1359.

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

Pivovarov, R., Albers, D. J., Sepulveda, J. L., and Elhadad, N. (2014). Identifying and mitigating biases in ehr laboratory tests. *Journal of biomedical informatics* **51,** 24–34.

Plank, J., Wutte, A., Brunner, G., Siebenhofer, A., Semlitsch, B., Sommer, R., Hirschberger, S., and Pieber, T. R. (2002). A direct comparison of insulin aspart and insulin lispro in patients with type 1 diabetes. *Diabetes Care* **25,** 2053–2057.

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

Raja-Khan, N. T., Warehime, S. S., and Gabbay, R. A. (2007). Review of biphasic insulin aspart in the treatment of type 1 and 2 diabetes. *Vascular Health and Risk Management* **3,** 919 – 935.

Ray, W. A. (2003). Evaluating medication effects outside of clinical trials: new-user designs. *American Journal of Epidemiology* **158,** 915–920.

Rosenbaum, P. R. (2010). *Design of observational studies*, volume 10. Springer.

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

Rubin, D. B. (2004). On principles for modeling propensity scores in medical research. *Pharmacoepidemiology and drug safety* **13,** 855–857.

Rubin, D. B., Stern, H. S., and Vehovar, V. (1995). Handling "don't know" survey responses: the case of the slovenian plebiscite. *Journal of the American Statistical Association* **90,** 822–828.

Sekhon, J. S. and Grieve, R. D. (2012). A matching method for improving covariate balance in cost-effectiveness analyses. *Health Economics* **21,** 695–714.

Smith, H. L. (1997). Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology* **27,** 325–353.

Steinwart, I. and Christmann, A. (2008). *Support vector machines*. New York: Springer-Verlag.

Steinwart, I. and Scovel, C. (2007). Fast rates for support vector machines using gaussian kernels. *The Annals of Statistics* pages 575–607.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science* **25,** 1–21.

Stuart, E. A. and Green, K. M. (2008). Using full matching to estimate causal effects in non-experimental studies: examining the relationship between adolescent marijuana use and adult outcomes. *Developmental psychology* **44,** 395.

Tao, C., Parker, C. G., Oniki, T. A., Pathak, J., Huff, S. M., and Chute, C. G. (2011). An owl meta-ontology for representing the clinical element model. In *AMIA Annu Symp Proc*, pages 1372–81. Citeseer.

Tian, L., Alizadeh, A. A., Gentles, A. J., and Tibshirani, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association* **109,** 1517–1532. PMID: 25729117.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer.

Wallace, M., Moodie, E., and Stephens, D. (2016). Comment on "personalized dose finding using outcome weighted learning". *Journal of the American Statistical Association* **111,** 1530–1534.

Wang, Y., Fu, H., and Zeng, D. (2018). Learning optimal personalized treatment rules in consideration of benefit and risk: with an application to treating type 2 diabetes patients with insulin therapies. *Journal of the American Statistical Association* **113,** 1–13.

Wang, Y., Wu, P., Liu, Y., Weng, C., and Zeng, D. (2016). Learning optimal individualized treatment rules from electronic health record data. In *Healthcare Informatics (ICHI), 2016 IEEE International Conference on*, pages 65–71. IEEE.

Watkins, C. J. and Dayan, P. (1992). Q-learning. *Machine learning* **8,** 279–292.

Watkins, C. J. C. H. (1989). *Learning from Delayed Rewards*. PhD thesis, King's College, Cambridge, UK.

Weiskopf, N. G., Hripcsak, G., Swaminathan, S., and Weng, C. (2013). Defining and measuring completeness of electronic health records for secondary use. *Journal of biomedical informatics* **46,** 830–836.

Weiskopf, N. G. and Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association* **20,** 144–151.

Wells, B. J., Nowacki, A. S., Chagin, K., and Kattan, M. W. (2013). Strategies for handling missing data in electronic health record derived data. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* **1,** 7.

Weng, C., Li, Y., Ryan, P., Zhang, Y., Liu, F., Gao, J., Bigger, J., and Hripcsak, G. (2014). A distribution-based method for assessing the differences between clinical trial target populations and patient populations in electronic health records. *Applied clinical informatics* **5,** 463.

Wu, P., Zeng, D., and Wang, Y. (2019). Matched learning for optimizing individualized treatment strategies using electronic health records. *Journal of the American Statistical Association* **0,** 1–23.

Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M., and Laber, E. (2012). Estimating optimal treatment regimes from a classification perspective. *Stat* **1,** 103–114.

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,** 681–694.

Zhang, K., Schölkopf, B., Muandet, K., and Wang, Z. (2013). Domain adaptation under target and conditional shift. In *Proceedings of the 30th International Conference on Machine Learning*. JMLR.

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,** 1106–1118.

Zhou, X., Wang, Y., and Zeng, D. (2018). Outcome-weighted learning for personalized medicine with multiple treatment options. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 565–574. IEEE.

Appendix A

Appendices to Chapter 2

A.1 Additional Simulations Evaluating M-learning for Discrete Outcomes

In this part, ordinal outcomes were generated by discretizing a latent continuous outcome generated similarly to those in Chapter 2.4. The underlying continuous outcomes were simulated as

$$S_1 : R = 2H_3 - H_4 + A(H_1 - H_2) + N(0,1)$$

and

$$S_2 : R = 1 + 2H_1 + H_2 + 0.5H_3 + A(H_2 + H_1^2 - 1) + N(0,1).$$

Furthermore, R was discretized into an ordinal clinical response outcome with 4 categories based on quantiles. In section A3, boxplots comparing value functions of M-learning, AOL and Q-learning under different sample sizes can be found. The true optimal value is 1.11 in S_1 and 2.23 in S_2.

In the first set of analyses, the distribution of A depended on H and no unmeasured confounder is present. Propensity model was specified by $P(A = 1|H) = \text{expit}(1 + 2H_1 + H_2)$ and a logistic

regression with H_1 and H_2 as linear covariates was used to estimate propensity scores. In the linear boundary setting, Q-learning ranks the best and M-learning is slightly worse than AOL. In S_2, M-learning is more robust to outliers and hence performs much better than AOL. The difference of value function between using $g(x) = x$ or $g(x) = 1$ is negligible.

In the second set of analyses, we specified propensity score model as $P(A = 1|H) = \text{expit}(1 + \exp(H_2))$. A logistic regression with H_1 and H_2 as linear covariates was used to estimate propensity scores and thus the model is misspecified. In S_1, AOL is better than M-learning in terms of value function and Q-learning still has a higher value. While in S_2, AOL performs much worse than M-learning since AOL is more sensitive to incorrectly specified propensity score model.

In the third set of analyses, we considered presence of unmeasured confounders. The clinical outcomes were simulated as

$$S_3 : R = 2H_3 - H_4 + A(H_1 - H_2 + X) + N(0, 1)$$

and

$$S_4 : R = 1 + 2H_1 + H_2 + 0.5H_3 + A(H_2 + H_1^2 + X - 1) + N(0, 1)$$

where $P(A = 1|H, X) = \text{expit}(1 + R^{(-1)} - R^{(1)} + 2X + H_1)$ and X is an unmeasured confounder (not included in any analysis in any method) and $R^{(-1)}, R^{(1)}$ are potential outcomes under each treatment. The propensity scores were estimated by a linear logistic regression with H_1 and H_2 as predictors. The true optimal value in these two settings are 1.34 and 2.52, respectively. All three methods deteriorate with the introduction of unmeasured confounders in both S_1 and S_2. In S_1, Q-learning outperforms the other two methods. In S_2, M-learning provides satisfactory results: the value function is close to the true optimal and has a high accuracy of identifying optimal treatment. In comparison, the other two methods performs much worse even with a large sample size.

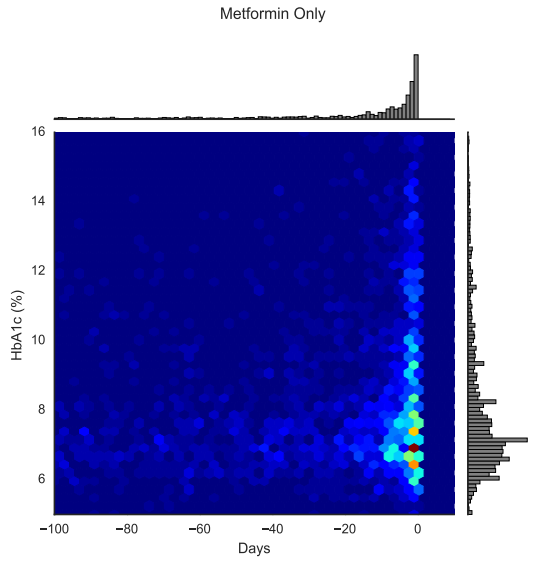
A.2 Additional Figures of Simulations and Real Data Analyses

Below is a brief description of each figure in Appendix A.

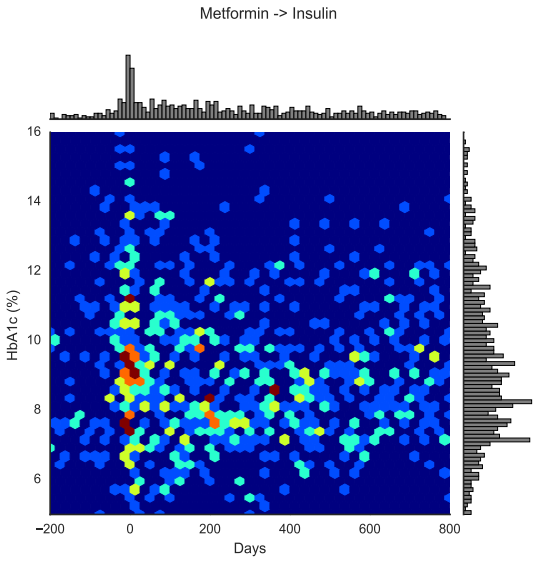
Figure A.1: Descriptive analysis of patterns of HbA1c measurement;

Figure A.2: Additional simulation results: value comparison of four methods with propensity scores correctly specified (top panel) and misspecified (bottom panel);

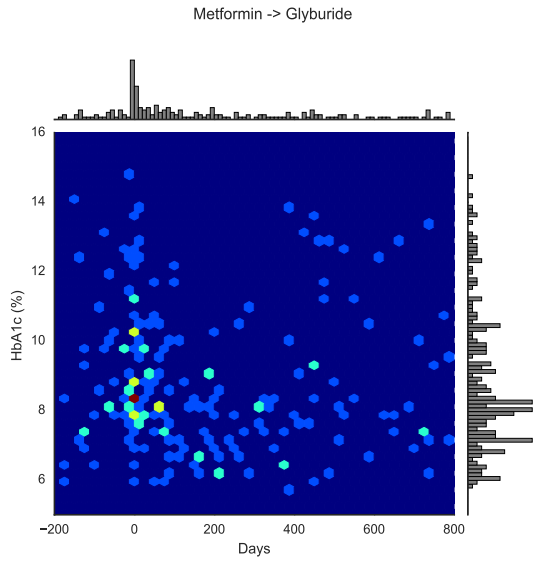
Figure A.3: Additional simulation results: value comparison of four methods in the presence of unmeasured confounders.



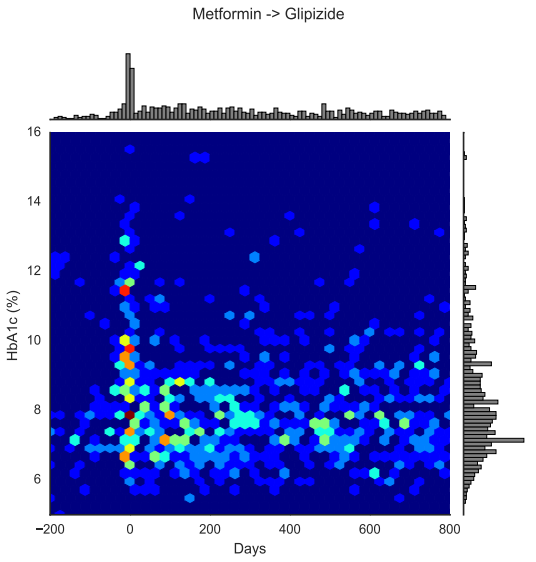
(a) Metformin Only



(b) Metformin + Insulin

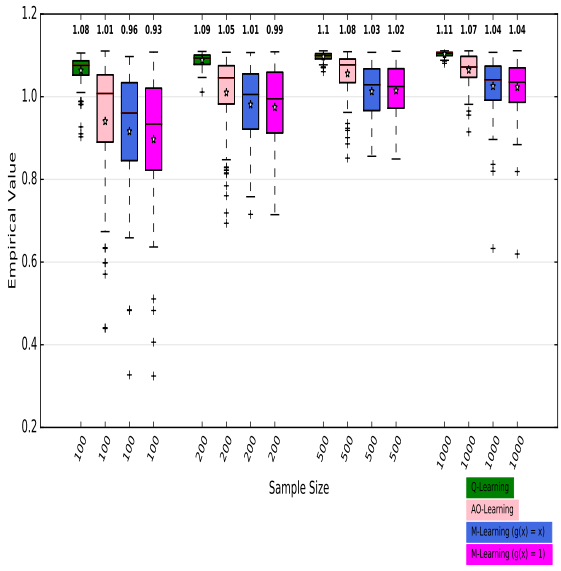


(c) Metformin + Glyburide

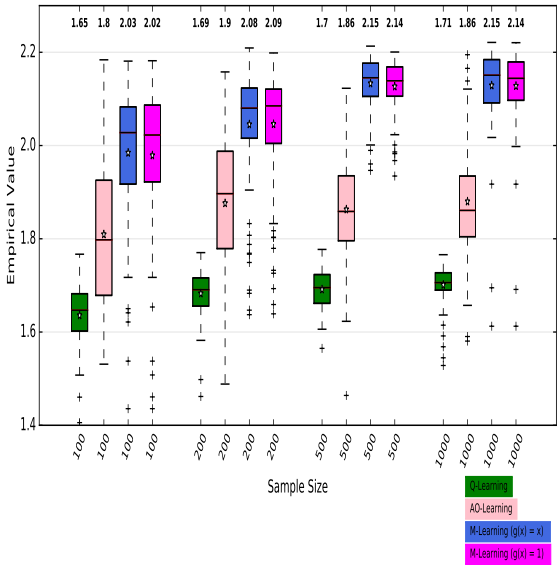


(d) Metformin + Glipizide

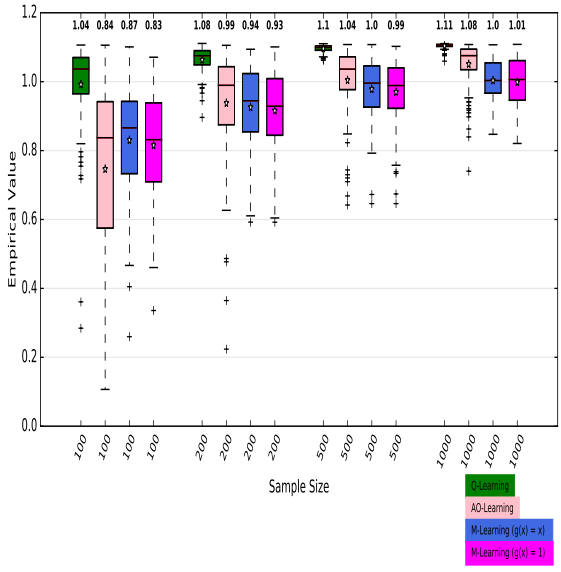
Figure A.1: HbA1c values and measurement intensity (Time 0: first stage treatment prescription)



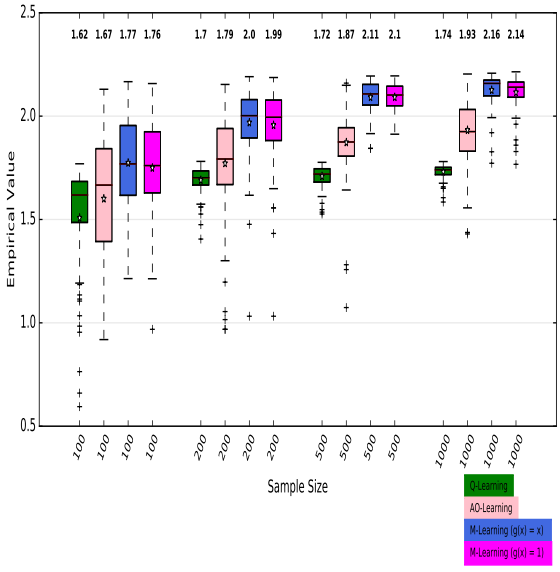
(a) Setting S1: propensity score model correctly specified



(b) Setting S2: propensity score model correctly specified

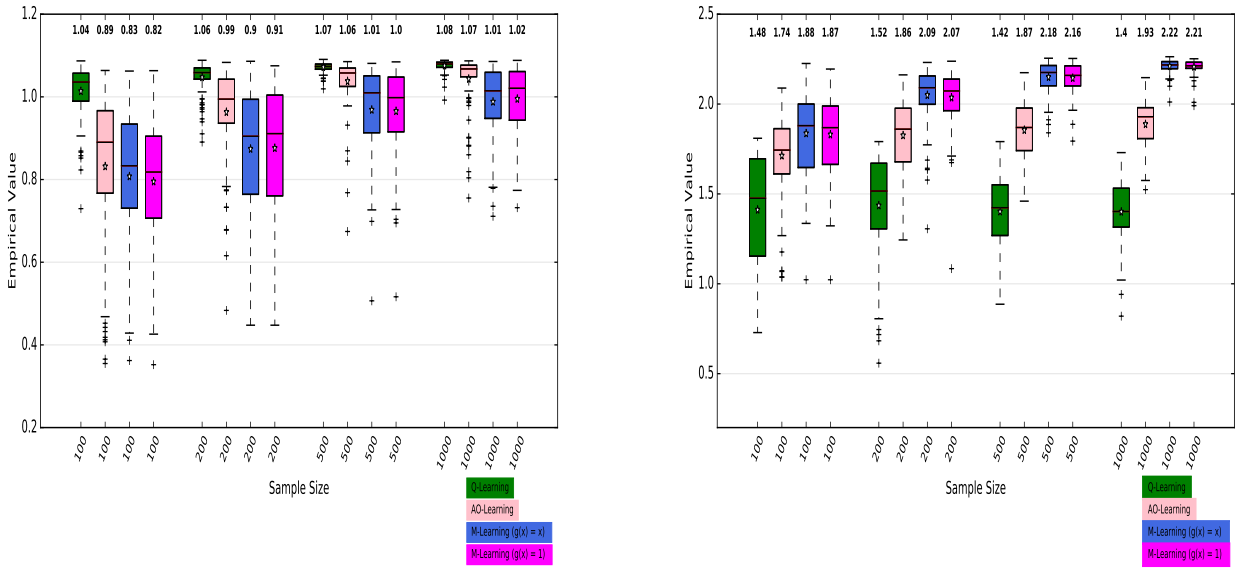


(c) Setting S1, propensity score model misspecified



(d) Setting S2, propensity score model misspecified

Figure A.2: Value comparison of four methods with propensity scores correctly specified (top panel) and misspecified (bottom panel). The numbers at the top of each subfigure are mean values.



(a) Setting S3: unmeasured confounders present

(b) Setting S4: unmeasured confounders present

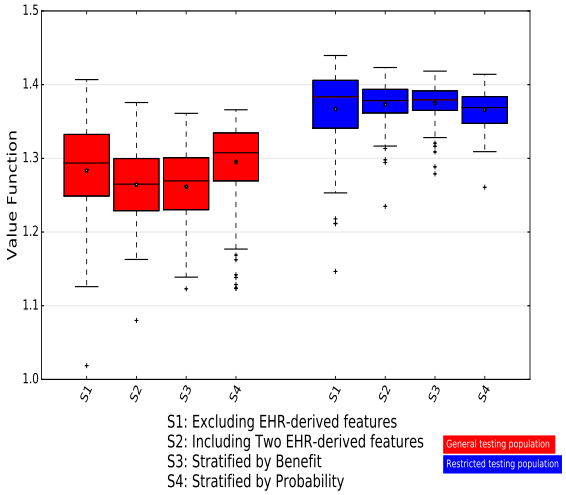
Figure A.3: Value comparison of four methods in the presence of unmeasured confounders. The numbers at the top of each subfigure are mean values.

Appendix B

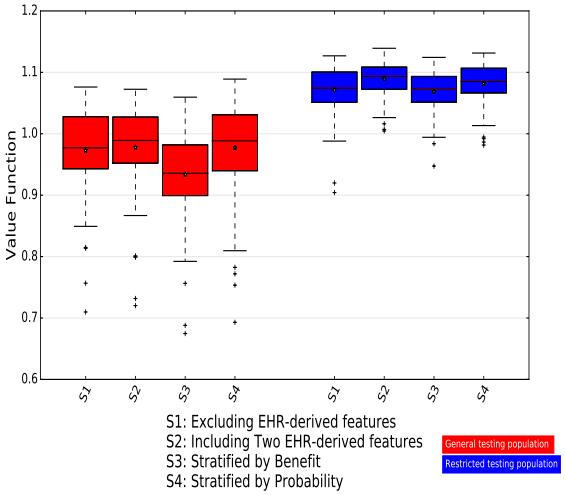
Appendices to Chapter 3

B.1 Additional Simulation Results for Q-learning

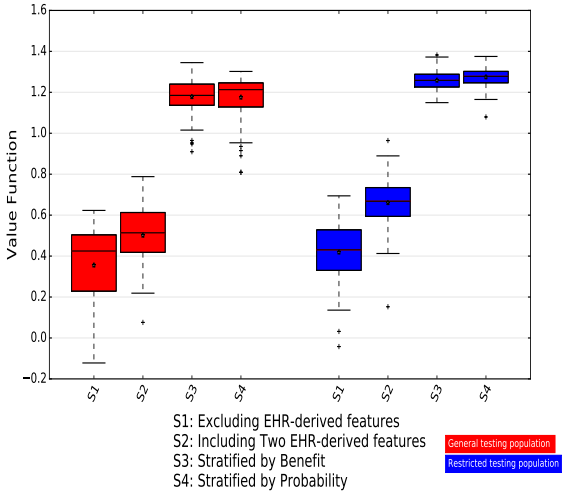
In this section, we present simulation results for Q-learning (Figure B.1). In scenario (i), using RCT feature variables alone achieves a value close to 1.28. Strategy S2 of directly adding super features reduces variability of value function for Q-learning, but does not improve its mean. This is consistent with our justification in Case I in Chapter 3 Section 2.3. In scenario (ii) with a latent tailoring variable, adding super features directly to Q-learning slightly improves the original value function of 0.98 when without H. If using linear regression instead of random forest regression for Q-learning (bottom panel), we observe that both stratification methods much improve results with larger values and smaller variability, which is consistent with our justification in Case II in Chapter 3 Section 2.3.



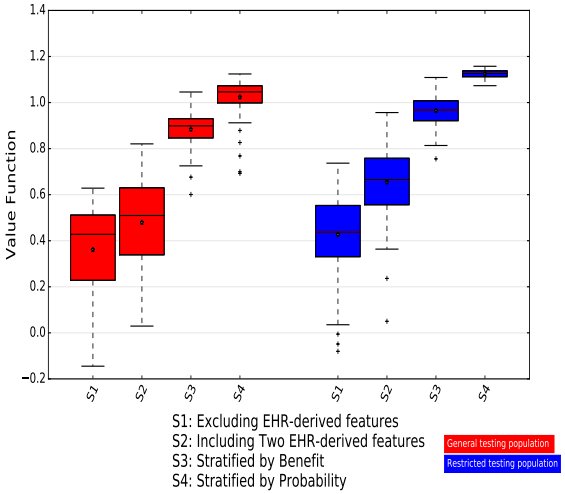
(a) Scenario (i), random forest regression



(b) Scenario (ii), random forest regression



(c) Scenario (i), linear regression



(d) Scenario (ii), linear regression

Figure B.1: Simulation Comparisons for Q-learning (evaluated on independent testing sets generated from general or restricted population; scenario (i) has no latent tailoring variables while scenario (ii) has a latent tailoring variable not used in learning)

B.2 Additional Simulation Results for Unmeasured Confounder

In this section, we present additional simulation results for kernel M-learning under presence of unmeasured confounders in three settings. In these settings, the outcome model is same as that in the main text and was evaluated on the general population. In setting 1 and setting 2, the true treatment assignment model was specified as $P(A = 1|X) = \text{expit}(1 + 2X_1 + X_2 + 0.5X_3)$ and $P(A = 1|X) = \text{expit}(1 + 2X_1 + X_2 + 2X_3)$ respectively, where X_3 is an unmeasured confounder. In setting 3, the true treatment assignment model was specified as $P(A = 1|X) = \text{expit}(1 + 2X_1 + 2X_2 + X_3 + R^{(1)} - R^{(-1)})$ where X_2 is an unmeasured confounder that determines potential outcomes $R^{(1)}$, $R^{(-1)}$ under each treatment and treatment assignment. In addition to the four strategies S1-S4, we include S5 which represents stratification by both super features H_1 and H_2 in the first two settings. Due to relative small sample size for one of the four strata, in setting 3 the result for stratifying by both super features is not available.

The results are summarized in Table A3. The proportion of patients receiving the optimal treatment is defined as "success rate". The value function of proposed approach with stratification remains to be better than not using any EHR-derived features. In S3 and S4, "success rate" is higher than S1. However, the performance decreases when the strength of confounding increases in setting 3 compared to setting 1 and 2. Stratification by both H_1 and H_2 leads to best performance in setting 1 and 2 with highest value function and "success rate".

Table B.1: Additional Results for Value Function and Success Rate

	Setting 1		Setting 2		Setting 3	
Strategy*	Value	Success Rate	Value	Success Rate	Value	Success Rate
S1	0.75 (0.16)	0.66 (0.04)	0.75 (0.16)	0.66 (0.04)	0.45 (0.32)	0.62 (0.07)
S2	0.71 (0.15)	0.66 (0.04)	0.72 (0.15)	0.66 (0.03)	0.81 (0.16)	0.67 (0.04)
S3	0.84 (0.13)	0.68 (0.03)	0.77 (0.14)	0.66 (0.05)	0.75 (0.14)	0.65 (0.03)
S4	0.83 (0.14)	0.73 (0.04)	0.83 (0.14)	0.73 (0.03)	0.79 (0.13)	0.68 (0.02)
S5	0.90 (0.12)	0.72 (0.03)	0.96 (0.07)	0.74 (0.02)	NA	NA

*: S1: RCT features only; S2: Augment RCT feature set by two EHR data-derived super features H_1, H_2; S3: Include H_1 in the feature set and stratify by H_2; S4: Include H_2 in the feature set and stratify by H_1; S5: Stratify by H_1 and H_2.

Appendix C

Appendices to Chapter 4

C.1 EHR Data Preprocessing Flowchart

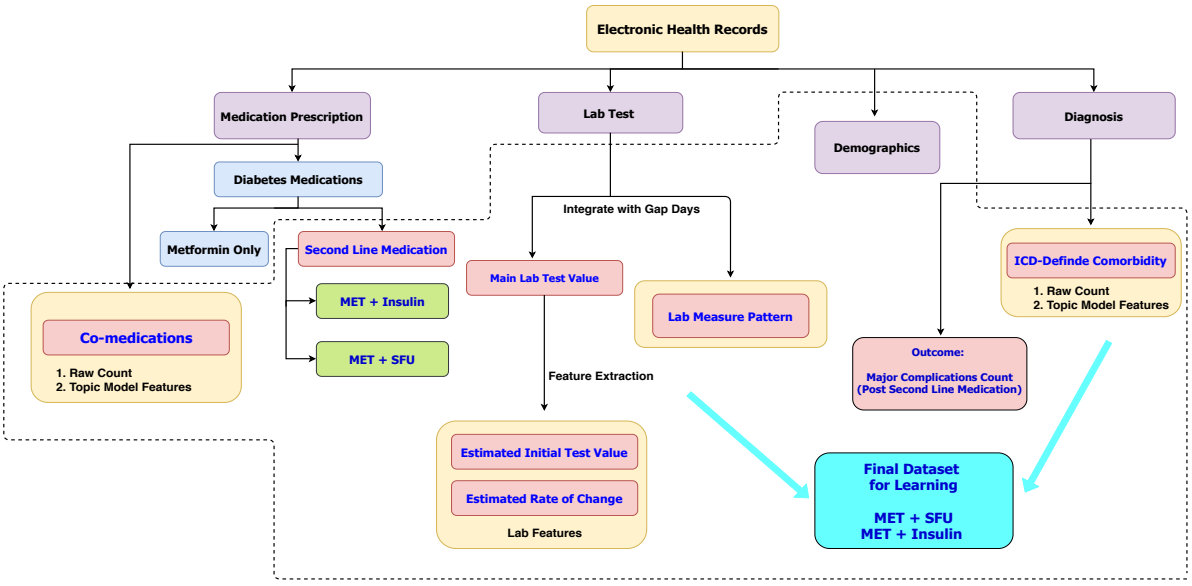


Figure C.1: EHR Data Preprocessing Chart for Different Domains

C.2 Measurement Pattern and Medication Figures

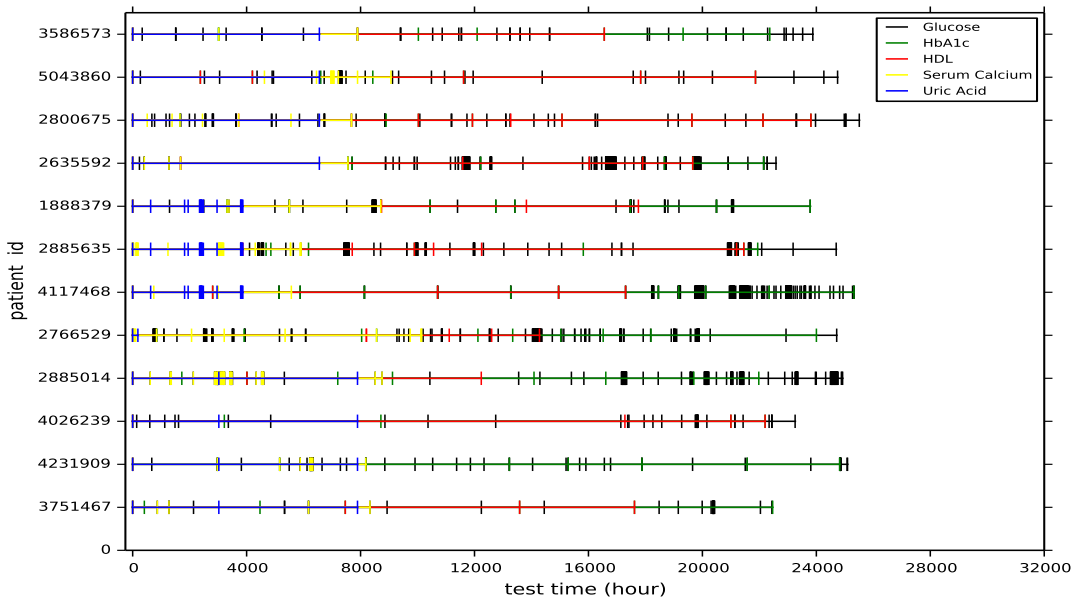


Figure C.2: Sample Patients Lab Tests Spikeplot

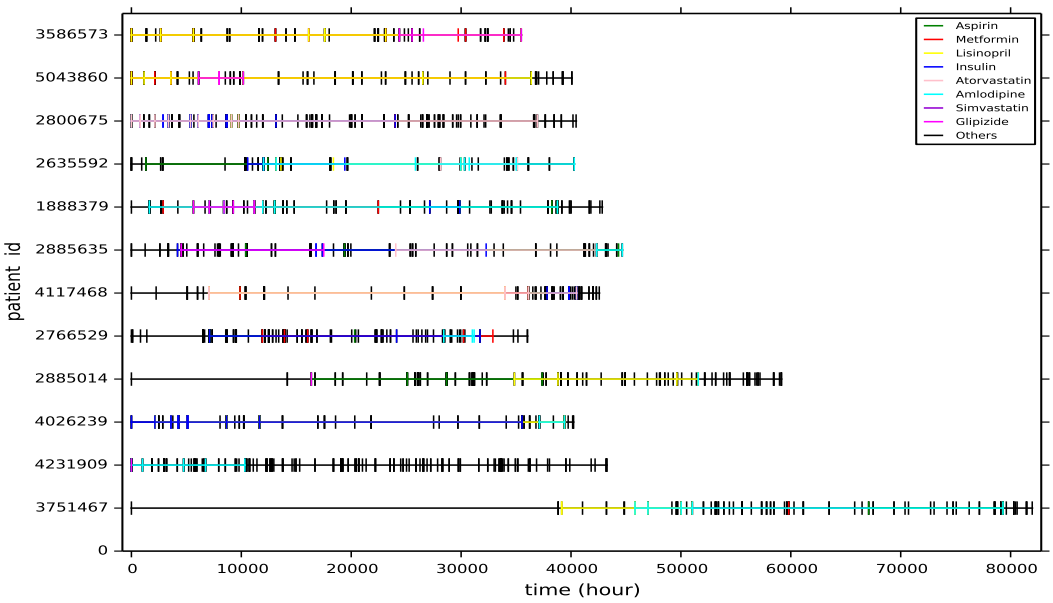


Figure C.3: Sample Patients Medications Spikeplot

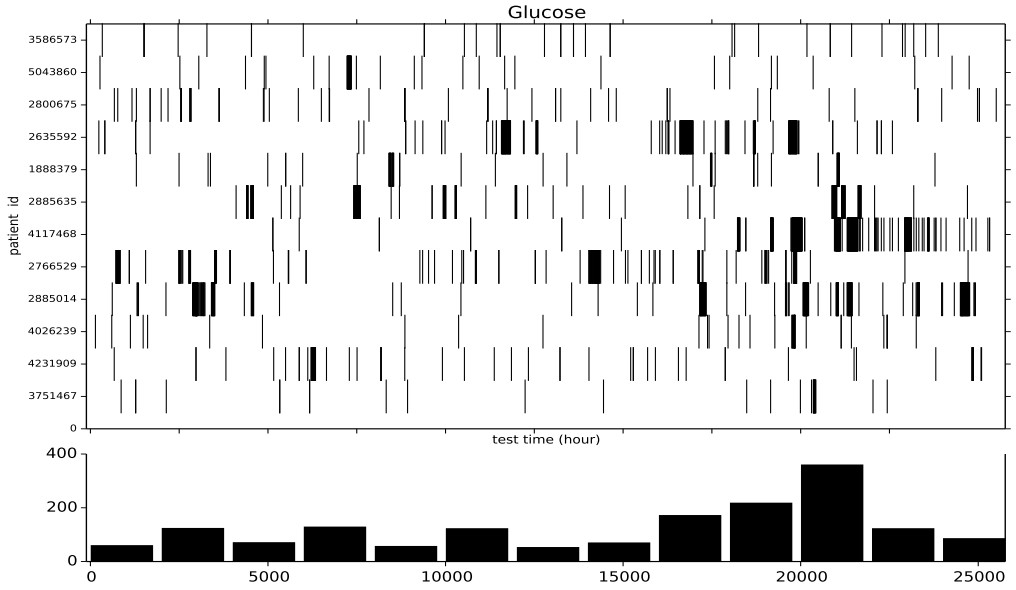


Figure C.4: Sample Patients Glucose Spikeplot

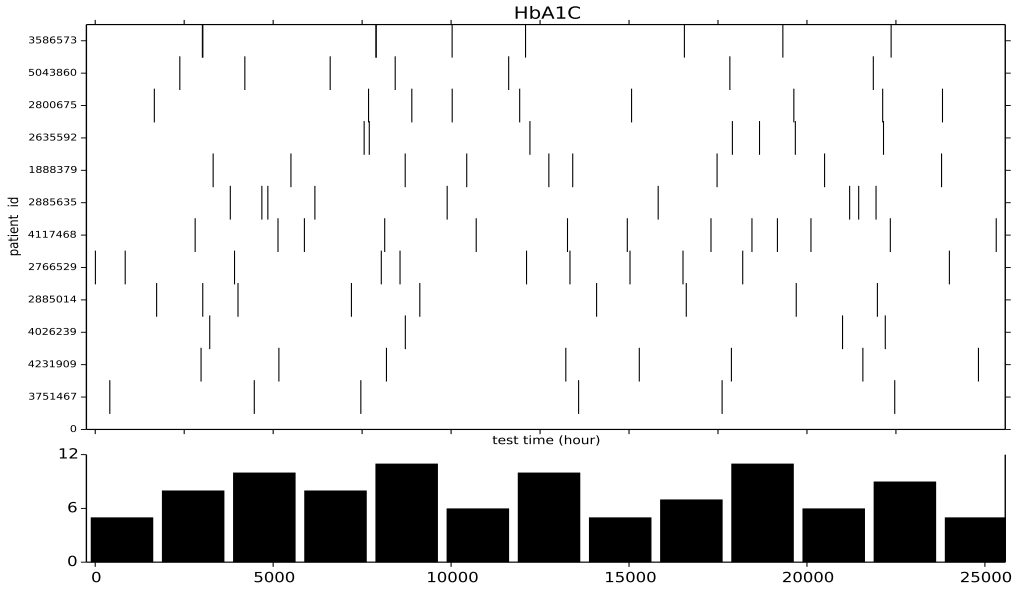
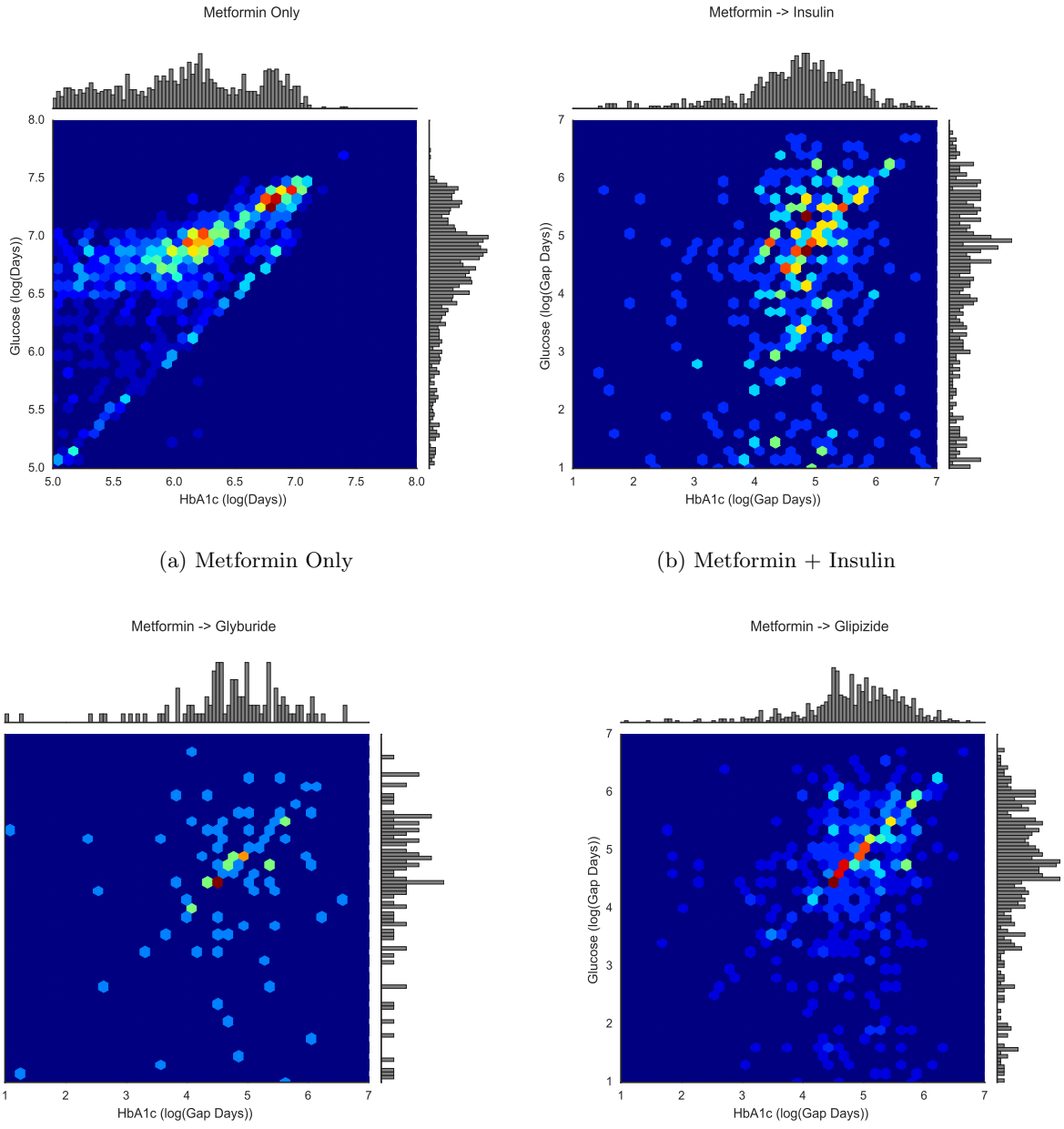


Figure C.5: Sample Patients HbA1c Spikeplot



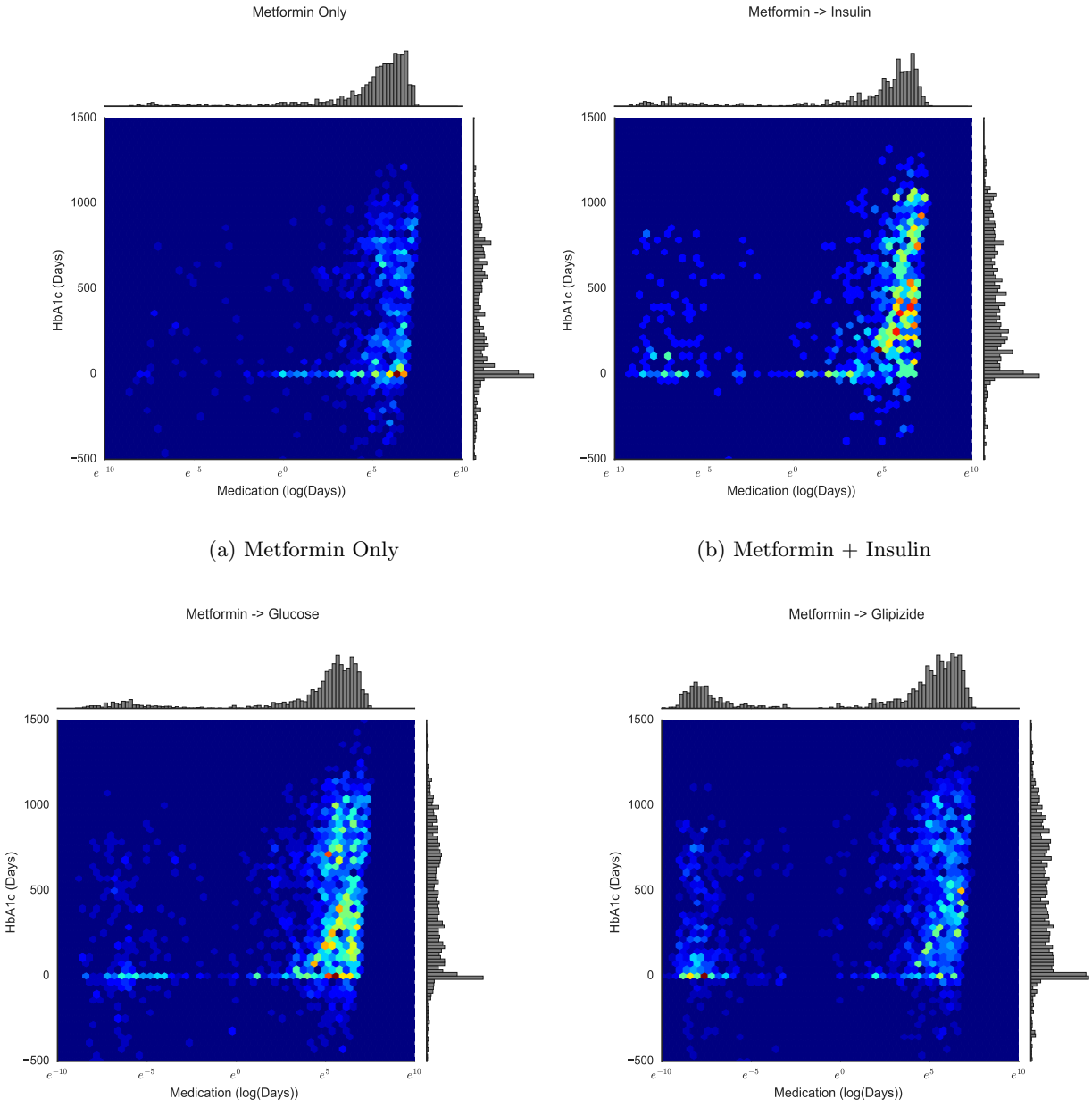
(a) Metformin Only

(b) Metformin + Insulin

(c) Metformin + Glyburide

(d) Metformin + Glipizide

Figure C.6: Longitudinal Measurement Pattern (Gap Days) of HbA1c vs. Glucose



(a) Metformin Only

(b) Metformin + Insulin

(c) Metformin + Glucose

(d) Metformin + Glipizide

Figure C.7: HbA1c Measurement Pattern vs. Medication (Gap Days)

C.3 Learned Topics Visualization from LDA Model

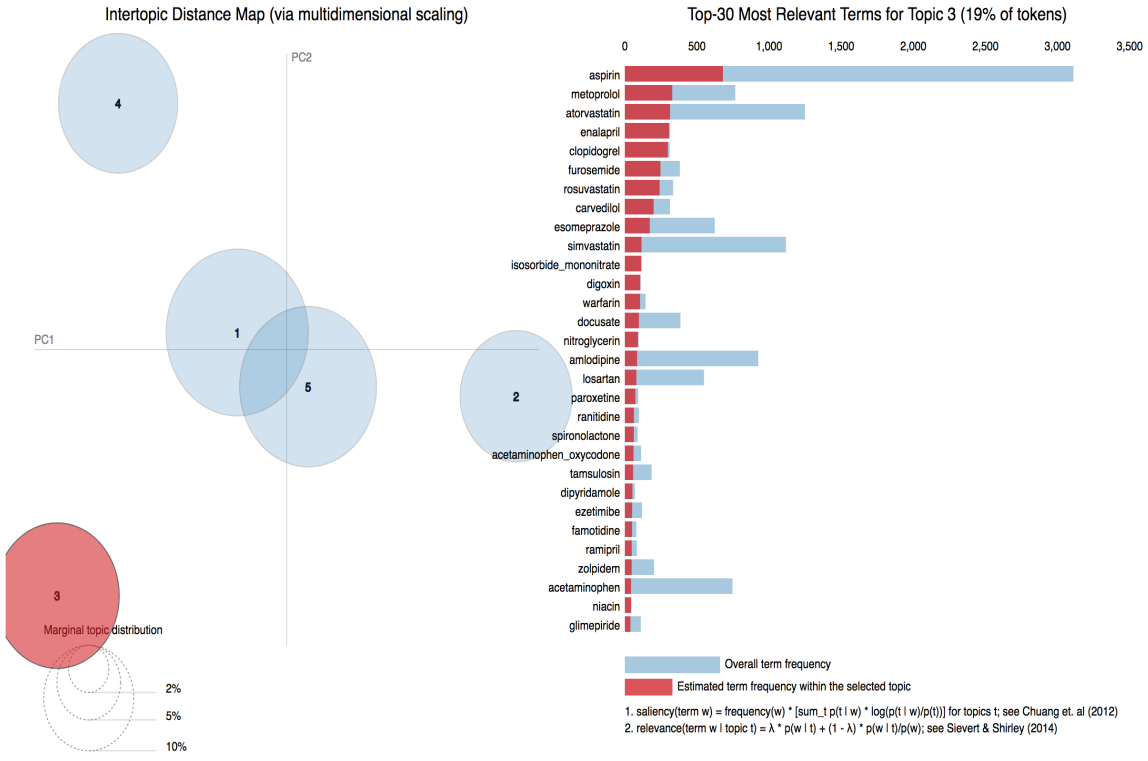


Figure C.8: Visualization of Co-medication Topic #3 and Their Most Relevant Terms

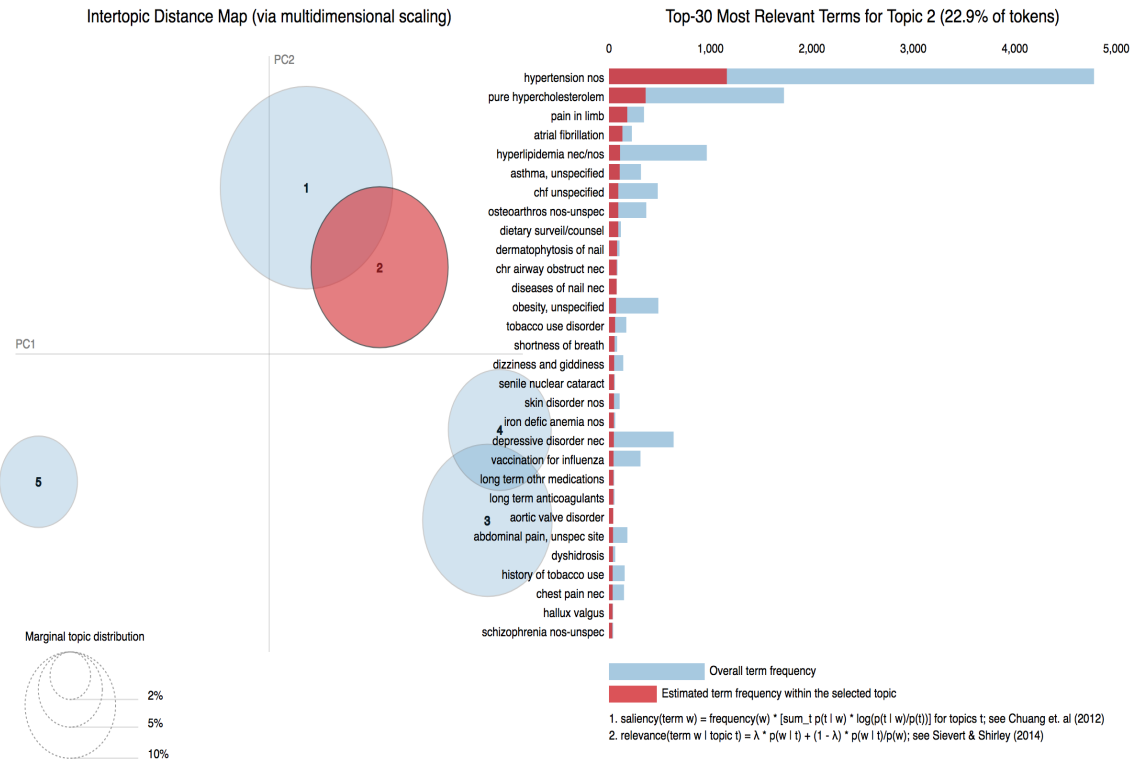


Figure C.9: Visualization of Condition Topic #2 and Their Most Relevant Terms

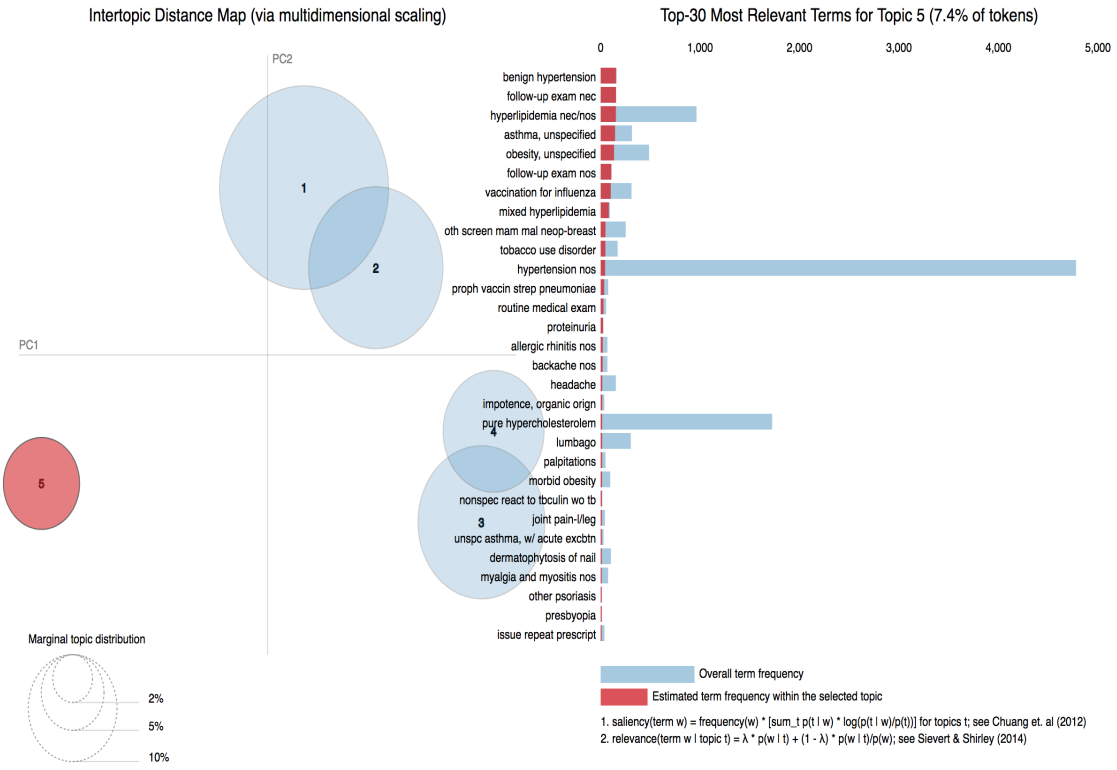


Figure C.10: Visualization of Condition Topic #5 and Their Most Relevant Terms