Learning Optimal Individualized Treatment Rules from Electronic Health Record Data

Yuanjia Wang*, Peng Wu, Ying Liu
Department of Biostatistics
Columbia University
*Email: yw2016@cumc.columbia.edu

Chunhua Weng
Department of Biomedical Informatics
Columbia University

Donglin Zeng
Department of Biostatistics
University at North Carolina at Chapel Hill

Abstract—Medical research is experiencing a paradigm shift from "one-size-fits-all" strategy to a precision medicine approach where the right therapy, for the right patient, and at the right time, will be prescribed. We propose a statistical method to estimate the optimal individualized treatment rules (ITRs) that are tailored according to subject-specific features using electronic health records (EHR) data. Our approach merges statistical modeling and medical domain knowledge with machine learning algorithms to assist personalized medical decision making using EHR. We transform the estimation of optimal ITR into a classification problem and account for the non-experimental features of the EHR data and confounding by clinical indication. We create a broad range of feature variables that reflect both patient health status and healthcare data collection process. Using EHR data collected at Columbia University clinical data warehouse, we construct a decision tree for choosing the best second line therapy for treating type 2 diabetes patients.

I. INTRODUCTION

During the past decade, medical research is experiencing a paradigm shift from "one-size-fits-all" strategy to a precision medicine approach where the right therapy, for the right patient, and at the right time, will be prescribed [1], [2]. Technological advances in data collection are enabling this shift by providing large-scale personal data (e.g., clinical assessments, electronic health records, genomic measures) to meet the promise of individualized health care. Statistical methods for estimating optimal medical treatment and disease management strategy according to patient-specific characteristics for each individual have received extensive attention in the recent literature [3], [4], [5], [6]. Powerful machine learning methods are proposed to estimate the optimal individualized treatment rule (ITR) based on high-dimensional subject-specific feature variables. In particular, reinforcement learning (e.g., Q-learning [3], [7]) can be used to discover the optimal treatment at each stage of critical decision point by a backward induction to maximize the estimated future clinical reward (Q-function). However, the regression based Q-learning selects the optimal treatment by modeling the Q-function and its contrasts that are not explicitly related to the optimization of the objective function (i.e., value function [7]), and thus may suffer from incorrect model assumptions. The mismatch between maximizing the Q-function and the value function may lead to suboptimal treatment rules due to overfitting the regression model [5]. Recent work by [5], [6] suggests an alternative framework, outcome-weighted learning (O-learning), to estimate the optimal ITR by directly optimizing the expected clinical outcome at the end of study. The resulting optimal treatment regimen is identified by a weighted support vector machine [8] and can estimate an ITR with any unconstrained nonparametric functional form.

Electronic health records (EHR) provide rich longitudinal data to address important clinical questions about the effectiveness of interventions and allows estimating ITRs in real world settings. Recently, EHR data collected in large scale networks are proved as useful resources to characterize treatment pathways in diverse populations [9]. However, subjects recorded in these databases rarely received random assignments of treatments, and this non-experimental feature poses major challenges in inferring causality from observational data, where confounding and selection bias may obscure a true effect or create a spurious one [10]. Thus combining medical domain knowledge with valid statistical methods is critical in order to identify the optimal ITR with least bias using observational data. A sound analysis needs to properly handle potentially high-dimensional confounding factors and systematic imbalance or heterogeneity among patients receiving different treatments. A wide range of statistical methods are developed to draw inferences from non-randomized data, including propensity score weighting [11], and structural nested models. However, none of these methods are combined with O-learning to discover ITR from large-scale observational EHR data. In addition, current O-learning methods [5], [6] provide fully nonparametric decision rules which are not easy to interpret.

In this work, we adapt O-learning approach to EHR data in order to estimate optimal ITR depending on a patient's personalized characteristics such as health history, disease risk factors, and predictive markers. To provide highly interpretable ITRs, we incorporate classification and regression trees (CART) in the O-learning. To mitigate confounding by clinical indication for the treatments under study as a result of physicians making treatment choices in light of available clinical information, we study treatment strategies where little evidence exists to guide choices between alternative treatments (e.g., second line treatment choices for type 2 diabetes [T2D] patients). To further mitigate confounding, we create a broad range of potentially relevant feature variables from the EHR data and apply propensity score adjustment using

random forest [12]. Using EHR data collected at Columbia University clinical data warehouse, we construct a decision tree to discover ITR for choosing the best second line therapy for T2D patients and demonstrate the advantage of O-learning in terms of achieving a greater clinical reward as compared to Q-learning.

II. METHODS

The main goal of O-learning is to achieve personalized medical decision making by estimating the optimal treatment choice that assigns individualized treatment at each decision stage according to a patient's personal characteristics and intermediate outcomes. The optimality of an ITR is measured by a final clinical outcome (or reward) and ITR is expected to be as effective or better than "one-size-fits-all" rule that do not tailor to a patient's personal features.

A. O-learning for Observational Data

Let R_i denote the reward or the clinical outcome for the ith subject, where a larger rewards may correspond to better functioning or fewer symptoms, depending on the clinical setting. Let H_i denote feature variables collected before the initiation of a treatment. Let A_i denote the received treatment assignment taking values $\{-1, 1\}$. An ITR is a decision function, $\mathcal{D}(H_i)$, that maps the domain of H_i to the treatment choices in $\{-1, 1\}$. The value function is defined as the expected clinical reward under the rule $\mathcal{D}(H)$ [7], $\mathcal{V}(\mathcal{D}) = E_{\mathcal{D}}[R_i]$. Under the positivity assumption, the value function can be obtained from observed data through [7]

$$\mathcal{V}(\mathcal{D}) = E\left[\frac{I(A_i = \mathcal{D}(H_i))R_i}{\pi(A_i, H_i)}\right], \tag{1}$$

where $\pi(A_i, H_i)$ is the probability of the ith patient receiving treatment A_i. An optimal ITR, $\mathcal{D}^*$, is the treatment rule that maximizes the value function, that is, $\mathcal{D}^* = \max_{\mathcal{D}} \mathcal{V}(\mathcal{D})$.

Q-learning indirectly estimates the optimal ITR through fitting a model for R_i. Specifically, Q-learning predicts outcomes as $\widehat{R}_i = \widehat{f}(H_i, A_i)$ using H_i and interactions between H_i and A_i as features, and selects the optimal treatment as $a^* = \max_{\{a=1,-1\}} \widehat{f}(H, a)$. Q-learning may suffer from incorrect model assumptions and overfitting when estimating f, especially when H_i contains high-dimensional features. Even if a nonparametric learning algorithm is used, the method models the Q-function and its contrasts that are not explicitly related to the optimization of the objective function in (1) and this limitation was shown to result in suboptimal ITRs due to overfitting the regression model in [5], [6].

In contrast, O-learning directly maximizes the value function in (1). Note that maximizing (1) is equivalent to minimizing $E\left[\frac{I(A_i \neq \mathcal{D}(H_i))R_i}{\pi(A_i, H_i)}\right]$, which is a weighted missclassification error rate with A_i as class labels and R_i as weights. O-learning capitalizes on this observation and turns learning optimal ITR into a weighted classification problem. Recently, [13] proposed an improved algorithm to the original O-learning, which is adopted here for further development. The basic idea is to replace weights R_i in (1) by some surrogate weights, denoted

as $\widetilde{R}_i$, which yields the same asymptotic optimal decision rule as [5] but is more efficient. Note that if we define $\widetilde{R} = R - \widehat{R}$, where the predicted $\widehat{R}$ depends only on H but not A, then the optimal decision function associated with $\widetilde{R}$ remains the same. Intuitively, because learning ITR is essentially to learn the qualitative interaction between A and H, removing any main effects of H on R has no influence. Thus, the first improvement compared to [5] is to remove the main effects by taking residuals from a prediction model (i.e., random forest).

A computational challenge of using R_i as weights in a weighted classification is that R_i can be negative, so the objective function is no longer convex and conventional classification algorithms based on convex optimization cannot be applied. Methods in [5] resolved this issue by subtracting an arbitrarily small constant from R in order to make it positive, which unfortunately increases the variability of the weights and could be unstable in practice. To this end, we propose the following procedure similar to [13]. First, note that by simple algebra, the value function for any decision function f under the new weights $\widetilde{R}_i = R_i - \widehat{R}_i$ is

$$E\left[\frac{\widetilde{R}_i}{\pi(A_i, H_i)} I(A_i \neq \text{sign}(f(H_i)))\right]$$

$$= E\left[\frac{|\widetilde{R}_i|}{\pi(A_i, H_i)} I(A_i \text{sign}(\widetilde{R}_i) \neq \text{sign}(f(H_i)))\right] - E\left[\frac{\widetilde{R}_i^-}{\pi(A_i, H_i)}\right],$$

where $x^- = -\min(0, x)$. Therefore, estimating the optimal ITR is equivalent to maximizing the first term on the right hand side of the above equation, which can be solved by a weighted classification algorithm with weights $|\widetilde{R}_i|/\pi(A_i, H_i)$ and class labels $A_i \text{sign}(\widetilde{R}_i)$. This guarantees a convex optimization problem that is easy to compute.

In a randomized trial, the probability of receiving a treatment, $\pi(A_i, H_i)$, is the randomization probability specified by design and thus is known in the analysis. However, for observational studies, the treatment assignment mechanism is usually unknown and needs to be estimated from data. To accommodate the observational feature of EHR data in clinical data warehouse, we pay special attention to the clinical records measurement patterns and create informative variables that are predictive to the observed measurement patterns and adjust for confounding (details in section II-B). These feature variables are then used to estimate $\pi(A_i, H_i)$.

To summarize, our proposed O-learning for EHR data performs the following three steps.
Step 1. Use feature variables H_i to predict treatment assignment probability (or propensity score), $\widehat{\pi}(A_i, H_i)$, for each individual.
Step 2. Use feature variables H_i to obtain a predicted outcome $\widehat{R}_i$ by fitting a random forest with inverse propensity score weighting by $\widehat{\pi}(A_i, H_i)$.
Step 3. Obtain $\widetilde{R}_i = R_i - \widehat{R}_i$ for each subject, and fit a weighted classification tree to estimate the decision function f, where the weights are $|\widetilde{R}_i/\widehat{\pi}(A_i, H_i)|$, and the class labels are $A_i \text{sign}(\widetilde{R}_i)$.
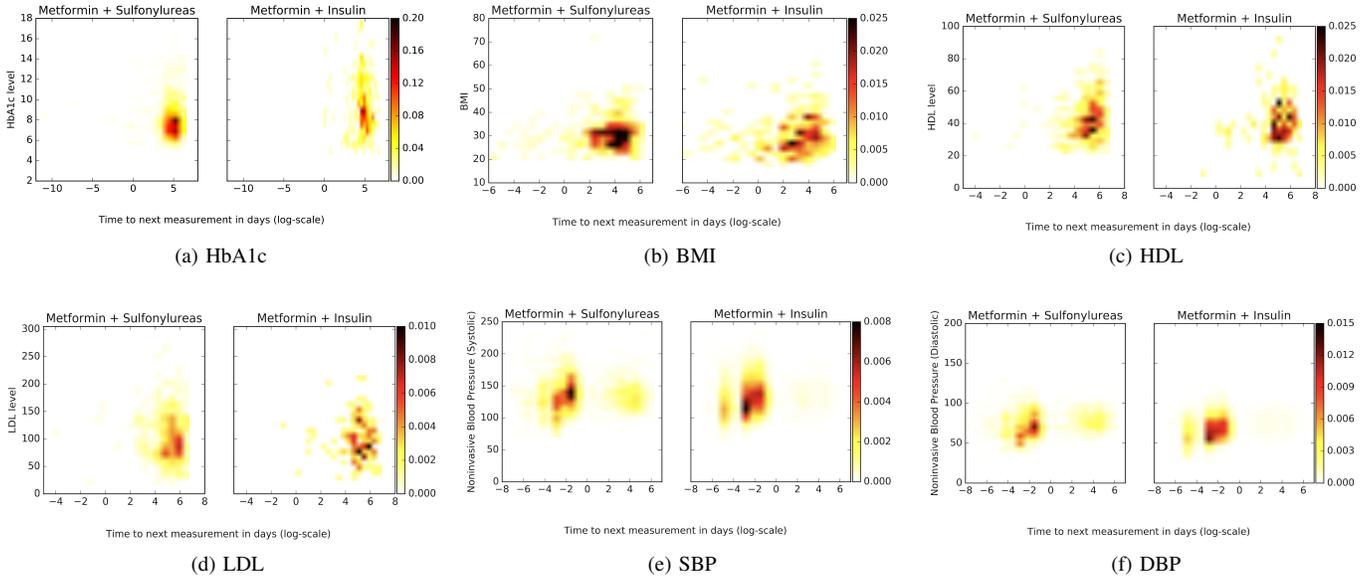
Fig. 1. Tests Measurement Patterns Discriminating Received Treatment

B. EHR Data

Current guidelines for managing T2D advocate metformin medication as the preferred first line therapy [14], but no longer recommend a particular second-line agent largely because of insufficient evidence on the long-term outcomes of different treatments [15]. Literature reveals uncertainties of timely insulin initiation in clinical practice and optimal sequence of treatments for insulin therapy versus oral hypoglycemic agents (OHA, e.g., sulfonylureas) for treating T2D. The impact of delayed insulin initiation and the comparative effectiveness of augmenting metformin with insulin therapy or a second line OHA is largely unknown, particularly in a real-world setting. We propose to apply the developed O-learning to explore the optimal second line treatment of T2D using data from Columbia University Medical Center's (CUMC) clinical data warehouse (CDW) and New York Presbyterian hospital patient electronic health records [16]. The CUMC CDW contains over 20 years of health information for about 4.5 million patients with diverse ethnicities. The CDW data resource is part of the New York City Clinical Data Research Network which include complete, comprehensive, and longitudinal data for at least 2.5 million patients from 22 organizations across seven systems in one of the nation's most populous and diverse regions.

However, EHR data are inherantly biased because they are collected in an uncontrolled environment and are not carefully curated for research purposes. Ignoring such biases in statistical analyses may lead to the detection of spurious effects, reversals of cause and effect [10], and errors when predicting optimal drug dosage [17]. We tackle these challenges by a two-stage analytic strategy. In the first stage, we extract feature variables that not only reflect patients' health status but also are informative of the healthcare and EHR

data recording process that can be used to mitigate bias. We characterize the raw data into three categories: patient demographics (age, gender, race), patient health history including prior ICD10 diagnosis, medication prescription, recent and overall average lab measures (HbA1c, LDL, HDL), recent and overall average vital signs (systolic blood pressure, SBP; diastolic blood pressure, DBP, body mass index, BMI), and measures of the CUMC healthcare system process. The latter category includes temporal clinical events and laboratory tests recording patterns, which are one of the most widely-used features to capture the characteristics of a healthcare system in EHR-related research [18]. However, capturing the context of laboratory testing is a challenge since EHR data do not directly offer an explicit reason of why a test was ordered. Thus, data-driven automatic feature extraction is recommended to discover contexts and patterns for lab measurements [19].

C. Feature Extraction and Implementing O-learning

We created temporal features of the rates of tests, numeric values of tests, and their bivariate association with the gap times between two consecutive measurements, which are used to adjust for bias and patient heterogeneity [18]. Each patient was aligned at their second line treatment initiation time, and their lab measures within the past month or during the entire recording period before treatment were calculated. The density and frequencies of the measures were obtained by bivariate smoothing.

Next, we used the aforementioned feature variables to construct propensity scores by random forest to predict probability (i.e., $\widehat{\pi}_i(H_i, A_i)$) of observing OHA or insulin initiation for each patient. To further handle potential bias caused by missing post-treatment outcome data, we used inverse probability weighting (IPW) of missing, which is a valid approach under missing at random (given observed features) assumption [20].

In the second stage, the discovery stage, we fitted O-learning to estimate the optimal ITR with extracted tailoring variables to choose OHA or insulin for patients already treated with metformin for at least 3 months. The clinical outcome to be optimized is glycemic control (level of HbA1c between 2 month post treatment and one year, R_i), and the tailoring variables include demographics, BMI, co-morbidity informed by ICD10 codes, SBP, DBP and other general health status variables. The negation of the post-treatment HbA1c was maximized (equivalently HbA1c was minimized). The value function was inversely weighted by treatment propensity and missing outcome probability as defined in (1) and computed on the full training sample.

III. RESULTS

We extracted EHR data from the CDW at CUMC for adults age 18 or older and having at least one T2D diagnosis ICD-9 code selected from the following set of codes: 250.00, 250.02, 250.10, 250.12, 250.20, 250.22, 250.30, 250.32, 250.40, 250.42, 250.50, 250.52, 250.60, 250.62, 250.70, 250.72, 250.80, 250.82, 250.90, 250.92, between 1/1/2008 and 12/31/2012. To study the choice of second line therapy for treating T2D, we further restricted our analyses sample to patients who had been on metformin for at least 3 months and who augmented treatment with either insulin or a second line OHA, sulfonylurea, and with at least one record post second-line treatment initiation. There were 357 patients in sulfonylureas group with 1725 longitudinal records and 203 in insulin group with 982 longitudinal records.

We summarized the bivariate association between the measurement intensity of several important laboratory tests/vital signs (measured before second line treatment initiation) and time to next measurement on the logarithm scale in Figure 1. The two-dimensional intensity plots show smoothed frequencies of subjects in the space defined by the numeric values of lab tests and measurement gaps. For example, Figure 1a shows that a high proportion of patients receiving sulfonylureas had a wider range of measurement gaps (between 3 and 6 on the log-scale) and lower HbA1c level as compared to patients receiving insulin (gap time between 4.5 and 6, greater frequency with higher HbA1c). Similarly, discriminant patterns are observed for other lab tests. Varying gaps between measurements of lab tests may associate with different health status of a patient [18] and explain some of the variability of different treatment initiation. Thus, feature variables created from these bivariate intensity plots are useful for predicting the probability of receiving alternative treatments. We discretized each density in Figure 1 into 9 subgroups based on low, medium, and high quartiles of lab tests and gap time (e.g., the first group is "low HbA1c and short measurement gap time"), and created indicator variables for each patient's group membership (intensity pattern group) as feature variables to estimate propensity scores. Other univariate summaries such as number of observations and length of observations were also constructed. We present a heat map of the standardized continuous feature variables in Figure 4. The features with
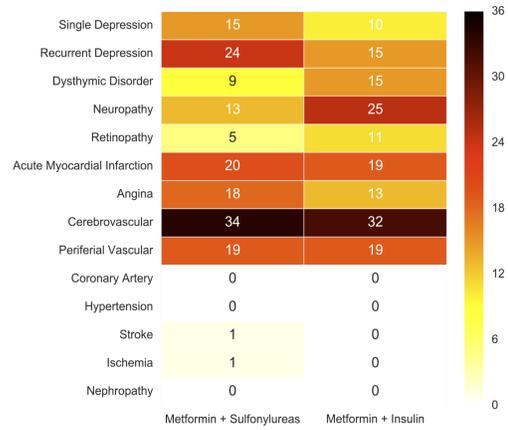
Fig. 2. Frequencies of Pre-treatment Co-morbidity Conditions Extracted from ICD9 Codes

(a) Sulfonylureas

(b) Insulin
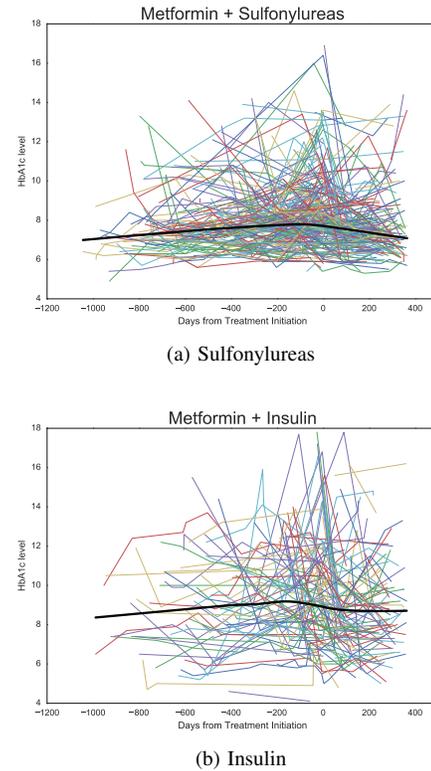
Fig. 3. Longitudinal Measurements of HbA1c for Individuals (colored lines) and Smoothed Overall Trajectory (black solid line)

most discriminant ability comparing treatment groups include length of observations, frequency of observations, bivariate intensity patterns, and some mean numeric values of lab tests (e.g., overall mean HbA1c, recent mean BMI). Discrete feature variables extracted include ICD9 co-morbidity codings (Figure 2). The frequency of occurrence is low in general, where the most prevalent condition is cerebrovascular disease.

Fig. 4. Heatmap of Extracted Continuous Feature Variables

A. Learning Optimal ITR

We present the longitudinal HbA1c measurements for all individuals included in the analyses in Figure 3. On average, there is an increasing trend (smooth solid black line) before treatment initiation and a decreasing trend after treatment. We included all the feature variables displayed in Figure 2 and 4 and general demographic variables measured prior to the treatment initiation to predict propensity of treatment assignment using random forest. Missing values of feature variables were imputed using proximity from the random forest. Consistent with observed in Figure 1 and 4, the most important features predictive of propensity scores include overall HbA1c mean value, bivariate SBP and DBP measurement patterns, length and number of measures of SBP and DBP. To examine whether propensity scores have balanced the distribution of pre-treatment feature variables between treatment groups, we compared group means of most important variables in Table I. Within each stratum defined by the quartiles of the predicted propensity scores, the means of feature variables are comparable between treatment groups, suggesting the propensity score balancing is effective.

The goal of estimating optimal ITR is to find a decision tree based on individual features so that for subjects who follow the ITR, the post-treatment HbA1c between two and twelve months will be under control (minimized). There were 85 subjects missing post treatment HbA1c, and we used IPW [20] to adjust for missing. The reciprocal of the product of treatment propensity and missing probability were used as

weights in the O-learning outlined in section II-A. A CART was fit using the feature variables for each subject except the measurement pattern variables which are informative for adjusting bias but may not directly relate to a future patient's health status to tailor treatment (a total of 36 features). As a comparison, we also fitted Q-learning [7] with regression tree and summarized the results in Table II. The values for O-learning and Q-learning in the table were computed using 2-fold cross validation: the regression model was fitted on the training samples and the value function was computed on the testing sample. The optimal ITR estimated by O-learning has a lower HbA1c than Q-learning for all tree levels. For the "one-size-fits-all" rules, the average pre-treatment HbA1c in sulfonylureas group is 8.56 and in insulin group is 9.82 (Table II). Augmenting all patients with OHA leads to a post-treatment HbA1c of 7.81 and augmenting all with insulin leads to 7.91. The 2-fold cross validation result shows optimal ITR estimated by O-learning with 4-level decision tree achieves a post-treatment HbA1c of 7.76, which approaches the normal level, and is lower than augmenting with medication or insulin for all patients. In conclusion, the optimal ITR estimated by O-learning leads to a lower HbA1c compared to Q-learning or non-individualized rules on the testing samples.

We present the 4-level decision tree fitted on all subjects in Figure 5. The most important tailoring variables are pre-treatment mean HbA1C (importance score = 0.299), recent 4-month mean HDL (importance score = 0.252), recent month mean DBP (importance score = 0.220), recent month

TABLE I
BALANCING FEATURE VARIABLE DISTRIBUTION WITHIN PROPENSITY
SCORE QUANTILE DEFINED STRATA

Features	0-25th quartile			25 -50th quartile		
	Sulf	Insul	p value	Sulf	Insul	p value
Average HbA1c	8.62	9.37	0.295	9.15	9.60	0.555
Length SBP	180.40	268.67	0.189	234.77	172.63	0.259
Length BMI	175.93	232.30	0.385	270.19	165.44	0.081
Length LDL	317.60	441.51	0.211	434.61	399.69	0.710
Length HbA1c	303.20	446.51	0.109	516.00	426.56	0.316
SBP (med-med)	3.83	3.59	0.765	0.58	2.06	0.030
Number of SBP	32.45	35.00	0.711	12.63	18.44	0.255
DBP (low-long)	1.91	2.27	0.597	0.47	1.60	0.037
DBP (high-med)	2.25	4.50	0.027	1.07	2.40	0.124
DBP (med-med)	5.33	2.74	0.015	1.27	1.67	0.543
	50-75th quartile			75 -100th quartile		
	Sulf	Insul	p value	Sulf	Insul	p value
Average HbA1c	8.29	8.99	0.338	7.50	8.84	0.105
Length SBP	341.87	360.33	0.830	374.98	300.14	0.180
Length BMI	353.08	289.44	0.405	366.17	310.43	0.349
Length LDL	582.87	480.67	0.437	584.51	502.43	0.193
Length HbA1c	586.79	514.44	0.518	578.34	546.57	0.644
SBP (med-med)	0.10	0.22	0.612	0.07	0.29	0.300
Number of SBP	7.23	6.33	0.687	5.54	3.86	0.075
DBP (low-long)	0.23	0.89	0.167	0.12	0.00	0.023
DBP (high-med)	0.39	0.22	0.556	0.17	0.14	0.867
DBP (med-med)	0.36	0.44	0.860	0.00	0.29	0.172

TABLE II
VALUE FUNCTION (HbA1c) OF FITTED TREATMENT RULES

	Individualized Treatment Rules	
Tree level	O-learning	Q-learning
2	7.72	7.88
3	7.79	8.13
4	7.76	7.92
5	7.73	7.97
	One-size-fits-all Rule	
	All Sulfonylureas	All Insulin
Pre-treatment	8.56	9.82
Post-treatment	7.81	7.91

mean LDL (importance score $= 0.129$), and pre-treatment neuropathy diagnosis (importance score $= 0.067$). Overall, about 70% patients were estimated to have sulfonylureas as the optimal second line treatment while 30% estimated to have insulin as optimal. As an example of optimal ITR, a patient with HbA1c ≤ 10.66, recent 4-month mean HDL ≤ 43.21 and recent 4-month LDL ≤ 84.0 should initiate insulin as the second-line therapy, while a patient with similar features but with LDL > 84.0 should use sulfonylureas.

IV. DISCUSSION AND CONCLUSION

In this work, we propose O-learning for EHR data to transform identifying optimal ITR to a classification framework to yield flexible and interpretable ITR as a decision tree. This transformation opens an avenue to merge powerful classification tools such as random forest, classification tree or support vector machine to optimize clinical decision making. The direct treatment choice optimization framework of O-
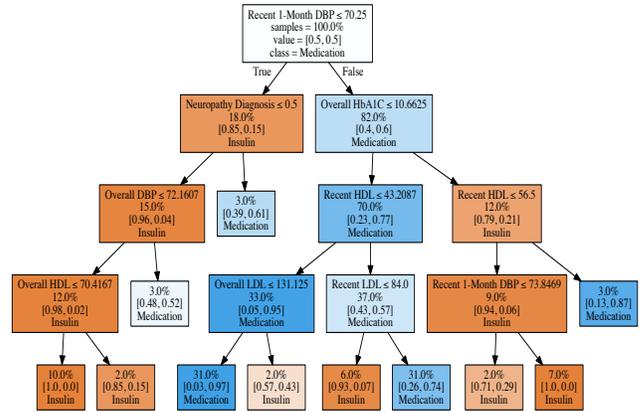
Fig. 5. Optimal ITR Decision Tree Estimated from O-learning

learning is shown to be superior than Q-learning on the training data. One reason of the improvement is that the treatment by feature variable interactions could be masked when fitting Q-learning to predict outcomes, especially in the presence of large number of features and interaction terms in complex real world setting. O-learning, in contrast, directly compares value function in (1) under alternative treatment choices for each individual and thus does not suffer from such limitation.

The estimated ITR provides insights on the effects of patients' characteristics on the outcome and the optimal choice of treatment. However, due to the non-experimental feature of EHR data, the estimated ITR needs to be interpreted with caution. Here, propensity score weighting is carried out to adjust for bias based on observed features. However, hidden confounding and bias due to unobserved characteristics may still be present. Thus the discovered optimal ITR needs to be confirmed in future randomized trials. Recent research on sequential multiple assignment randomized trials [21] provides experimental methodologies for designing randomized trials to study multiple-stage ITRs or dynamic treatment rule. Current single-stage O-learning can be extended to multiple-stage settings where the treatment strategies are dynamically adapted depending a subject's time-varying feature variables and build optimal treatment sequences. However, a challenge for applications to EHR data is to handle unbalanced random treatment switching time points.

Cohort identification using EHR data can be challenging [22]. More sophisticated approach for identifying true T2D patients from EHR data using both ICD codes and clinical notes is desirable. Currently we used mean laboratory test values as feature variables to learn an ITR. Better temporal abstraction methods may be used to improve accuracy [23]. Lastly, using double-robust augmentation to handle missing data [24] and exploring methods to handle complications from missing-not-at-random mechanisms [25] are of interest.

REFERENCES

[1] M. A. Hamburg and F. S. Collins, "The path to personalized medicine," *New England Journal of Medicine*, vol. 363, no. 4, pp. 301–304, 2010.

[2] F. S. Collins and H. Varmus, "A new initiative on precision medicine," *New England Journal of Medicine*, vol. 372, no. 9, pp. 793–795, 2015.

[3] S. A. Murphy, D. W. Oslin, A. J. Rush, and J. Zhu, "Methodological challenges in constructing effective treatment sequences for chronic psychiatric disorders," *Neuropsychopharmacology*, vol. 32, no. 2, pp. 257–262, 2006.

[4] P. W. Lavori and R. Dawson, "Dynamic treatment regimes: practical design considerations," *Clinical trials*, vol. 1, no. 1, pp. 9–20, 2004.

[5] Y. Zhao, D. Zeng, A. J. Rush, and M. R. Kosorok, "Estimating individualized treatment rules using outcome weighted learning," *Journal of the American Statistical Association*, vol. 107, no. 499, pp. 1106–1118, 2012.

[6] Y. Zhao, D. Zeng, E. Laber, and M. R. Kosorok, "New statistical learning methods for estimating optimal dynamic treatment regimes," *Journal of the American Statistical Association*, p. DOI: 10.1080/01621459.2014.937488, 2014.

[7] M. Qian and S. A. Murphy, "Performance guarantees for individualized treatment rules," *Annals of Statistics*, vol. 39, no. 2, pp. 1180–1210, 2011.

[8] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[9] G. Hripcsak, P. B. Ryan, J. D. Duke, N. H. Shah, R. W. Park, V. Huser, M. A. Suchard, M. J. Schuemie, F. J. DeFalco, A. Perotte *et al.*, "Characterizing treatment pathways at scale using the ohdsi network," *Proceedings of the National Academy of Sciences*, p. 201510502, 2016.

[10] G. Hripcsak, D. J. Albers, and A. Perotte, "Exploiting time in electronic health record correlations," *Journal of the American Medical Informatics Association*, vol. 18, no. Supplement 1, pp. i109–i115, 2011.

[11] D. Scharfstein, A. Rotnitzky, and J. Robins, "Rejoinder to comments on "adjusting for non-ignorable drop-out using semiparametric nonresponse models?"," *Journal of the American Statistical Association*, vol. 94, pp. 1121–1146.

[12] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[13] Y. Liu, Y. Wang, M. Kosorok, Y. Zhao, and D. Zeng, "Robust hybrid learning for estimating personalized dynamic treatment regimens," *Technical Report*.

[14] A. Qaseem, L. L. Humphrey, D. E. Sweet, M. Starkey, and P. Shekelle, "Oral pharmacologic treatment of type 2 diabetes mellitus: a clinical practice guideline from the american college of physicians," *Annals of internal medicine*, vol. 156, no. 3, pp. 218–231, 2012.

[15] W. L. Bennett, N. M. Maruthur, S. Singh, J. B. Segal, L. M. Wilson, R. Chatterjee, S. S. Marinopoulos, M. A. Puhan, P. Ranasinghe, L. Block *et al.*, "Comparative effectiveness and safety of medications for type 2 diabetes: an update including new drugs and 2-drug combinations," *Annals of internal medicine*, vol. 154, no. 9, pp. 602–613, 2011.

[16] S. B. Johnson, "Generic data modeling for clinical rrepositories," *Journal of the American Medical Informatics Association*, vol. 3, no. 5, pp. 328–339, 1996.

[17] H. Sagreiya and R. B. Altman, "The utility of general purpose versus specialty clinical databases for research: warfarin dose estimation from extracted clinical variables," *Journal of biomedical informatics*, vol. 43, no. 5, pp. 747–751, 2010.

[18] R. Pivovarov, D. J. Albers, J. L. Sepulveda, and N. Elhadad, "Identifying and mitigating biases in ehr laboratory tests," *Journal of biomedical informatics*, vol. 51, pp. 24–34, 2014.

[19] G. Hripcsak and D. J. Albers, "Next-generation phenotyping of electronic health records," *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 117–121, 2013.

[20] R. J. Little, "Survey nonresponse adjustments," *International statistical review*, vol. 54, no. 1, p. 3, 1986.

[21] S. A. Murphy, L. Collins, and A. J. Rush, "Customizing treatment to the patient: Adaptive treatment strategies," vol. 88, Supplement 2, no. 0, pp. S1 – S3, 2007.

[22] C. Weng, Y. Li, P. Ryan, Y. Zhang, F. Liu, J. Gao, J. Bigger, G. Hripcsak *et al.*, "A distribution-based method for assessing the differences between clinical trial target populations and patient populations in electronic health records," *Applied clinical informatics*, vol. 5, no. 2, pp. 463–479, 2014.

[23] A. R. Post, T. Kurc, R. Willard, H. Rathod, M. Mansour, A. K. Pai, W. M. Torian, S. Agravat, S. Sturm, and J. H. Saltz, "Temporal abstraction-based clinical phenotyping with eureka!" in *AMIA Annual Symposium Proceedings*, vol. 2013. American Medical Informatics Association, 2013, p. 1160.

[24] J. M. Robins, "Optimal structural nested models for optimal sequential decisions," in *Proceedings of the second seattle Symposium in Biostatistics*. Springer, 2004, pp. 189–326.

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