Statistics and Its Interface Volume 16 (2023) 503–515

Estimating individualized treatment rules for multicategory type 2 diabetes treatments using electronic health records

Jitong Lou*, Yuanjia Wang, Lang Li, and Donglin Zeng

In this article, we propose a general framework to learn optimal treatment rules for type 2 diabetes (T2D) patients using electronic health records (EHRs). We first propose a joint modeling approach to characterize patient's pretreatment conditions using longitudinal markers from EHRs. The estimation accounts for informative measurement times using inverse-intensity weighting methods. The predicted latent processes in the joint model are used to divide patients into a finite of subgroups and, within each group, patients share similar health profiles in EHRs. Within each patient group, we estimate optimal individualized treatment rules by extending a matched learning method to handle multicategory treatments using a one-versus-one approach. Each matched learning for two treatments is implemented by a weighted support vector machine with matched pairs of patients. We apply our method to estimate optimal treatment rules for T2D patients in a large sample of EHRs from the Ohio State University Wexner Medical Center. We demonstrate the utility of our method to select the optimal treatments from four classes of drugs and achieve a better control of glycated hemoglobin than any one-size-fits-all rules.

Keywords and phrases: Electronic health records, Individualized treatment rules, Latent process, Machine learning, Multicategory treatments, Type 2 diabetes.

1. INTRODUCTION

Type 2 diabetes (T2D) is the most common type of diabetes which causes millions of people to suffer from severe diabetes-related complications such as heart attacks, stroke, blindness, and kidney failure [40]. To treat T2D, American Diabetes Association (ADA) recommended to use metformin monotherapy as the initial treatment and select additional therapies based on patient-centered considerations [4]. A treatment guideline from the United Kingdom also suggested metformin as the first-line drug, unless it is contraindicated or not tolerated [31]. Palmer et al. [37] summarized 301 clinical trials (1.4 million patient-month) in which metformin and other 8 available classes of glucose-lowering drugs were compared. This meta analysis reported that when compared with other drugs given as monotherapy, metformin only had better or similar effects on managing glycated hemoglobin (HbA1c) levels among adults with T2D. Nevertheless, there was no significant difference in all-cause mortality or other complications between any glucose-lowering drugs alone or combined. There is a lack of conclusive evidence for the best T2D management strategy from clinical trials.

With the emergence of large-scale electronic systems such as electronic health records (EHRs), which usually contain patient demographics, vital signs, laboratory test results, medications, diagnoses, and medical insurances documented at the point of care, there has been an increasing trend of using EHRs as an observational database to study T2D treatment patterns in real world practices. For example, Montvida et al. [34] selected 1.02 million adults with T2D from the U.S. Centricity Electronic Medical Records and concluded that, from 2005 to 2016, first-line use increased for metformin (60% to 77%) and decreased for sulfonylureas (20% to 8%). Canivell et al. [6] used a 5-year-EHR for 15,205 patients with T2D from the SIDIAP database and assessed glycemic controls after treatment intensification. Compared to experiments such as randomized controlled trials (RCTs), observational studies use real-world information from larger patient populations and contain a longer duration of observations, and thus may offer valuable complements to RCTs. Studies using large-scale EHRs may reflect real-world patterns of treatment pathways which can neither be conducted nor observed in RCTs [22]. More importantly, EHRs provide a great opportunity to study heterogeneity of treatment responses in a large population so that we can learn optimal individualized treatment rules (ITRs) for T2D patients to fulfill the goal of precision medicine, a medical paradigm that utilizes individual patient's characteristics such as demographics, lab test results, and genetic information, to optimize treatments [16].

There has been intensive methods development for precision medicine in the fields of statistics and machine learning over the last decade [32]. These methods include regression model-based methods such as Q-learning [45, 36, 38], A-learning [35, 39], regret-regression [19], and subgroup analysis [14, 27, 15]. Through directly optimizing ITR-related value functions, Zhao et al. [48] proposed an outcome

*Corresponding author.

weighted learning approach that converted the estimation of ITRs to a weighted classification problem. Similar methods were later developed in contrast weighted learning [43] and augmented outcome weighted learning [28]. More recently, Wu et al. [47] proposed a matched learning approach, called M-learning, to learn ITRs based on pairs of patients who shared similar pre-treatment health profiles. This approach was demonstrated to be more robust than weighting methods.

However, the above methods are confronted by the following challenges when applied to EHRs. First, characterizing individual patient's pretreatment condition is difficult since their health markers measured over time are multivariate and the measurements can be continuous (e.g., lab measures), binary (e.g., disease diagnoses) or counts (e.g., number of medications). Moreover, these measurements are taken at patient's clinical encounters which potentially depend on their underlying health status. Thus, not accounting for informative measurement patterns of health markers may cause selection bias [17, 18]. Second, there are often many observed treatment options and patient's propensity to receive one specific treatment is complex and heterogeneous, which may not be captured by parametric models. Furthermore, there presents substantial heterogeneity among patients in terms of treatments and outcomes that need to be accounted for when learning ITRs. Standard weighting methods suffer from numerical instability due to low representation of patients with some treatments.

In this work, to address the challenges in EHRs, we propose a general framework for learning ITRs for T2D patients and use one concrete dataset as an example to demonstrate the framework. Specifically, we propose a multivariate longitudinal model to model the time-trajectory of different types of health markers through a generalized exponential family of distributions, while accounting for their dependence through a latent multivariate Gaussian temporal process. We also adopt inverse-intensity weighting to adjust for potential informative times of measurements. Through the joint models, we can identify several T2D patient subgroups using clustering algorithm to summarize patient's health profiles based on their pre-treatment EHRs. To learn ITRs within each subgroup, we create a few classes of treatments and apply random forests to estimate treatment propensity scores. Finally, to handle the challenge of multiple treatments, we extend the matched learning method in [47] to multicategory treatments. Particularly, we develop a one-versus-one matched learning method to estimate ITRs. The derived rules are further validated through cross-validation.

The remaining part of this paper is organized as follows. In Section 2, we provide the details of the proposed models and learning methods. In Section 3, we demonstrate an implementation of our methods to EHRs from the EHR and Information Warehouse of the Ohio State University Wexner Medical Center (OSU-WMC). In Section 4, we describe the estimated ITRs for T2D patients and compare with the observed treatments in EHRs. Concluding remarks are given in Section 5.

2. A GENERAL FRAMEWORK TO LEARN OPTIMAL ITRS USING EHRS

We use A, $\boldsymbol{Z}$ and R to denote a T2D patient's treatment at a decision time (referred to as time zero), pre-treatment features, and reward outcome, respectively. We assume no unobserved confounding and stable unit treatment value assumption, which are two crucial assumptions to allow using the EHRs for learning the optimal treatment rules. The first assumption implies that the treatment assignment is independent of potential outcomes given $\boldsymbol{Z}$, so there will not be any hidden bias due to unobserved confounding; while the second assumption implies that there is no treatment interference between the patients. The assumptions are not testable due to the observational nature of the EHRs but may be plausible if $\boldsymbol{Z}$ contains sufficient information about why each patient received one particular treatment and one patient's response does not depend the other patients' treatments or responses. Under these assumptions, it is known that the optimal ITR is a function mapping $\boldsymbol{Z}$ to A's domain and it is given as the treatment that yields the maximum value of $\mathbb{E}(R|\boldsymbol{Z}, A = a)$. Many methods have been developed to estimate such optimal ITR using RCTs, but our goal is to instead use EHRs to estimate the optimal ITR.

Data from EHRs consist of patient's health marker measurements, for example, body mass index, cholesterol level, and HbA1c for T2D patients, as well as received medications, at clinical encounters over a span of calendar time windows. Time zero is usually set to be the index date when a patient received treatment A, and the reward outcome, R, is a pre-defined measure indicating disease improvement since time zero (for example, HbA1c reduction within 6 months after taking the treatment). However, obtaining a reasonable set of feature variables for $\boldsymbol{Z}$ is challenging, since they not only include patient's demographics (age, gender, race), but more importantly, should reflect patient's preconditions that are useful for the treatment decision. The latter must be extracted from patient's longitudinal health markers before time zero.

In the following sections, we first use the method in our earlier work [29] to extract patient's pre-treatment health profiles using EHRs that will be included as feature variables for learning ITRs. We then propose a matching-based learning algorithm to estimate optimal treatment rules that will maximize patient's outcomes.

2.1 Characterizing patient's pre-treatment health conditions using multivariate latent processes

Given the heterogeneity among patients in EHRs, it is important to characterize patient's pre-treatment conditions

based on longitudinal marker measurements in EHRs. We present methods to handle two challenges: the first challenge is that patterns of measurement time points may depend on patient's underlying health status and thus are informative; the second challenge is that health markers are of mixed types and collected at different time points.

We firstly model the measurement time pattern of health markers using a counting process, and we want to adjust for the informative measurement patterns. Suppose EHR data are obtained from n patients. Following our earlier work [29], we let $\boldsymbol{X}_i$ be m-dimensional pre-treatment covariates for the ith patient. Among p health markers, let $Y_{ik}(t)$ denote the measurement of the kth marker at time t, where t is time since the first record date of EHRs but no later than the date of time zero. The total number of observations up to time t for marker k can be represented by a counting process $N_{ik}(t) \equiv \sum_j I(t_{ikj} \le t)$, where $I(\cdot)$ is the indicator function, and t_{ikj} is the jth observation time point of kth health marker for the ith patient. We model the intensity of $N_{ik}(t)$ conditional on all observed data by time t as

$$\begin{aligned}(1)\qquad &\mathbb{E}\left[dN_{ik}(t)|\boldsymbol{Y}_i(s), s \le t\right] \\ &= \lambda_k(t)\exp\left\{\boldsymbol{X}_i^T\boldsymbol{\gamma}_k + \boldsymbol{L}_{ik}^T(t)\boldsymbol{\eta}_k\right\}dt,\end{aligned}$$

where $\lambda_k(t)$ is a baseline intensity function, $\boldsymbol{L}_{ik}(t)$ is a vector of observed health history up to time t, and $\boldsymbol{\gamma}_k$ and $\boldsymbol{\eta}_k$ are intensity parameters.

Next, we model the longitudinal trajectory of each health marker using generalized linear models with latent processes. Particularly, we assume that $Y_{ik}(t)$ follows a generalized exponential family model with conditional density given latent variables as

$$f_{ik}(y;\theta_{ik},\phi_{ik},t) = \exp\left\{\frac{y\theta_{ik}(t) - b_k(\theta_{ik}(t))}{a_k(\phi_{ik}(t))} + c_k(y,\phi_{ik}(t))\right\},$$

where $\theta_{ik}(t)$ is the patient- and process-specific canonical parameter, and $\phi_{ik}(t)$ is the dispersion parameter. Let $\theta_{ik}(t) = g_k(\mu_{ik}(t))$, where $g_k(\cdot)$ is the canonical link, and $\mu_{ik}(t)$ is the mean of $Y_{ik}(t)$. To capture patient heterogeneity and the dependence among $\{Y_{i1}(t),\ldots,Y_{ip}(t)\}$, we assume

$$(2)\quad g_k(\mu_{ik}(t)) = \widetilde{\boldsymbol{X}}_i^T\boldsymbol{\beta}_k(t) + \epsilon_{ik}(t), \quad \boldsymbol{\epsilon}_i(t) \sim \mathcal{N}_p\left(\boldsymbol{0}, \boldsymbol{\Omega}(t)\right),$$

where $\widetilde{\boldsymbol{X}}_i^T = [1, \boldsymbol{X}_i^T]$. $\boldsymbol{\epsilon}_i(t) = \{\epsilon_{i1}(t),\ldots,\epsilon_{ip}(t)\}^T$ are multivariate normal latent effects and they are independent of $\boldsymbol{X}_i$. The covariance matrix $\boldsymbol{\Omega}(t)$ reflects the dependence among the processes at time t and is allowed to be fully nonparametric.

In (1) and (2), we need to estimate parameters $\boldsymbol{\gamma}_k$, $\boldsymbol{\eta}_k$, $\boldsymbol{\beta}_k(t)$, and $\boldsymbol{\Omega}(t)$. Clearly, intensity parameters $\boldsymbol{\gamma}_k$ and $\boldsymbol{\eta}_k$ can be estimated by fitting a standard Andersen-Gill proportional intensity model [2]. We denote the coefficient estimators as $\widehat{\boldsymbol{\gamma}}_k$ and $\widehat{\boldsymbol{\eta}}_k$. Subsequently, we define $d\widetilde{N}_{ik}(s) = dN_{ik}(s)/\exp\left\{\boldsymbol{X}_i^T\widehat{\boldsymbol{\gamma}}_k + \boldsymbol{L}_{ik}^T(s)\widehat{\boldsymbol{\eta}}_k\right\}$ and use $d\widetilde{N}_{ik}(s)$ in the parameter estimation of $\boldsymbol{\beta}_k(t)$ and $\boldsymbol{\Omega}(t)$. As the standardized intensity of $\widetilde{N}_{ik}(s)$ no longer depends on each patient's covariates, we can remove the heterogeneity of informative measurement time points among patients. The rationale of using this inverse-intensity weighting to adjust for selection bias is similar to inverse-probability weighting [26, 18], whereas the latter is used to balance the treatment groups but here we use $d\widetilde{N}_{ik}(s)$ to standardize the measurement time processes.

To estimate $\boldsymbol{\beta}_k(t)$ and $\boldsymbol{\Omega}(t)$, we employ a kernel-weighted local estimation that was proposed by Cao et al. [7]. Sparse data is a common challenge for implementing EHRs data. In parameter estimation, if a time point has limited data, kernel functions can assign weights to the data nearby and pool information across times and patients to handle the sparse data. Specifically, we solve the kernel-weighted local estimating equation (3) for each health marker to estimate $\boldsymbol{\beta}_k(t)$:

$$\begin{aligned}(3)\quad \boldsymbol{0} &= U_{n,k}(\boldsymbol{\beta}_k(t)) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int K_{h_{1n}}(s-t)\widetilde{\boldsymbol{X}}_i[Y_{ik}(s) - E_{ik}(t)]d\widetilde{N}_{ik}(s),\end{aligned}$$

where $E_{ik}(t) = \mathbb{E}\left[g_k^{-1}\left(\widetilde{\boldsymbol{X}}_i^T\boldsymbol{\beta}_k(t) + \epsilon_{ik}(t)\right)\middle|\boldsymbol{X}_i\right]$, $K_h(u) = K(u/h)/h$ with $K(u)$ being a symmetric kernel function and h is its bandwidth. $K(u)$ often takes the Epanechnikov kernel or Gaussian kernel. Noting that $E_{ik}(t)$ serves as the first moment of $Y_{ik}(t)$, we estimate $\boldsymbol{\beta}_k(t)$ by minimizing the distance between observed values of health markers and their first moments. Denote the estimators obtained by solving (3) as $\widehat{\boldsymbol{\beta}}_k(t)$.

Next, for each element in $\boldsymbol{\Omega}(t) = (\sigma_{kl}(t))$, we estimate them by plugging $\widehat{\boldsymbol{\beta}}_k(t)$ in and solving the following kernel-weighted local estimating equations:

$$\begin{aligned}(4)\quad 0 &= U_{n,k,l}(\sigma_{kl}(t)) \\ &= \frac{1}{n^2}\iint \widetilde{K}_{h_{2n}}(s-t, s'-t)\left[Y_{ik}(s)Y_{il}(s') - E_{ikl}(t)\right] \\ &\quad \times d\widetilde{N}_{ik}(s)d\widetilde{N}_{il}(s'),\end{aligned}$$

where the double integration excludes $s = s'$ if $k = l$,

$$\begin{aligned}&E_{ikl}(t) \\ &= \mathbb{E}\left[g_k^{-1}\left(\widetilde{\boldsymbol{X}}_i^T\widehat{\boldsymbol{\beta}}_k(t) + \epsilon_{ik}(t)\right)g_l^{-1}\left(\widetilde{\boldsymbol{X}}_i^T\widehat{\boldsymbol{\beta}}_l(t) + \epsilon_{il}(t)\right)\middle|\boldsymbol{X}_i\right],\end{aligned}$$

and $\widetilde{K}_h(u_1,u_2) = \widetilde{K}(u_1/h, u_2/h)/h^2$. Here, $\widetilde{K}(u_1,u_2)$ is a bivariate kernel function, usually taken to be the product of univariate Epanechnikov or Gaussian kernel, and h is its bandwidth. In (4), $E_{ikl}(t)$ plays the role of the second moment of $Y_{ik}(t)$. Denote the solved estimators to $\widehat{\sigma}_{kl}(t)$.

We provide the Gauss-Hermite quadrature method [1] for calculating $E_{ik}(t)$ and $E_{ikl}(t)$ in Section S.4 of the supplementary material. The bandwidths of kernel functions in

(3) and (4) are determined using a data-driven approach in Cao et al. [7]. In Lou et al. [29], the proposed estimators $\widehat{\boldsymbol{\beta}}_k(t)$ and $\widehat{\sigma}_{kl}(t)$ are shown to be consistent and asymptotically normal, and numerical studies demonstrate the good performance of the proposed estimation method on finite samples.

Finally, to characterize patient's pre-treatment health status and identify informative patient subgroups, we compute the similarity distance between each pair of patients and then perform a hierarchical clustering based on the distance matrix. Specifically, to account for between-marker and time-varying dependence, we define a Mahalanobis distance [30], which is based on the estimated latent processes, as

$$S_{ij} = \left\{ \sum_{t=t_1}^{t_N} \left[\widehat{\boldsymbol{\epsilon}}_i(t) - \widehat{\boldsymbol{\epsilon}}_j(t)\right]^T \widehat{\boldsymbol{\Omega}}^{-1}(t) \left[\widehat{\boldsymbol{\epsilon}}_i(t) - \widehat{\boldsymbol{\epsilon}}_j(t)\right] \right\}^{1/2}, \tag{5}$$

where $t_1, ..., t_N$ are pre-specified equally spaced time points, $\widehat{\boldsymbol{\epsilon}}_i(t) = \mathbb{E}\left[\boldsymbol{\epsilon}_i(t) \middle| \boldsymbol{Y}_i(t), \widehat{\boldsymbol{\beta}}_k(t), \widehat{\boldsymbol{\Omega}}(t)\right]$, and $\boldsymbol{Y}_i(t)$ is imputed by last-value carried forward if there is no measurement at time t. The numeric method for calculating $\widehat{\boldsymbol{\epsilon}}_i(t)$ can be found in Section S.4 of the supplementary material.

The major merit of subgrouping patients is patients in the same group have similar health profiles that are distinct from other groups. The homogeneity of patients in the same group can thus alleviate the confounding bias when estimating optimal treatment rules. Because the Mahalanobis distance naturally accounts for the between-marker correlation, the use of such distance can effectively remove the redundant information regarding the patient's health status. Moreover, the use of $\boldsymbol{\epsilon}(t)$, rather than $\boldsymbol{Y}$, unifies the different types of health markers and result in a metric that is comparable among patients. Otherwise, it is unreasonable to represent the distance between two patients using a continuous and a binary health features. We choose the hierarchical clustering because it is a powerful approach to identify homogeneous, interpretable groups of the patients. Thus, even though the original health markers are measured irregularly and are of very different data types, our joint models enable one to combine them using the latent processes on the same scales and account for dependence over time. With the estimated subgroups from clustering, the subsequent ITRs will be estimated for each subgroup separately.

2.2 Matched learning for multicategory treatments

When estimating ITRs for binary treatments, Wu et al. [47] showed that M-learning could outperform other commonly used methods for observational databases. Thus, we generalize M-learning to handle multicategory treatments which are commonly seen in EHRs. First, in each patient subgroup $s \in \{1, \ldots, S\}$ that was identified before, the comparison among a total of K treatments can be converted to $K(K-1)/2$ comparisons between two treatments, which can be integrated using the one-versus-one method to yield an optimal ITR for all treatments.

Specifically, for each patient i in each subgroup s, let $\boldsymbol{Z}_i$ denote the baseline covariate $\boldsymbol{X}_i$ and some additional pre-treatment health marker information, for example, the average BMI in the past year. We let A_i and R_i be the treatment at time zero and the reward outcome post-treatment, respectively. For each treatment pair (u, v), let $T_i = 1$ if $A_i = u$ and $T_i = -1$ if $A_i = v$. Assume there are $N^s_{u,v}$ patients who received treatment u or v in group s. Antonelli et al. [3] and Wu et al. [47] proposed a doubly robust matching method to improve the efficiency of matching methods. This method uses not only covariates but also propensity scores and prognostic scores, denoted by $\pi(\boldsymbol{Z}_i) \equiv P(T_i = 1|\boldsymbol{Z}_i)$ and $\psi(\boldsymbol{Z}_i) \equiv \mathbb{E}[R_i|\boldsymbol{Z}_i]$, respectively, to create matched sets. Thus, for the ith patient, the improved matched set, denoted by M_{is}, has an expression as follows:

$$M_{is} = \{j : A_j \neq A_i, d(\boldsymbol{H}_j, \boldsymbol{H}_i) \leq \delta\}, \tag{6}$$

where $d(\cdot, \cdot)$ is a distance function, δ is a threshold which may vary with i, and $\boldsymbol{H}_i = \left(\boldsymbol{Z}_i, \widehat{\pi}(\boldsymbol{Z}_i), \widehat{\psi}(\boldsymbol{Z}_i)\right)$. In our implementation, we use random forests to perform a multicategory classification for T_i given $\boldsymbol{Z}_i$ to obtain $\widehat{\pi}(\boldsymbol{Z}_i)$ and use gradient boosting machines to estimate R_i given $\boldsymbol{Z}_i$ to obtain $\widehat{\psi}(\boldsymbol{Z}_i)$. As suggested by [3] and [47], the doubly robust method may lead to optimal treatment rules even if the model for the propensity score, or the model for the prognostic score is misspecified, but not both, and including prognostic scores was empirically shown to perform better than the methods without using them.

Following [47], we adopt a weighted support vector machine (SVM) with weights for estimating ITR when comparing treatments u and v. Specifically, we minimize the following objective function

$$V^s_{u,v}(f; g) = (N^s_{u,v})^{-1} \sum_{i \in \text{group } s} |M_{is}|^{-1} \sum_{j \in M_{is}} g(R_j - R_i) \times \phi(-f(\boldsymbol{Z}_i) T_i \text{sign}(R_j - R_i)) + \lambda^s_{u,v} \|f\|_{\mathcal{H}_k}, \tag{7}$$

where $|M_{is}|$ is the size of matched set M_{is}, $g(\cdot)$ is a function determining the weights in the weighted SVM, $\phi(x)$ is the hinge loss given by $\max(1 - x, 0)$, $f(\cdot)$ is a function such that the decision rule $D(\boldsymbol{Z}) = \text{sign}(f(\boldsymbol{Z}))$, $\lambda^s_{u,v}$ is a tuning parameter, and $\mathcal{H}_k$ is a reproducing kernel Hilbert space with a kernel function $k(\cdot, \cdot)$. Using the weight $g(R_j - R_i)$ ensures that the estimated treatment rule is driven by comparing the pairs of patients who have large outcome differences. Using a weighted SVM to minimize $V^s_{u,v}(f; g)$, we obtain $D^{*s}_{u,v}$, the optimal ITR for comparing treatment u to v in group s. Similarly, for the remaining treatment pairs, we estimate the corresponding decision rules. Thus, for the ith patient in group s, we derive the

optimal ITR, $D^{*s}(\boldsymbol{Z}_i)$, as the majority vote recommended by $\left\{D^{*s}_{u,v}(\boldsymbol{Z}_i) : u, v \in \{1, \ldots, K\}, u \neq v\right\}$.

In the above learning algorithm, the tuning parameters are chosen using cross-validation. The ITRs estimated from a training sample are evaluated using an independent testing sample by calculating an empirical value function defined as

$$\frac{\sum_{i\in\text{test sample}} \mathbb{I}(A_i = \widehat{D}^{*s}(\boldsymbol{Z_i}))R_i/\widehat{P}(A_i|\boldsymbol{Z}_i)}{\sum_{i\in\text{test sample}} \mathbb{I}(A_i = \widehat{D}^{*s}(\boldsymbol{Z_i}))/\widehat{P}(A_i|\boldsymbol{Z}_i)}, \tag{8}$$

where $\widehat{P}(A_i|\boldsymbol{Z}_i)$ is obtained from the propensity score estimation.

We further illustrate the significance of the latent group information to the proposed M-learning method using a simulation study, which can be found in the Section S.1 of the supplementary material.

3. IMPLEMENTATION USING OSU-WMC EHRS

3.1 Data preparation

The EHRs from OSU-WMC contain demographics, laboratory test measures, vital signs, and diagnosis codes for 58,490 patients diagnosed with T2D between 2011 and 2018. We set time for the treatment decision for each patient (time zero) as the last clinical encounter in year 2016 when he or she received T2D medications. This choice was based on two facts in order to learn ITRs from data with limited time periods: we needed a sufficient time window of the longitudinal history before time zero to precisely characterize patient's health status and subgroups; and we needed a reasonable follow-up period after time zero to precisely calculate the outcome variable, the HbA1c level at 6 months after time zero.

In the EHRs, there were four health markers $Y_{ik}(t)$ associated with T2D: systolic blood pressure (SBP), HbA1c, high-density lipoproteins (HDL), and body mass index (BMI). After checking normal ranges for the health markers [41, 46, 4], we removed missing, duplicated, and extreme entities such as SBP$\geq$250 mmHg, HbA1c$\leq$3 or $\geq$20%, HDL $\leq$0 or $\geq$120 mg/dL, and BMI$\leq$10 or $\geq$60 kg/m^2. In addition, we created a binary variable to denote whether diabetic drugs were prescribed at a clinical encounter (labeled as DD) and a continuous variable as the logarithm of the number of medications prescribed at each clinical encounter (labeled as logMed). Both variables were considered as an important indication of T2D patient's comorbidity status. Therefore, our analysis included one binary longitudinal marker (DD) and other 5 continuous health markers over time (SBP, HbA1c, HDL, BMI, logMed). We required there must be at least one measurement for at least one marker is available before time zero. With this restriction, there were a total of 8,456 subjects with 497,763 longitudinal records before time zero date and they were used to learn patient's pre-treatment subgroups using the method in Section 2.1. Among these patients, 53.43% were female, 62.03% were white, and their ages in years ranged from 17.89 to 100.84 with a mean of 59.67. The number of records for SBP, HbA1c, HDL, BMI, and DD was 9.8, 2.3, 4.0, 2.2, 14.6, and 29.3, respectively, when averaged over all patients.

The medications at time zero were considered to be treatments, A_i, for learning ITRs. There were 3,978 types of medications observed in the EHRs, and we classified the medications to either the 163 diabetic drugs for T2D [11] or the remaining non-T2D medications before further grouping them. As stated in Section 1, metformin is commonly considered as the first-line drug for T2D and it may have a better control of HbA1c levels than other T2D drugs given as monotherapy. On the other hand, there is some evidence that basal insulin also serves as an important T2D treatment. Thus, we compared four classes of treatments: metformin monotherapy, insulin monotherapy, other T2D monotherapy or combinations of other T2D drugs, and combinations of at least two classes of treatment among the aforementioned three classes. We referred the third class as "other T2D drugs" and the fourth class as "multiple T2D drugs". Thus, A_i was one of four treatments including metformin, insulin, other T2D drugs and multiple T2D drugs.

We were interested in the treatment effects on reducing HbA1c level after time zero. For the ith patient, we constructed the outcome variable R_i as the expected HbA1c level 6 months after the date of time zero. In particular, we first collected all available HbA1c measures from lab tests, which were conducted from time zero to up to one year after. For each patient, we performed a linear interpolation model as $\alpha_i+\beta_i(t_{ij}-t_{i,\text{baseline}})$, where t_{ij} was the date (in years) for the jth measurement for patient i, $t_{i,\text{baseline}}$ was the date for time zero, and α_i and β_i were respectively the intercept and slope for the patient's trajectory. Finally, based on the least-square estimates, we defined the outcome as the expected HbA1c level at 6 month after treatment for each patient, which was given as $R_i = \widehat{\alpha}_i + 0.5\widehat{\beta}_i$. In OSU-WMC data, only 5,458 patients had at least two HbA1c measurements during the year after time zero so their outcomes could be calculated. Furthermore, we excluded 333 patients whose estimated slope coefficient was either greater than 5 or less than -5, which were not sensible clinically.

To construct the feature variables, $\boldsymbol{Z}_i$, we first used the proposed method in Section 2.1 to obtain subgroups of all patients using all available health markers before time zero. More specifically, when fitting the joint models, the covariates entering the intensity model (1) for the measurement times included the demographics and time-dependent covariates indicating whether there was any measurement of longitudinal marker k in the past 6 months and if there was, what was the average marker value. Using about 2-year data prior to time zero, we estimated the intensity parameters $\boldsymbol{\gamma}_k$ and $\boldsymbol{\eta}_k$, along with $\boldsymbol{\beta}_k(t)$ and $\boldsymbol{\Omega}(t)$ at 25 equally spaced time points (in days), where the time length between two

consecutive time points were 30 days. Given $\widehat{\boldsymbol{\beta}}_k(t)$ and $\widehat{\boldsymbol{\Omega}}(t)$, in this particular application, we integrated health markers over time by calculating $\widehat{\boldsymbol{\epsilon}}_i(t)$ through the 20-point Gauss-Hermite quadrature. Using the between-patient similarity matrix, we performed a hierarchical clustering analysis on all 8,456 patients. To determine the optimal number of patient groups, we calculated the point-biserial correlation (PBC) coefficient [20, 33] using R package `WeightedCluster` [42]. PBC measures the capacity of the clustering to reproduce the distance matrix. We estimated the clustering quality for $2, 3, \ldots, 10$ groups, and, finally, we set the number of group to be 5, which maximized the absolute values of PBC. Section S.2 of the supplementary material presents the dendrogram of Mahalanobis distances between patients and discusses the consistency of the identifying patient subgroups.

The derived groups using our models represented patient's chronic preconditions, so they might not capture patient's most recent health status before the treatment. Therefore, we also included in $\boldsymbol{Z}_i$ the average values of the health markers during the most recent year before time zero. Consequently, $\boldsymbol{Z}_i$ consisted of the derived group membership, the average values of SBP, HbA1c, HDL, BMI, DD, and logMed in the past one year before time zero, as well as age, gender and race variables. Finally, we had data of $(A_i, \boldsymbol{Z}_i, R_i)$ from 5,125 patients for learning optimal ITRs.

3.2 Learning ITRs using proposed method

Figure 1 describes the flow chart of our proposed framework, along with methods used at each step. As illustrated in Figure 1, we used EHRs before time zero (last clinical encounter in year 2016) to fit joint models and learn subgroups for these patients. After using the 12-month-data of HbA1c after time zero to define the outcome, we applied the multicategory matched learning method in Section 2.2 to estimate optimal ITRs in each subgroup.

Before estimating propensity scores, prognostic scores, and ITRs, we normalized all the continuous variables to alleviate the bias introduced by scaling. Propensity scores,

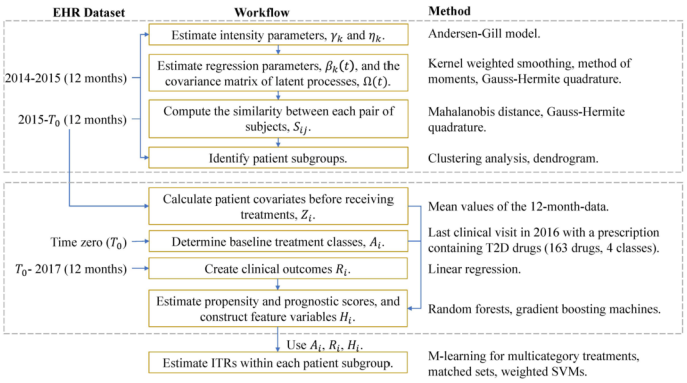

Figure 1. Flow chart of learning optimal ITRs using OSU-WMC EHRs.

$\pi(\boldsymbol{Z}_i)$, were estimated by a 10-fold cross-validation random forest with 3 repeats. The misclassification rates on the whole training data were 3.1%, $<$ 0.001%, 16.4%, 11.0% and 15.2% for each of groups 1 to 5, respectively. The 5th and 95th percentiles of estimated propensities are around 0.01 and 0.80. To avoid extreme weights in the calculation of value functions, we truncated probabilities less than 1% or greater than 80%. Similarly, prognostic scores, $\psi(\boldsymbol{Z}_i)$, were estimated by a gradient boosting model with 5,000 trees of which maximum depth was 4 in each patient subgroup. The model provided a good fit to the clinical outcome with a mean square error less than 10^{-4}. The most important covariate in estimating both propensity and prognostic scores is the recent one-year HbA1c.

In the doubly robust matching step in Section 2.2, we used the Euclidean metric as the distance function $d(\cdot, \cdot)$ in (6) and set δ such that only the 1-nearest neighbors of patients were included into their matched sets (pairs). In other words, for the ith patient, among other patients with the same gender and race but different baseline treatments, we searched for the patient who has the closest Euclidean distance to them in terms of demographics variables, recent measurements of health markers, and two estimated scores. Since the M-learning model was designed to maximize $\mathbb{E}(R|\boldsymbol{Z}, A)$, we plugged the negative values of expected HbA1c levels into the estimation of ITRs to control for the glucose level. We let $g(R_j - R_i) = |R_j - R_i|$ and applied the radial basis function (RBF) kernel to optimize the objective function (7) in the M-learning model, and 2-fold cross-validation with 100 repeats were used to learn optimal ITRs. The tuning parameter for the cost of constraint violation was selected from $\{2^k : k = 0, \pm 1, \ldots, \pm 15\}$ using 2-fold cross-validation. We calculated the bandwidth parameter of each RBF kernel according to a data-driven method which could be implemented using the `kernlab` package [23] in R.

3.3 Comparison with alternative methods

Given the optimal ITR estimated from the proposed method, we used (8) in Section 2.2 to compute empirical HbA1c values and compare the optimal ITR with the four universal rules for each class of the treatments and Q-learning. In Q-learning, we used the same health features and treatments as those in the proposed method, and we applied the support vector regression with RBF kernel to obtain parameter estimates of Q-learning. Furthermore, we evaluated the performance of M-learning and Q-learning without using the latent group information. Lastly, we explored the feasibility of using simpler health features – such as the Charlson comorbidity score [8] and the last observation of health markers – to summarize the pre-treatment health condition of patients and replace the latent group information. We computed the Charlson scores using diagnosis codes in 12-month EHR data prior to baseline treatment dates. Similar to the proposed method, we implemented the aforementioned variations of M-learning and Q-learning

with 100 repeats using R packages `kernlab`, `caret`, and `DynTxRegime` [21]. We selected tuning parameters of the support vector regression model and kernel function by cross-validations. We also incorporated estimated propensity and prognostic scores to all the alternative methods. All continuous features were standardized before parameter estimation.

4. ANALYSIS RESULTS OF OSU-WMC EHRS

4.1 Identified latent subgroups

We identified 5 subgroups in the EHR datasets and patients in each cluster have similar health profiles. Figure 2 shows the averages of normalized health marker measurements for patients in each cluster. The value in each cell is averaged across all patients and clinical visits for the corresponding health marker and patient subgroup. A higher value of HDL and a lower value of HbA1c, HDL, and BMI indicate a healthier status. The value of DD and logMed do not directly reflect the disease state. However, a lower value indicates that physicians tend to prescribe less medications to this group of patients, and thus a less severe state. We compared these values to the sample average of each health marker. A healthier T2D status is indicated in blue and a severe condition is indicated in red.

Group 5 is comprised of 1,231 patients. All the health markers convey the information that this group of people is relative healthier. Compared to other groups, patients in this group did not take excessive number of medications. Group 2 contains 3,360 patients whose HDL is slightly lower than the average. This suggests that they might have some difficulties in controlling the cholesterol level. However, other health markers reflect a relatively healthy status of patients in this group. The SBP of 737 patients in group 1 is apparently higher than the average, while other health markers are below or around the averages. This pattern indicate that these patients might have hypertension and a moderate status of T2D. Group 3 have 2,446 patients and their BMIs are above the average. Besides BMI, the value of HDL in this group represent a bad signal as well. Thus, patients in group were at a moderately severe state. For the 682 patients in group 4, almost all the health markers show the most severe severe state of T2D. In particular, their HbA1c level is much higher. Another interesting fact is physicians had prescribed more-than-average diabetic drugs and ancillary medications to this group of patients, but their diseases were not controlled well. This result implies that this group might not have received the optimal treatments, and, therefore, we would focus on this group in the following analysis.

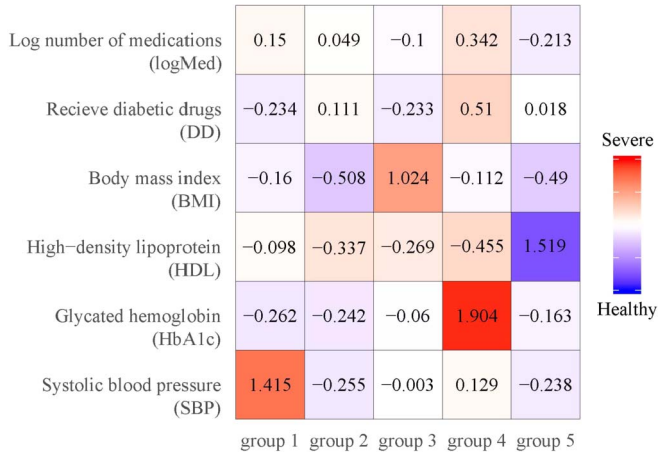

Figure 2. Averages of normalized measurements by health markers and patient subgroups using 24-month-data before baseline treatment dates. Red: more severe status than the overall sample average in terms of a health marker; Blue: healthier status than the overall sample average in terms of a health marker; White: overall sample average status in terms of a health marker.

4.2 ITRs for multicategory treatments

As described in Section 3.1, we used 5,125 patients out of 8,456 patients in the finalized dataset for learning optimal ITRs. Before estimating optimal ITRs on the whole dataset, we compared the performance of the proposed method with the alternative methods in Section 3.3. The results of 100 cross-validation repetitions are displayed in Figure 3, and the summary statistics are listed in Table 1.

In general, the empirical HbA1c value of estimated ITRs is lower than any universal rules (i.e., "one-size-fits-all" rules) in any of the 100 repetitions. Compared to other versions of M-learning and Q-learning, the proposed method has lower average empirical values in the majority of cases as well. For example, in group 4, the weighted mean outcomes in (8) are 8.479, 9.199, 8.866, and 9.121 for patients prescribed metformin only, insulin only, other T2D monother-

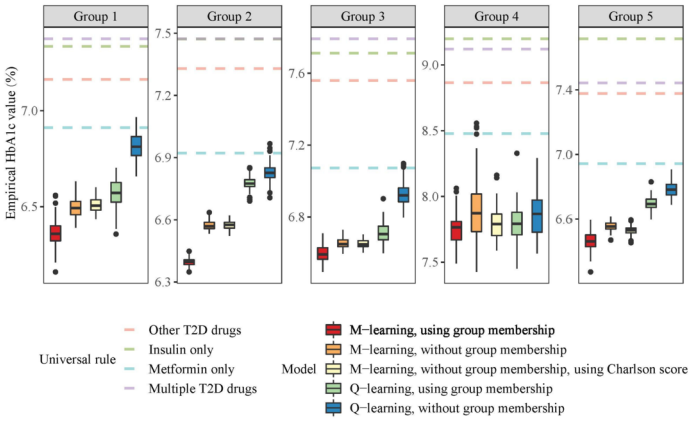

Figure 3. The empirical value function for the expected HbA1c level (%) using 2-fold cross-validation with 100 repeats (a lower value is more beneficial).

Table 1. The comparison of ITRs using M-learning and Q-learning in terms of the empirical value function for the expected HbA1c level (%) using 2-fold cross-validation with 100 repeats (a lower value is more beneficial)

Group	Fit model	Mean (sd)	Median (Q1,Q3)
Group 1	**M-learning, using group membership**	**6.362 (0.070)**	**6.359 (6.322, 6.400)**
	M-learning, without group membership	6.496 (0.048)	6.493 (6.458, 6.527)
	M-learning, without group membership, using Charlson score	6.508 (0.037)	6.505 (6.482, 6.536)
	Q-learning, using group membership	6.596 (0.073)	6.572 (6.524, 6.626)
	Q-learning, without group membership	6.812 (0.067)	6.812 (6.766, 6.865)
Universal rules: Metformin only: 6.911, Insulin only: 7.335, Other T2D drugs: 7.162, Multiple T2D drugs: 7.375			
Group 2	**M-learning, using group membership**	**6.397 (0.018)**	**6.398 (6.384, 6.407)**
	M-learning, without group membership	6.572 (0.023)	6.569 (6.557, 6.588)
	M-learning, without group membership, using Charlson score	6.575 (0.021)	6.576 (6.560, 6.588)
	Q-learning, using group membership	6.776 (0.031)	6.775 (6.761, 6.794)
	Q-learning, without group membership	6.826 (0.045)	6.827 (6.804, 6.851)
Universal rules: Metformin only: 6.922, Insulin only: 7.472, Other T2D drugs: 7.329, Multiple T2D drugs: 7.473			
Group 3	**M-learning, using group membership**	**6.597 (0.045)**	**6.592 (6.564, 6.629)**
	M-learning, without group membership	6.654 (0.028)	6.649 (6.635, 6.673)
	M-learning, without group membership, using Charlson score	6.651 (0.023)	6.647 (6.635, 6.668)
	Q-learning, using group membership	6.712 (0.057)	6.705 (6.673, 6.749)
	Q-learning, without group membership	6.927 (0.060)	6.921 (6.884, 6.962)
Universal rules: Metformin only: 7.073, Insulin only: 7.711, Other T2D drugs: 7.560, Multiple T2D drugs: 7.792			
Group 4	**M-learning, using group membership**	**7.752 (0.117)**	**7.765 (7.670, 7.810)**
	M-learning, without group membership	7.906 (0.234)	7.872 (7.733, 8.018)
	M-learning, without group membership, using Charlson score	7.794 (0.120)	7.790 (7.702, 7.866)
	Q-learning, using group membership	7.794 (0.131)	7.793 (7.708, 7.879)
	Q-learning, without group membership	7.860 (0.155)	7.867 (7.728, 7.972)
Universal rules: Metformin only: 8.479, Insulin only: 9.199, Other T2D drugs: 8.866, Multiple T2D drugs: 9.121			
Group 5	**M-learning, using group membership**	**6.464 (0.057)**	**6.462 (6.429, 6.502)**
	M-learning, without group membership	6.554 (0.028)	6.554 (6.538, 6.575)
	M-learning, without group membership, using Charlson score	6.529 (0.026)	6.530 (6.514, 6.544)
	Q-learning, using group membership	6.697 (0.041)	6.694 (6.672, 6.724)
	Q-learning, without group membership	6.783 (0.045)	6.782 (6.750, 6.815)
Universal rules: Metformin only: 6.943, Insulin only: 7.717, Other T2D drugs: 7.378, Multiple T2D drugs: 7.443			
Overall	**M-learning, using group membership**	**6.572 (0.024)**	**6.571 (6.559, 6.589)**
	M-learning, without group membership	6.611 (0.020)	6.607 (6.595, 6.624)
	M-learning, without group membership, using Charlson score	6.610 (0.015)	6.610 (6.600, 6.620)
	Q-learning, using group membership	6.774 (0.026)	6.774 (6.758, 6.792)
	Q-learning, without group membership	6.890 (0.040)	6.887 (6.865, 6.913)
Universal rules: Metformin only: 7.208, Insulin only: 7.727, Other T2D drugs: 7.506, Multiple T2D drugs: 7.872			

apy, and multiple treatments, respectively. ITRs estimated by the proposed approach achieve a mean empirical HbA1c value of 7.752, which is lower than that for M-learning without group membership (7.906), M-learning without group membership but using Charlson score (7.794), Q-learning using group membership (7.794), and Q-learning without membership (7.860). The differences in other groups are even greater. Another point worth noting is the M-learning and Q-learning using group memberships consistently perform better than their variations which ignore the group information and/or use the Charlson score. The last finding is that using the most recent observations of health markers can obtain close empirical HbA1c values to the method using the average values of the recent 1-year data. This result may be due to the change in health markers in our EHR data are gentle, and one year is a relatively short time period. In conclusion, the proposed M-learning model outperforms the alternative methods on muticategory treatment recommendations in this application. The latent subgroup information is crucial to the proposed method, and it can alleviate the confounding bias as well as improving the accuracy of the M-learning model.

After evaluating the model performance, we estimated the ITRs using the whole dataset. The distributions of four treatment classes in baseline assignments and ITRs are displayed in Figure 4. Taking group 4 as an example, there are 53 (14.2%) patients who received metformin only; 152 (40.8%) patients received insulin only; 88 (23.6%) received other T2D monotherapy; and 80 (21.4%) received at least two treatments. However, in other patient subgroups, 25% to 35% of patients were assigned metformin monotherapy, insulin monotherapy, and other T2D drugs, respectively.

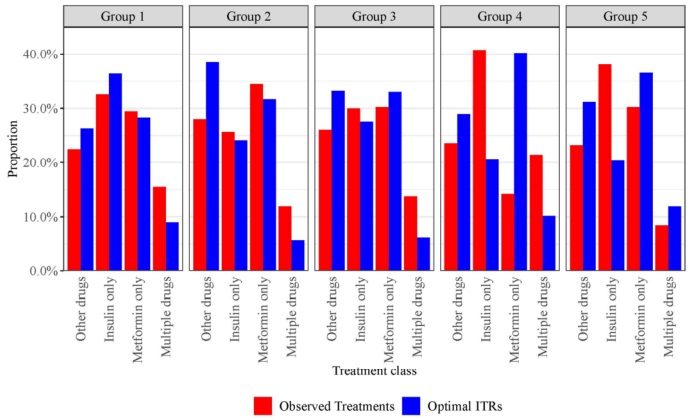

Figure 4. The distribution of observed treatments versus treatments recommended by estimated ITRs within each subgroup.

The proportion of multiple treatments is around 8% to 15%. Thus, the proportion of patients in group 4 receiving metformin monotherapy is much lower than that in other groups, while the normalized HbA1c level of group 4, as shown in Figure 2, is the highest among all groups.

Compared to the observed baseline treatments, the proportion of assigning either metformin monotherapy or other T2D drugs increases in almost every patient group. In contrast, the estimated ITRs suggest to prescribe insulin to less patients than observed. The proportions of metformin monotherapy and other T2D drugs recommended by the ITRs are fairly close; however, in group 4, metformin has a drastic increment in the assignment proportion from 14.2% to 40.2%. Similarly, in group 5, the proportion of insulin monotherapy decreases from 38.1% to 20.4%, together with observable increases in the assignment of metformin monotherapy and other T2D monotherapy. Table S3 in the supplementary file displays the contingency table of observed treatments and the estimated ITRs recommendations. For metformin monotherapy, observed treatments and ITRs are matched by about 50% to 60% times. The proportion of other T2D monotherapy ranges from 40% to 60% across patient groups. Nevertheless, the observed insulin monotherapy merely has about a 30% to 40% match rate with ITRs. Particularly, in group 3, 4 and 5 of which patients have the highest HbA1c measurements, ITRs tend to assign metformin and other T2D monotherapy to more than 65% of patients who originally received insulin.

To further investigate the superiority of estimated ITRs over observed treatments, we stratified the patients by baseline HbA1c levels and summarized the empirical values as shown in Table 2.

These five strata were selected by considering: 1. 6.5% is a stringent goal for glycemic control and it is usually used to separate non-diabetic and diabetic patients. 2. 7.5% is a threshold being used by AACE/ACE to initiate dual therapy, whereas ADA suggests considering dual therapy if a patient's entry HbA1c level is $\geq 9\%$. 3. 10.5% is a common upper limit of HbA1c-based inclusion criteria in clinical trials for T2D patients [44]. Since metformin monotherapy is the most effective universal rule for this data, we also included it in Table 2 for comparison. Some of the summary statistics in Table 2 are missing because, in those strata, there is no or only one patient of whom observed treatment and predicted optimal treatment are matched.

In the overall group, the estimated ITRs outperform observed assignments and metformin monotherapy across the population. For "healthy" people (baseline HbA1c between 3.0% and 6.5%), the mean empirical value of estimated ITRs is around 6 with a decrease of 0.2 from observed assignments and metformin monotherapy. For "relatively healthy" patients (baseline HbA1c between 6.5% and 7.5%), the estimated ITRs can achieve mean empirical values of 6.5 across patient subgroups, indicating an achievement of the stringent goal of HbA1c control. The differences between estimated ITRs and alternatives are around 0.5, which are greater than the standard deviation of emiprical values of estimated ITRs. For patients with the medium level of baseline HbA1c (between 7.5% and 10.5%), the estimated ITRs can decrease the HbA1c to 7.5 for group 4 and below 7 for other groups. Similarly, the differences between estimated ITRs and other two treatment rules are greater than the standard deviation of emiprical values of estimated ITRs, expect for the patients in group 4 with baseline HbA1c between 7.5% and 9%. Lastly, the estimated ITRs can control the empirical value to 6.13 in group 3 and 7.37 in group 4 for patients with baseline HbA1c level $\geq 10.5\%$. In this stratum, metformin monotherapy is much more effective than observed assignments, but it still has a difference of 1.17 in overall, compared to the estimated ITRs.

Bringing all the comparisons together, we can conclude that metformin monotherapy has the optimal effect on HbA1c control, i.e., with the smallest HbA1c level at 6 months, especially for patients with low or moderate HbA1c levels. Whereas, the insulin monotherapy does not have noticeable advantages over other T2D monotherapy, and may even have worse HbA1c management when the baseline HbA1c level is high. Also, the insulin monotherapy may not be optimal to people who are relatively healthy and do not suffer from T2D complications. In this case, physicians may consider to prescribe either metformin monotherapy or other T2D monotherapy. Compared to metformin monotherapy, our ITRs lead to a better control of the HbA1c level around 0.5% for patients with a baseline HbA1c between 6.5% and 7.5%. While the baseline HbA1c is greater than 7.5%, the treatment difference is at least 0.5% and can go up to 3%. As 0.5% is a signal of clinically significant difference [5, 12, 25], this result suggests our method is practical and the treatment effect far exceeds the eight alternative glucose-lowering drugs summarized by a large-scale meta analysis [37]. Therefore, the proposed model can potentially assist in the prescription of T2D treatments for HbA1c management in real world business.

Table 2. Empirical values of estimated ITRs and observed treatments by patient subgroup and baseline HbA1c level

Group	Baseline HbA1c(%)	Sample size	Estimated ITRs	Observed assignments	Metformin only
			Mean empirical value (sd)		
Overall	(3.0, 6.5)	1220	6.07(0.51)	6.28(0.72)	6.27(0.69)
	[6.5, 7.5)	1697	6.59(0.52)	7.14(0.79)	7.06(0.74)
	[7.5, 9.0)	1427	6.82(0.83)	7.91(1.05)	7.70(1.11)
	[9.0, 10.5)	526	7.31(1.08)	9.02(1.52)	8.75(1.60)
	[10.5, 20.0)	255	7.19(1.18)	9.57(1.98)	8.36(1.82)
Group 1	(3.0, 6.5)	126	5.98(0.51)	6.31(0.85)	6.36(0.72)
	[6.5, 7.5)	166	6.48(0.49)	7.13(0.78)	6.96(0.71)
	[7.5, 9.0)	123	6.32(0.83)	7.79(1.02)	7.37(1.37)
	[9.0, 10.5)	28	6.76(0.63)	8.63(1.27)	8.81(1.78)
	[10.5, 20.0)	2	-	10.69(1.11)	10.05(−)
Group 2	(3.0, 6.5)	531	6.07(0.49)	6.29(0.73)	6.39(0.71)
	[6.5, 7.5)	765	6.58(0.49)	7.14(0.78)	7.05(0.74)
	[7.5, 9.0)	561	6.71(0.58)	7.87(1.05)	7.65(0.99)
	[9.0, 10.5)	120	6.67(0.84)	8.96(1.45)	8.65(1.41)
	[10.5, 20.0)	12	-	9.60(1.62)	8.95(0.50)
Group 3	(3.0, 6.5)	345	6.08(0.53)	6.32(0.72)	6.28(0.65)
	[6.5, 7.5)	455	6.60(0.58)	7.16(0.84)	7.11(0.79)
	[7.5, 9.0)	446	6.75(0.72)	7.95(1.07)	7.72(1.16)
	[9.0, 10.5)	176	6.54(1.08)	9.14(1.53)	8.60(1.47)
	[10.5, 20.0)	56	6.13(0.65)	9.04(2.11)	6.66(1.08)
Group 4	(3.0, 6.5)	3	6.38(0.94)	6.64(0.85)	7.03(−)
	[6.5, 7.5)	10	6.51(0.44)	7.13(1.52)	7.67(2.14)
	[7.5, 9.0)	67	7.52(0.95)	8.05(1.25)	8.03(1.34)
	[9.0, 10.5)	135	7.63(0.99)	8.83(1.54)	8.66(1.60)
	[10.5, 20.0)	158	7.37(1.16)	9.64(1.99)	8.80(1.84)
Group 5	(3.0, 6.5)	215	6.09(0.51)	6.16(0.61)	6.17(0.67)
	[6.5, 7.5)	301	6.65(0.44)	7.13(0.69)	7.06(0.54)
	[7.5, 9.0)	230	6.65(0.48)	7.91(0.98)	7.74(1.02)
	[9.0, 10.5)	67	6.52(0.26)	9.21(1.55)	9.63(2.44)
	[10.5, 20.0)	27	-	10.26(1.53)	8.38(1.28)

4.3 Stability of ITRs

To evaluate the stability of estimated ITRs, we followed the rationale that using different but similar training data and checked the consistency of treatment recommendations on a test dataset. Meanwhile, we noticed that, in cross-validation, the whole dataset was randomly split into a training set and a test set, which were different from the original dataset but inherited the data structure. Thus, we counted the majority vote of recommendations of 100 2-fold cross-validated repetitions for each patient when they were assigned to the test set, and then we compared the majority vote of recommendations with the recommendation estimated from the whole dataset. From Table S1, for 61.66% of the patients, the two types of treatment recommendations are matched. The Cohen's Kappa statistic [9] is 0.4361 with an asymptotic standard error of 0.0097, so it indicates a moderate agreement between the two recommendations and partially demonstrates the stability of the estimated ITRs.

5. DISCUSSION

In this work, we propose a general framework to estimate optimal ITRs for multicategory treatments using EHRs. Our first contribution is to create a novel latent process model, which jointly analyzes different types of health markers and accounts for informative measurement patterns. Using patient similarities estimated from the latent process model, we identify subgroups of patients with homogeneous health profiles. The identified patient subgroups show different health patterns, and the cluster membership of patients is an important feature in M-learning to match patients among a more homogeneous pool.

The second contribution is the generalization of M-learning to multicategory treatments. We reduce the problem of selecting the optimal option among multicategory treatments into multiple binary classification problems using the one-versus-one strategy, and we use the majority voting to integrate the results of the binary classification problems. Using this doubly robust multicategory M-learning, which incorporates propensity scores and prognostic scores, we reduce the confounding due to both covariates and recent patterns of health markers. Thus, our approach tackles the challenges in EHRs and takes full advantage of information available from the health markers.

To determine the group membership of new patients, users firstly compute the average of latent processes $\boldsymbol{\epsilon}(t)$ for each identified patient group as centroids. Next, standardize the data of new patients. In particular, each numeric covariate X is subtract by the mean and is divided by the standard deviation of that covariate for the old patient population. Then, use the standardized data, $\widehat{\boldsymbol{\beta}}(t)$ and $\widehat{\boldsymbol{\Omega}}(t)$, and Gauss-Hermite quadrature method (see Section S.4 of the supplementary material) to compute $\widehat{\boldsymbol{\epsilon}}(t)$ of new patients. Finally, select the patient group whose centroid is the closest to $\widehat{\boldsymbol{\epsilon}}(t)$ as the group membership for a new patient.

As shown in Figure 1, we proposed a general analysis pipeline to learn the optimal treatment rules using the EHR data. However, there are certainly alternative ways to use each component of this pipeline. For example, the outcome, HbA1c level at 6 months was interpolated based on a linear function using one-year data since the HbA1c value changes slowly and smoothly. More sophisticated interpolation models such as splines may be useful if the measurements are taken intensively or over multiple years. Moreover, our proposed framework focuses on finding the optimal treatment for the outcome in a short period of time at a single decision point (e.g., at a patient visit). The method is not designed to optimize the long-term outcome over years and after multiple treatment phases. Thus, the proposed work does not account for the delayed effect in sequential decision making or long-term health management. By substituting the value function in (7) with a matching-based value function for multiple stages and using backward learning methods, our method can be extended to estimate dynamic treatment regimes. To further improve the stability of estimated ITRs, we can include additional features into the model and perform model-based feature selection to include more informative confounding factors. On the other hand, when interpretability, but not accuracy, of ITRs is more of interest, there is a class of tree-based models [13, 24, 10] that can be applied to learning ITRs. However, tree-based methods are often highly variable and their performance may be affected by the complexity of EHR data and true decision regimes.

In future research, we will consider additional data resources as they become available. For example, we can utilize patients' diet information or glucose level measurements collected by mobile devices to recommend individualized treatments. Another extension is to take other aspects of T2D control into consideration. For instance, besides lowering the level of HbA1c, we will also consider to balance control of adverse events such as hypoglycemia.

SUPPLEMENTARY MATERIAL

Supplementary material includes additional analyses on simulation and EHRs data. It is available for this article online (http://intlpress.com/site/pub/files/_supp/sii/2023/0016/0004/SII-2023-0016-0004-s001.pdf). The coding examples for the proposed methods and analyses are accessible on the GitHub repositories https://github.com/jitonglou/MultiMlearn and https://github.com/jitonglou/IdSubgroup.

ACKNOWLEDGEMENTS

This research work was supported by the National Institutes of Health grants GM124104, NS073671, and MH117458.

Received 10 February 2021

REFERENCES

[1] Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables. Applied mathematics series.* Dover Publications, New York, NY, USA. MR0415956

[2] Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* **10** 1100–1120. MR0673646

[3] Antonelli, J., Cefalu, M., Palmer, N. and Agniel, D. (2018). Doubly robust matching estimators for high dimensional confounding adjustment. *Biometrics* **74** 1171–1179. MR3908135

[4] American Diabetes Association (2018). Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes–2018. *Diabetes Care* **41** 73–85.

[5] American Diabetes Association (2021). Standards of Medical Care in Diabetes—2021. *Diabetes Care* **44** S73–S150.

[6] Canivell, S., Mata-Cases, M., Real, J., Franch-Nadal, J., Vlacho, B., Khunti, K. and et al. (2019). Glycaemic control after treatment intensification in patients with type 2 diabetes uncontrolled on two or more non-insulin antidiabetic drugs in a real-world setting. *Diabetes, Obesity and Metabolism* **21** 1373–1380.

[7] Cao, H., Zeng, D. and Fine, J. P. (2015). Regression analysis of sparse asynchronous longitudinal data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 755–776. MR3382596

[8] Charlson, M. E., Pompei, P., Ales, K. L. and MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases* **40** 373–383.

[9] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20** 37–46.

[10] Doubleday, K., Zhou, H., Fu, H. and Zhou, J. (2018). An Algorithm for Generating Individualized Treatment Decision Trees and Random Forests. *Journal of Computational and Graphical Statistics* **27** 849–860. MR3890875

[11] Drugs. com (2019). Drugs used to treat diabetes, type 2. https://www.drugs.com/condition/diabetes-mellitus-type-ii.html.

[12] National Institute for Health and Care Excellence (UK) (2015). *Type 2 diabetes in adults: management*. National Institute for Health and Care Excellence (UK), London, UK.
[13] Foster, J. C., Taylor, J. M. G. and Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine* **30** 2867–2880. MR2844689
[14] Foster, J. C., Taylor, J. M. G. and Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Stat. Med.* **30** 2867–2880. MR2844689
[15] Fu, H., Zhou, J. and Faries, D. E. (2016). Estimating optimal treatment regimes via subgroup identification in randomized control trials and observational studies. *Stat. Med.* **35** 3285–3302. MR3528258
[16] Ginsburg, G. S. and Phillips, K. A. (2018). Precision medicine: from science to value. *Health Affairs* **37** 694–701.
[17] Haneuse, S. (2016). Distinguishing selection bias and confounding bias in comparative effectiveness research. *Medical care* **54** e23.
[18] Haneuse, S. and Daniels, M. (2016). A general framework for considering selection bias in EHR-based studies: what data are observed and why? *EGEMS (Washington DC)* **4** 1203.
[19] Henderson, R., Ansell, P. and Alshibani, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics* **66** 1192–1201. MR2758507
[20] Hennig, C. and Liao, T. F. (2010). Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification Technical Report, Research Report No. 308, Department of Statistical Science, University College London.
[21] Holloway, S. T., Laber, E. B., Linn, K. A., Zhang, B., Davidian, M. and Tsiatis, A. A. (2020). DynTxRegime: Methods for Estimating Optimal Dynamic Treatment Regimes R package version 4.9. MR4175018
[22] Hripcsak, G., Ryan, P. B., Duke, J. D., Shah, N. H., Park, R. W., Huser, V. and et al. (2016). Characterizing treatment pathways at scale using the OHDSI network. *Proceedings of the National Academy of Sciences of the United States of America* **113** 7329–7336.
[23] Karatzoglou, A., Smola, A., Hornik, K. and Zeileis, A. (2004). kernlab – An S4 package for kernel methods in R. *Journal of Statistical Software* **11** 1–20.
[24] Laber, E. B. and Zhao, Y. Q. (2015). Tree-based methods for individualized treatment regimes. *Biometrika* **102** 501–514. MR3394271
[25] Lameijer, A., Fokkert, M. J., Edens, M. A., Slingerland, R. J., Bilo, H. J. G. and van Dijk, P. R. (2020). Determinants of HbA1c reduction with FreeStyle Libre flash glucose monitoring (FLARE-NL 5). *Journal of Clinical & Translational Endocrinology* **22** 100237.
[26] Lin, H., Scharfstein, D. O. and Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 791–813. MR2088782
[27] Lipkovich, I., Dmitrienko, A., Denne, J. and Enas, G. (2011). Subgroup identification based on differential effect search—a recursive partitioning method for establishing response to treatment in patient subpopulations. *Stat. Med.* **30** 2601–2621. MR2815438
[28] Liu, Y., Wang, Y., Kosorok, M. R., Zhao, Y. and Zeng, D. (2018). Augmented outcome-weighted learning for estimating optimal dynamic treatment regiments. *Stat. Med.* **37** 3776–3788. MR3869154
[29] Lou, J., Wang, Y., Li, L. and Zeng, D. (2021). Learning latent heterogeneity for type 2 diabetes patients using longitudinal health markers in electronic health records. *Statistics in Medicine* **40** 1930–1946. MR4229830
[30] De Maesschalck, R., Jouan-Rimbaud, D. and Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems* **50** 1–18.
[31] McGuire, H., Longson, D., Adler, A., Farmer, A. and Lewin, I. (2016). Management of type 2 diabetes in adults: summary of updated NICE guidance. *British Medical Journal* **353** i1575.
[32] Mesko, B. (2017). The role of artificial intelligence in precision medicine. *Expert Review of Precision Medicine and Drug Development* **2** 239–241.
[33] Milligan, G. and Cooper, M. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50** 159–179.
[34] Montvida, O., Shaw, J., Atherton, J., Stringer, F. and Paul, S. (2018). Long-term trends in antidiabetes drug usage in the U.S.: real-world evidence in patients newly diagnosed with type 2 diabetes. *Journal of the American Medical Association* **41** 69–78.
[35] Murphy, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 331–366. MR1983752
[36] Murphy, S. A. (2005). A generalization error for Q-learning. *J. Mach. Learn. Res.* **6** 1073–1097. MR2249849
[37] Palmer, S. C., Mavridis, D. and Nicolucci, A. (2016). Comparison of clinical outcomes and adverse events associated with glucose-lowering drugs in patients With type 2 diabetes: a meta-analysis. *Journal of the American Medical Association* **316** 313–324.
[38] Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. MR2816351
[39] Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics. Lect. Notes Stat.* **179** 189–326. Springer, New York. MR2129402
[40] Roglic, G. (2016). WHO global report on diabetes: a summary. *International Journal of Noncommunicable Diseases* **1** 3-8.
[41] Stone, N. J., Robinson, J. G., Lichtenstein, A. H., Bairey Merz, C. N., Blum, C. B., Eckel, R. H. and et al. (2014). 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association task force on practice guidelines. *Journal of the American College of Cardiology* **63** 2889–2934.
[42] Studer, M. (2013). WeightedCluster library manual: a practical guide to creating typologies of trajectories in the social sciences with R Technical Report, LIVES Working Papers 24.
[43] Tao, Y. and Wang, L. (2017). Adaptive contrast weighted learning for multi-stage multi-treatment decision-making. *Biometrics* **73** 145–155. MR3632360
[44] Thrasher, J. (2017). Pharmacologic Management of Type 2 Diabetes Mellitus: Available Therapies. *The American Journal of Medicine* **130** S4–S17.
[45] Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning* **8** 279–292.
[46] Whelton, P. K., Carey, R. M., Aronow, W. S., Casey Jr., D. E., Collins, K. J., Dennison Himmelfarb, C., and et al. (2018). 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. *Journal of the American College of Cardiology* **71** e127–e248. MR3293618
[47] Wu, P., Zeng, D. and Wang, Y. (2020). Matched learning for optimizing individualized treatment strategies using electronic health records. *J. Amer. Statist. Assoc.* **115** 380–392. MR4078470
[48] Zhao, Y., Zeng, D., Rush, A. J. and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. MR3010898

Jitong Lou
135 Dauer Drive
3101 McGavran-Greenberg Hall
Chapel Hill, NC 27599
USA
E-mail address: jitong@live.unc.edu

Yuanjia Wang
722 West 168th Street
Rm 210
New York, NY 10032
USA
E-mail address: yw2016@columbia.edu

Lang Li
250 Lincoln Tower
1800 Cannon Drive
Columbus, OH 43210
USA
E-mail address: li.8958@osu.edu

Donglin Zeng
135 Dauer Drive
3103B McGavran-Greenberg Hall
Chapel Hill, NC 27599
USA
E-mail address: dzeng@email.unc.edu