HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2021 December 01.

Published in final edited form as:

Biometrics. 2020 December ; 76(4): 1075–1086. doi:10.1111/biom.13288.

On Using Electronic Health Records to Improve Optimal Treatment Rules in Randomized Trials

Peng Wu[1], **Donglin Zeng**[2], **Haoda Fu**[3], **Yuanjia Wang**[1]

[1]Department of Biostatistics, Mailman School of Public Health, Columbia University

[2]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill

[3]Eli Lilly and Company

Summary:

Individualized treatment rules (ITRs) tailor medical treatments according to patient-specific characteristics in order to optimize patient outcomes. Data from randomized controlled trials (RCTs) are used to infer valid ITRs using statistical and machine learning methods. However, RCTs are usually conducted under specific inclusion/exclusion criteria, thus limiting their generalizability to a broader patient population in real-world practice settings. Because electronic health records (EHRs) document treatment prescriptions in the real world, transferring information in EHRs to RCTs, if done appropriately, could potentially improve the performance of ITRs, in terms of precision and generalizability. In this work, we propose a new domain adaptation method to learn ITRs by incorporating information from EHRs. Unless we assume that there is no unmeasured confounding in EHRs, we cannot directly learn the optimal ITR from the combined EHR and RCT data. Instead, we first pre-train "super" features from EHRs that summarize physician treatment decisions and patient observed benefits in the real world, as these are likely to be informative of the optimal ITRs. We then augment the feature space of the RCT and learn the optimal ITRs by stratifying by super features using subjects enrolled in RCT. We adopt Q-learning and a modified matched-learning algorithm for estimation. We present heuristic justification of our method and conduct simulation studies to demonstrate the performance of super features. Finally, we apply our method to transfer information learned from EHRs of patients with type 2 diabetes to learn individualized insulin therapies from RCT data.

Keywords

Domain adaptation; Machine learning; Matching; Personalized medicine; Real-world evidence

Data Availability Statement

The EHR data that support the findings in this paper are not publicly available due to privacy or ethical restrictions.

Supporting Information

Web Appendices, Tables, and Figures referenced in Section 3, Section 4, Section 5 and Section 6 are available with this paper at the Biometrics website on Wiley Online Library.

1 Introduction

Personalized medicine based on individualized treatment rules (ITRs) has been proposed as an alternative to the universal, “one-size-fits-all” treatment strategy that is currently employed by physicians and other healthcare providers (Collins and Varmus, 2015). Personalized medical decision making is becoming closer to reality due in part to recent advances in modern technologies that produce a large amount of patient-specific data. In particular, patient electronic health records (EHRs), which contain longitudinal data on medical history, laboratory measures, and disease diagnoses, provide rich information about each patient comorbidity, treatment history, and health outcomes in a real-world setting. The best approach for incorporating such real-world information to learn ITRs represents a pressing research challenge in today’s modern era of personalized medicine.

Recently developed methods for optimizing ITRs aim to assist healthcare providers to prescribe the right therapy to the right patient at the point of care (Collins and Varmus, 2015). These methods include regression-based approaches that estimate the conditional mean of treatment outcomes or their contrasts given treatments and patient feature variables, such as Q-learning (Qian and Murphy, 2011), A-learning (Murphy, 2003), and G-computation (Moodie et al., 2007). Alternatively, one can directly optimize a value function or its equivalence to learn optimal ITRs. These methods include outcome-weighted learning (O-learning, Zhao et al. (2012)), a doubly-robust version of O-learning (Liu et al., 2018) and matching-based M-learning (Wu et al., 2018).

Many existing methods for ITRs are developed in the context of randomized controlled trials (RCTs), for which the estimated ITRs are consistent. While RCTs have high internal validity, they are usually conducted under specific inclusion/exclusion criteria, thus potentially limiting the generalizability of the resultant ITRs to a broader real-world patient population. In fact, evidence of treatment efficacy from RCTs may not translate to a general real-world patient population, even with patients who have the same condition as those included in the trial (Haynes, 1999). Real-word data such as that available in EHRs document medical practices in usual care settings. The incorporation of EHR evidence into RCTs should improve generalizability when learning ITRs from RCTs. However, challenges due to potentially unmeasured confounding in EHRs and differences between patients in RCTs and those in EHRs make incorporation of EHR evidence to RCTs complicated.

In this paper, we propose a novel method to learn ITRs from RCTs by borrowing information from EHRs. First, we pre-train two “super” feature mappings from the EHR data: one is a probability feature that estimates the propensity that a physician treatment prescription, as documented in the EHR, is optimal when predicted from the EHR data; the other is a benefit feature that reflects the observed real-world benefit under the optimal treatment. Because treatment prescriptions in EHR data are likely to be at least somewhat beneficial (e.g., better than random assignment) (Wallace et al., 2016), the “super” features learned from EHR data can be potentially informative of the optimal ITRs for patients in real-world medical practice settings, as well as those in RCTs. Next, we augment the feature space of the RCT data by super features when learning final ITRs using only subjects enrolled in the RCT. To enhance the signal from super features, we propose stratified

learning to estimate the optimal ITRs separately within each stratum defined by the super features. Particularly, we apply Q-learning and a modified M-learning to estimate the optimal ITRs.

We provide heuristic reasoning of several advantages of the proposed method. First, because the final optimal ITRs are estimated using RCT data, they remain consistent due to the virtue of randomization, regardless of whether EHR super features are used and whether there is unmeasured confounding in the EHR data. Second, because the super features are informative of the treatment benefit and optimal ITRs, our learning method, which is based on stratification by super features, should yield more precise estimation of ITRs compared to methods that do not include super features (e.g., super features constrain the search space of RCT to a more informative feature domain). Finally, the super features are learned using EHR data, so they are likely to be correlated with the true optimal treatments for the EHR population. Although our optimal ITRs are fit to subjects enrolled in an RCT, they are partially directed by the super features and thus potentially generalizable to the EHR population. As a note, the proposed method is related to a general framework of transfer learning in machine learning, which refers to domain adaptation by allowing different assumptions in different domains with a single task in hand (Pan and Yang, 2010). Reweighting or data transformation methods are commonly used techniques to handle challenges in domain adaptation (Zhang et al., 2013). However, weighting or transformation approach does not directly relate to the goal of maximizing value function for ITRs. In contrast, our proposed method extracts the most informative features (super feature) from an external data source (i.e., EHRs) that is directly related to the goal of optimizing value function to improve ITR estimation on the RCT data. Our ultimate goal is to learn an ITR that could also be more generalizable to a lager target population.

We conduct simulation studies to demonstrate performance of domain adaptation learning compared to methods that do not use information from EHRs. We also apply our method to derive super features from EHRs of patients with type 2 diabetes (T2D) to learn individualized insulin therapies from a T2D RCT, namely, the DURABLE study (Fahrbach et al., 2008). We show that directly applying ITRs learned from EHRs to RCT data performs even worse than a one-size-fits-all strategy on RCT samples, while the proposed super features and domain adaptation lead to improvement in the value function on the RCT. We conclude the paper with discussion and potential future extensions of our work.

2 Method to Improve ITRs by Incorporating Information from EHRs

2.1 Learning the optimal ITR using RCT data

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

$A = \mathcal{D}(X)$. When the treatment assignment mechanism is known (e.g., for an RCT), this expectation can be equivalent expressed as $V(\mathcal{D}) = E\left[\frac{R}{\pi(A,X)} I\{A = \mathcal{D}(X)\}\right]$, where $\pi(a, x)$ is the randomization probability for $A = a$ given $X = x$. Thus, the goal of learning the optimal ITR is to find $\mathcal{D}^*(\cdot)$ that maximizes $V(\mathcal{D})$. The corresponding empirical value function using n i.i.d. observations collected from an RCT can be expressed as

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

Various methods have been developed to estimate the optimal ITR. They can be unified under the framework of maximizing some surrogate of function of (1), in which different weights and loss functions are used to replace R and $I(A = \mathcal{D}(X))$, respectively. For example, in terms of the surrogate weight for R, Q-learning replaces R in (1) by the estimated treatment benefit based on a regression model (Qian and Murphy, 2011), doubly-robust O-learning replaces R by a doubly-robust augmented residual of $R - E(R|X)$ (Liu et al., 2018), and M-learning uses R subtracting the averaged outcomes for matched subjects who receive opposite treatments (Wu et al., 2018). For the choice of surrogate loss, doubly-robust O-learning and M-learning use the hinge-loss to replace the zero-one loss in (1).

2.2 Domain adaptation to improve learning ITRs

Suppose that in addition to the RCT data, we also observe data from patients in EHRs, including feature variables, received treatments, and real-world outcomes. Our goal is to extract information from the EHR data and include it when learning the optimal ITR from the trial data. We refer to this information extraction as “domain adaptation” from EHR to RCT; this framework is illustrated in Figure 1.

Unless we assume that there is no unmeasured confounding in the EHR data, we cannot directly learn the optimal ITR from the combined EHR and RCT data. Instead, we pre-train useful feature mappings from the EHR data to augment the feature space of the RCT data. Because physician treatment decisions, as documented in EHRs, are likely to carry clinical insights and deemed to be beneficial to patients in consideration of all available options, they are likely closer to optimal than random assignments (as done in RCTs). Thus, features that can summarize physician prescription patterns are informative of the optimal treatment for a patient. Furthermore, because physicians may prescribe a treatment based on many considerations, including efficacy, risk of complications, and cost, their prescription patterns may not be sufficient when the goal is to learn an ITR to maximize a specific outcome (e.g., efficacy). Thus, the observed benefit under the optimal treatment, as predicted from EHR data is also useful.

In the first step of domain adaptation, we capture information available in EHRs, but not in RCTs (i.e., a physician’s judgment of which treatment is beneficial and a patient’ observed benefit). We create feature mappings predictive of optimal ITRs and observed benefits, referred to as “super features”. The first super feature is an optimal treatment probability measure, denoted as H_1 that estimates the probability that a physician prescribes a particular

treatment a_0 as optimal based on a patient's covariates in the EHR. This variable captures both the clinician-based treatment decision documented in EHRs and algorithm-based optimal treatment computed by ITRs.

To construct this feature, we pre-train an ITR from EHR data, denoted as $\widehat{\mathcal{D}}(X_c)$, from a common set of pre-treatment covariates in the EHR and RCT data, denoted as X_c, and X_c is a subset of RCT covariates X using the methods described in Section 3. Next, we fit a classification model (e.g., by random forest classifier or logistic regression) to estimate the probability of a physician prescribing a_0 given X_c using the subset of patients in the EHRs who received optimal treatments, as predicted by $\widehat{\mathcal{D}}(X_c)$. That is, we obtain the prediction function of the first super feature, $\hat{f}_1(X_c) = \hat{P}\left(A = a_0 \mid A = \widehat{\mathcal{D}}(X_c), X_c\right)$, which is a function of X_c. For information transfer, we apply the fitted function to subjects in the RCT with observed feature variables x_c to obtain their predicted optimal treatment probability as the first super feature $H_1 = \hat{f}_1(x_c)$.

The second super feature is a benefit feature, denoted as H_2, which measures a patient's observed gain or loss on an EHR outcome under the optimal treatment. That is, H_2 is the expected difference in the outcome when a subject receives the optimal treatment compared to the outcome under the non-optimal treatment. To obtain H_2, we first fit a random forest regression model for outcomes under the optimal treatment to estimate $E\left(R \mid A = \widehat{\mathcal{D}}(X_c), X_c\right)$, using the subset of patients in the EHR who received the optimal treatment $\widehat{\mathcal{D}}(X_c)$. Similarly, we fit a model for outcomes under the non-optimal treatment, $E\left(R \mid A = -\widehat{\mathcal{D}}(X_c), X_c\right)$, using patients who received non-optimal treatments. The prediction function of the second super feature is the benefit, $f_2(X_c) = E\left(R \mid A = \widehat{\mathcal{D}}(X_c), X_c\right) - E\left(R \mid A = -\widehat{\mathcal{D}}(X_c), X_c\right)$, which is a function of X_c. We apply the fitted function to predict benefit for subjects in the RCT as their second super feature $H_2 = \hat{f}_2(x_c)$.

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

The procedure of our method is summarized in Algorithm 1 and Figure 1.

Algorithm 1 The Algorithm for Domain Adaptation Learning

Step 0. Use machine learning methods in Section 3 to pre-train ITRs, denoted as $\widehat{\mathcal{D}}(X_c)$, from patients in the EHRs.

Step 1. Construct H_1 and H_2 for patients in the EHRs by:

1a. Learn $H_1 = \widehat{P}(A = a_0 | A = \widehat{\mathcal{D}}(X_c), X_c)$ by random forest classification among patients in the EHRs who have received the optimal treatment predicted by $\widehat{\mathcal{D}}(X_c)$, where a_0 is a pre-specified treatment.

1b. Learn $H_2 = \widehat{E}(R|A = \widehat{\mathcal{D}}(X_c), X_c) - \widehat{E}(R|A = -\widehat{\mathcal{D}}(X_c), X_c)$ by random forest regression separately for patients in the EHRs who have received optimal treatments, as predicted by $\widehat{\mathcal{D}}(X_c)$, and non-optimal treatments.

1c. Predict H_1, H_2 on subjects in the RCT using their features X_c.

Step 2. Learn the final optimal ITR for the RCT using the methods described in Section 3 by:

(1) use all RCT subjects and features (X, H_1, H_2); OR

(2) stratify RCT subjects by H_1 and/or H_2, and learn separate ITRs for each strata.

2.3 Heuristic reasoning of the domain adaptation learning

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

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

where H represents (H_1, H_2), and $\tilde{R}$ is the surrogate outcome for R depending on which learning method is used. Note that $\tilde{R} = E[R \mid A = 1, X, H] - E[R \mid A = -1, X, H]$ in Q-learning, $\tilde{R} = R$ in the original O-learning, $\tilde{R} = R - E[R \mid X, H]$ in an augmented doubly-robust O-learning, and $\tilde{R}$ is $R(a) - E[R|A = -a, X, H]$ in M-learning, where a is the treatment actually received. Correspondingly, let $\widehat{\mathcal{D}}^*$ be the estimated final optimal ITR using the empirical data, as given in Algorithm 1. Let $\mathcal{D}^0$ be the optimal ITR that maximizes $E[\check{R} I(A = \mathcal{D}(X))/\pi(A, X)]$ without super features, where $\check{R}$ is similar to $\tilde{R}$ except that the former is calculated without stratification by H and the latter is calculated with stratification, and let $\widehat{\mathcal{D}}^0$ be its estimator.

Case I.—When there is no structural assumption on $\mathcal{D}(x)$, because H is a function of X, it is clear that $\mathcal{D}^*(x)$ maximizes $V(\mathcal{D})$. Thus, both $\mathcal{D}^*(x)$ and $\mathcal{D}^0$ yield the same optimal rule. Furthermore, following the same derivation as in Liu et al. (2018), we have

$$V\left(\widehat{\mathcal{D}}^*\right) - V(\mathcal{D}^*) \leqslant \left(E\tilde{R}^2\right)^{\alpha} a_n + b_n, \quad V\left(\widehat{\mathcal{D}}^0\right) - V(\mathcal{D}^*) \leqslant \left(E\check{R}^2\right)^{\alpha} a_n + b_n,$$

where α is a constant depending on the underlying distribution and the dimension of X, and a_n and b_n are constants that depend on n and the geometric noise index, respectively, in the underlying distribution. Because $\tilde{R}$ is obtained and centered in each stratum, but $\check{R}$ is not, we expect $E\tilde{R}^2 \leqslant E\check{R}^2$, and we conclude that the domain adaptation ITR, $\widehat{\mathcal{D}}^*$, leads to a more efficient value than the ITR without H. In addition, when H is more predictive of R, which is likely to hold because H contains the predictive benefit feature H_2 using the EHR data, more efficiency gain is expected when using the domain adaptation ITR.

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

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

where $R^{(1)}$ and $R^{(-1)}$ denote the potential outcomes for treatment 1 and −1 in the RCT population, respectively. In the EHR population, the treatments may act similarly in terms of qualitative effects (i.e., sign of the treatment effect), but the magnitude of the treatment effects (i.e., benefits) may not be as large as what is seen in the RCT population, which is well managed and performed under controlled conditions. As such, we assume only that the direction of the conditional treatment effect is the same for subjects in the RCT and patients in the EHRs. That is, the potential outcomes for the EHR population, denoted by $\tilde{R}^{(a)}$, $a = 1$, −1, satisfy

$$E\left(R^{(1)} - R^{(-1)} \mid X\right) = E\left(\tilde{R}^{(1)} - \tilde{R}^{(-1)} \mid X\right)g(X), \tag{2}$$

where $g(X) > 0$ is an arbitrary positive function and reflects the heterogeneous ratios of the treatment effects across subgroups defined by X between the two populations. Hence, the domain adaptation optimal rule maximizes

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

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

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

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

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

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

Note that because (2) implies that the conditional treatment effects in the RCT and EHR data have the same signs, the nonparametric rules learned from the RCT data are also the optimal rules for the EHR data. Thus, generalizability only becomes an issue when parametric rules (e.g., linear rules), which may not be the same as the sign of the true conditional treatment effect, are learned from the RCT data (Case II discussion is also relevant to parametric rules).

In summary, we obtain the following conclusions:

1. If ITRs are learned nonparametrically, then the domain adaptation rule leads to the optimal treatment rule with a more efficient value estimation, as compared to the optimal rule without the EHR super features H.
2. If ITRs are learned for some parametric class (e.g., linear rules), then the domain adaptation rule leads to a higher value than the one without H; moreover, ITRs are more generalizable to the EHR population when this population has different magnitudes of treatment effects, when compared to the RCT population.

Remarks: The above conclusion (1) shows that because H is constructed from features common to both the RCT and EHR data, then if the ITRs are learned nonparametrically from RCTs, we may not expect a higher value function than that obtained when not using H (especially with a large sample). However, by carefully constructing H to be predictive of the optimal rules and observed outcomes from an external source (i.e., EHRs), the precision of value estimation can be improved. Moreover, from conclusion (2), when estimating parametric ITRs, the inclusion of H may lead to higher value and precision, with greater generalizability (under the assumptions).

3 Algorithms for Estimating ITRs

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

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

The second learning algorithm is M-learning (Wu et al., 2018), with an improvement we develop in this work. This choice of algorithm is based on the fact that it is a general method that includes O-learning as a special case. In matched learning (Wu et al., 2018), the value function is re-expressed in a matching-based alternative form as

$$\frac{1}{n}\sum_i \frac{1}{|\mathscr{M}_i|}\sum_{j \in \mathscr{M}_i}\left[I\{R_j \geq R_i, \mathscr{D}(X_i) = -A_i\} + I\{R_j \leq R_i, \mathscr{D}(X_i) = A_i\}\right]u(|R_j - R_i|), \tag{3}$$

where $\mathcal{M}_i$ is a matching set for subject i that consists of subjects with similar covariates, defined under a suitable distance metric, but opposite treatments. The function $u(\cdot)$ is a monotonically increasing function to weight different subject outcomes in our implementation, we chose $u(x) = x$. The rationale of (3) is that for two subjects who are matched in confounders or propensity scores of treatments, but are observed to receive different treatments, the subject with a higher outcome should be more likely to have received the optimal treatment. Furthermore, doubly-robust matching through using prognostic scores in M-learning can further improve efficiency of ITR estimation, and it is expected to perform better, especially when the treatment assignment is imbalanced, as in observational patient data.

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

$$V_n(f) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} I(A_i \neq A_j)k_{a_n}\{s(X_i, X_j)\}I\{f(X_i)A_i\text{sign}(R_j - R_i) \geqslant 0\}u(|R_j - R_i|), \tag{4}$$

where $k_{a_n}(\cdot)$ is a kernel function with bandwidth a_n, e.g., with Gaussian kernel $k_{a_n}\{s(X_i, X_j)\} = \exp(-a_n\|X_i, X_j\|^2)$, where $\|\cdot\|$ denotes some suitable distance function. Note that in (4), subject pairs with different treatments and more similar feature variables or confounding variables and larger differences in clinical outcomes will receive higher weights. Kernel M-learning thus extracts information from all possible pairs, but adjusts their contribution to the objective function based on their similarity in confounding variables and differences in clinical outcomes. To solve the optimization problem based on $V_n(f)$ in (4), one can replace the zero-one loss by another surrogate loss function. Specifically, when replacing by the hinge-loss, the optimization problem is transformed to a weighted support vector machine (SVM) for kernel-weighted pairs

$$L_n(f) = n^{-2}\sum_{i=1}^{n}\sum_{\{j: A_i \neq A_j\}}\phi\{-f(X_i)A_i\text{sign}(R_j - R_i)\}k_{a_n}\{s(X_i, X_j)\}u(|R_j - R_i|) + \lambda_n\|f\|_{\mathcal{X}_K}, \tag{5}$$

where $\phi(x) = (1 - x)_+$, λ_n is a tuning parameter, and $\mathcal{X}_K$ is a reproducing kernel Hilbert space with kernel function $K(\cdot, \cdot)$. The dual problem of (5) is a quadratic programming problem that can be solved by quadratic programming packages (e.g., through a weighted SVM with matched pairs), similar to Wu et al. (2018). We include the derivations in the Supporting Information A1.

4 Simulation Studies

We generated two separate domains of data source to simulate a scenario involving data from both an RCT and EHRs. To accommodate treatment benefit heterogeneity across patient populations observed in real-world data, we considered the underlying true optimal ITR with a piece-wise linear tree structure. Specifically, outcome data were simulated as $C1 : R = \eta(X) + \phi(X) * A + \epsilon, \epsilon \sim N(0, 0.25)$, where $\phi(X) = 0.5I(X_1 + X_2 > 0) - 0.5I(X_1 + X_2 < = 0)[1 + I(X_2 < = -0.5)] + X_3^2 - X_2^2$, feature variables X_k were generated from a standard normal distribution, and $\eta(X) = X_1 - 0.5X_2$. Here, $\eta(X)$ is the main effect, and the sign of $\phi(X)$ defines the true optimal ITR. The true treatment propensity model in the observational study was specified as $P(A = 1|X) = \text{expit}(1 + 2X_1 + X_2)$, and the treatment assignment probability for the RCT was 0.5.

To accommodate heterogeneity between subjects in the RCT and patients in the EHRs, as well as an unobserved tailoring variable, we considered two scenarios:

- **i.** All features are observed but X_1 has a different distribution in the RCT, where X_1 is restricted to $[-0.4, 0.4]$;
- **ii.** Same as (*i*) but with an additional feature variable X_3 as an unobserved tailoring variable.

Scenarios (*i*) and (*ii*) mimic RCT in which subjects are recruited from subpopulations under certain restrictive inclusion criteria, while the EHR data represent the more general real-world patient population.

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

- **(S1)** Use RCT data and RCT features only;
- **(S2)** Augment the RCT feature set by EHR super features H_1, H_2 (Section 2.2);
- **(S3)** Include H_1 in the feature set and stratify by H_2;
- **(S4)** Include H_2 in the feature set and stratify by H_1.

The super feature H_1 was estimated from a random forest classification model, and H_2 was estimated from a random forest regression model. To speed up computation, the tuning parameter a_n in kernel M-learning was chosen so that the matched pairs with distance less than a_n was fixed as a proportion of pairs (e.g., 25%). In S3, we stratified the training dataset and testing dataset based on a median split of the average predicted benefit H_2, and we included H_1 as an extra feature variable in learning ITR. Similarly, in S4, we stratified the data by the predicted optimal probability H_1 (median split), and we included H_2 as an additional feature. We considered stratifying by both H_1 and H_2 in a simulation in the Supporting Information. That simulation shows that if the sample size is large, then stratification on both super features may lead to a better performance, but with a higher computational cost.

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

Simulation results for M-learning with fitted nonparametric ITRs evaluated on both general (EHR) and restricted (RCT) populations are summarized in Figure 2. The optimal empirical value function is 1.48. When testing on the general population, the addition of the super features directly reduces the variability in scenario (*i*) and also improves the value function to 1.26 from the original value function of 1.10. Domain adaptation by stratification also performs better than without consideration of super features in both (*i*) and (*ii*), which is consistent with our heuristic reasoning given in Section 2.3 (Case I). When testing on the restricted distribution (RCT), the addition of the super features directly achieves the highest mean value in both scenarios (*i*) and (*ii*) (Figure 2a and 2b). The value functions evaluated on the general population are lower than those for the restricted population, demonstrating that the optimal rule fit from the restricted RCT population may not achieve the same effect as that derived from the general EHR population. However, the difference between populations is smaller for the domain adaptation rules, showing that they may be more generalizable.

In another set of analyses, we fit linear rules to scenarios (*i*) and (*ii*), which are misspecified in these settings because the true underlying optimal decision function has a piece-wise linear structure. The results are summarized in Figure 2c and 2d. Both stratification methods lead to a large improvement in scenario (*i*), supporting our heuristic reasoning in Section 2.3 (Case II). Furthermore, domain adaptation rules reach a similar value on the general (EHR) population and the restricted (RCT) population. Thus, we show that even though domain adaptation rules are fit from the RCT population, they behave as if learned from the EHR population with better generalizability. This is consistent with our heuristic reasoning that domain adaptation learning is more generalizable due to the fact that super features are highly correlated with optimal EHR rules (Section 2.3, Case II).The results for Q-learning show similar trends (Supporting Information A2) and simulation results for unmeasured confounder are summarized in Supporting Information A3.

To summarize, domain adaptation learning assisted by EHR super features (i.e., by directly including H or stratifying by H) improves the performance of ITR estimation (i.e., a higher value or a smaller variability). The improvement is observed when evaluated both on the general population and the restricted population. The difference in the value function between the populations is smaller with domain adaptation than without it. The simulations suggest that the information gained from the EHR data may be transferred to ITR estimation on the RCT data.

5 Applications

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

We extracted EHRs from patients with T2D in the New York Presbyterian Hospital (NYPH) clinical data warehouse (CDW). The CDW has implemented a well-defined quality control process, and studies were launched to investigate and improve data quality, including completeness, correctness, concordance, plausibility, and currency (Weiskopf and Weng, 2013). The main information contained in the CDW includes demographics, in-patient and out-patient medication prescriptions, ICD diagnostic codes, and laboratory tests, which are longitudinally documented (Wu et al., 2018).

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

We applied four strategies S1 – S4, as examined by simulations. Our goal was to estimate an optimal ITR to select the best second-line T2D treatment (lispro mix vs. insulin glargine). The outcome measure was a reduction in HbA1c level 12 months post insulin initiation. Baseline features extracted from the EHR data included age, gender, race, baseline value and rate of change of HbA1c, glucose, systolic blood pressure (SBP), diastolic blood pressure (DBP), and body mass index (BMI) estimated from a linear mixed effects model.

When creating super features from the EHR data, we used inverse probability weighting to adjust for missing outcomes. The weights were obtained by a logistic regression model to predict whether a patient had any post insulin treatment HbA1c measure. For kernel M-learning, baseline and rate of change of laboratory test values and demographics variables were used in creating the matching set. Other details are in the Supporting Information A5.

To explore the information available in features from two data sources, in Figure 3, we present the t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize the feature

space. The left panel shows two-dimensional features embedded from t-SNE for patients in the EHRs. The embedded features show a large overlap between the patients who received different treatments in both dimensions. Higher dimensional t-SNE figures suggest similar overlap. Therefore, most patients in the EHR can find matched neighbors based on their feature variables for M-learning, and few extreme subjects are present. The right panel shows two-dimensional features embedded for the RCT data, labeled by the median split of prognostic scores estimated from a linear regression model fitted under the insulin glargine group. The red dots represent subjects in the high prognostic score group (larger than median), while the blue stars represent subjects in the low prognostic score group. This figure appears in color in the electronic version of this article, and any mention of color refers to that version. Although there is some overlap, the two groups clearly separate into two centers. This suggests that the RCT features have power in predicting prognostic scores and that matching on prognostic scores or including them in the learning step of the ITR estimation could improve efficiency.

When pre-training super features on the EHRs, Q-learning was fit by random forest regression, while kernel M-learning was fit using weighted SVM with a Gaussian kernel. The bandwidth for the Gaussian kernel was selected based on maximization of the empirical value function computed from the entire sample. In the next step, we trained random forest models with inverse propensity weighting on the subgroup of patients who received the optimal or non-optimal treatment assignment according to the estimated ITRs. The features included in these models were a common subset of baseline features in the EHR and RCT cohorts, including baseline HbA1c, glucose, SBP, DBP and BMI. We used the trained models to predict two super features on each subject in the RCT: predicted probability of lispro mix being optimal for an individual (H_1); and predicted benefit under the optimal treatment (H_2).

To shed light on the available information from the EHR-derived super features, Figure 4 displays t-SNE embedding of RCT features labeled by dichotomized EHR super features. In the top panel, subjects cluster into two groups based on the dichotomized optimal benefit (H_2) and optimal probability (H_1) features, suggesting likely information gain in estimating ITRs if stratified by these features. In the lower left panel, the embedded RCT features slightly separate the treatment outcome (HbA1c reduction). On the lower right panel, the addition of the EHR-derived probability feature (H_1) further separates subjects in terms of the treatment outcome, which suggests that the optimal treatment probabilities are informative (“high, optimal” represents subjects in the RCT who received the EHR-predicted optimal treatment and had high HbA1c reduction; “low, non-optimal” represents subjects in the RCT who received the EHR-predicted non-optimal treatment and had low HbA1c reduction).

There were 18 baseline covariates in the RCT with EHR super features included to estimate ITRs from 1,945 subjects. All covariates were standardized before fitting the model, and the empirical value function of HbA1c reduction was estimated by 100 repetitions of 3-fold cross-validation. In strategy S2, Q-learning included super features as additional features, and M-learning included them in matching, in addition to a prognostic score. In S3 and S4,

the RCT sample was stratified by one of the two super features, and the other was included in the learning or matching step. Details on the matching are in Supporting Information A5.

The results for the nonparametric rule are displayed in Table 1 and Figure 5. For the non-personalized universal rules, HbA1c reduction is 1.827 for those assigned lispro and 1.672 for those assigned glargine. Q-learning does not provide improvement compared to the universal rule of lispro, and the direct incorporation of super features barely helps. In S3, when stratifying by benefit feature, Q-learning tends to have a worse empirical value. In S4, when stratifying by probability, Q-learning has a higher value function than the baseline model. Kernel M-learning achieves a more significant HbA1c reduction compared to Q-learning when directly including super features with a mean value function of 1.833. In M-learning, both stratification methods (S3 and S4) further improve the mean value function. In particular, stratification by probability provides a large improvement, with a mean value function of 1.849. This suggests that the incorporation of EHR-derived super features transfers useful qualitative information (i.e., optimal treatment probability) to improve performance of ITR estimation in the RCT data. The increased variability of the value function when stratifying by probability is partially due to a higher within-group variability for subgroups defined by probability than by benefit for several important covariates (e.g., baseline HbA1c, baseline glucose). The results for linear rules and stratfication by PC are presented in Supporting Information A6.

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

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

6 Discussion

In this paper, we propose domain adaptation learning to transfer information from observational data to RCTs. In this framework, super features are pre-trained from EHR data to carry information and inform learning ITRs from RCTs. The probability feature is shown to be more robust than the benefit feature on the real-world EHR data because it is more difficult to estimate the benefit of the optimal treatment than to assess the direction of the treatment effects in subgroups. To improve the efficiency of the benefit feature, other approaches for estimating the contrast function can also be used (e.g., Tian et al., 2014). To further improve the performance of domain adaptation, one may consider iteratively deriving informative features from the RCT data and validate on the EHR data, and vice versa.

Our method can provide performance gains when the super features learned from EHR data are informative of treatment responses in RCTs. In real-world practice, valuable but unobserved tailoring variables that are informative of the optimal treatments for real-world patients (e.g., physician insights, observations on patients) may be present and correlate with observed features in the EHR data. Thus, EHR super features fit with observed variables may be informative of these latent factors and also predictive of the optimal treatment. The higher this correlation, the more expected gain in efficiency and generalizability for domain adaptation learning. When the EHR treatments are not beneficial, our approach may not provide a gain in efficiency, but it still remains consistent (i.e., it converges to the true optimal treatment rule) because only RCT data are used for treatment rule learning. In this case, one may consider subgroups in the EHRs for which the ITRs are beneficial.

The proposed pre-training in step 0 of Algorithm 1 uses the same pre-treatment covariates in the EHRs and the RCT. Thus, one consideration of a good EHR/RCT pair for application of our method is the breadth of features captured in the EHRs and RCT. Another consideration is the similar direction of treatment response in subgroups of two populations. While our work focuses on the analysis of existing data, it can be conjectured to use EHRs to assist recruiting patients in future RCTs or conduct point-of-care trials (Angus, 2015).

The proposed domain adaptation framework can be implemented in practice by following the three steps in Algorithm 1. Details regarding computational issues are in the Supporting Information A7. Lastly, external validation on different study samples is needed to validate our method.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by U.S. NIH grants NS073671, GM124104, and MH117458.

References

Angus DC (2015). Fusing randomized trials with big data: the key to self-learning health care systems? JAMA 314, 767–768. [PubMed: 26305643]

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

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

Haynes B (1999). Can it work? does it work? is it worth it?: The testing of healthcare interventions is evolving. BMJ: British Medical Journal 319, 652. [PubMed: 10480802]

Liu Y, Wang Y, Kosorok MR, Zhao Y, and Zeng D (2018). Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. Statistics in Medicine page In press.

Moodie EE, Richardson TS, and Stephens DA (2007). Demystifying optimal dynamic treatment regimes. Biometrics 63, 447–455. [PubMed: 17688497]

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

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

Plank J, Wutte A, Brunner G, Siebenhofer A, Semlitsch B, Sommer R, Hirschberger S, and Pieber TR (2002). A direct comparison of insulin aspart and insulin lispro in patients with type 1 diabetes. Diabetes Care 25, 2053–2057. [PubMed: 12401756]

Qian M and Murphy SA (2011). Performance guarantees for individualized treatment rules. The Annals of Statistics 39, 1180. [PubMed: 21666835]

Raja-Khan NT, Warehime SS, and Gabbay RA (2007). Review of biphasic insulin aspart in the treatment of type 1 and 2 diabetes. Vascular Health and Risk Management 3, 919–935. [PubMed: 18200811]

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

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

Weiskopf NG and Weng C (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. Journal of the American Medical Informatics Association 20, 144–151. [PubMed: 22733976]

Wu P, Zeng D, and Wang Y (2018). Matched learning for optimizing individualized treatment strategies using electronic health records. Journal of the American Statistical Association page In press.

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

Zhao Y, Zeng D, Rush AJ, and Kosorok MR (2012). Estimating individualized treatment rules using outcome weighted learning. Journal of the American Statistical Association 107, 1106–1118. [PubMed: 23630406]

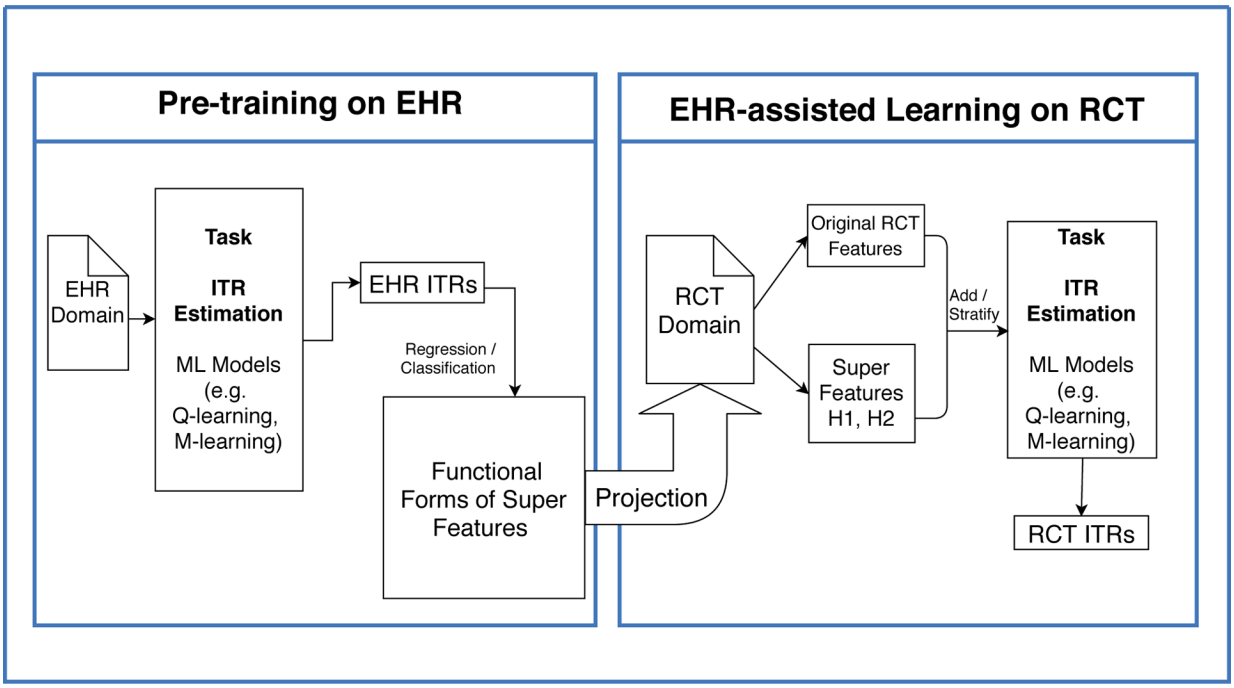
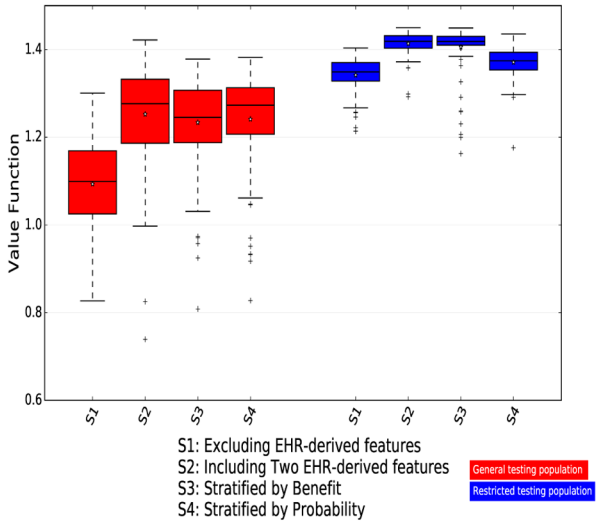
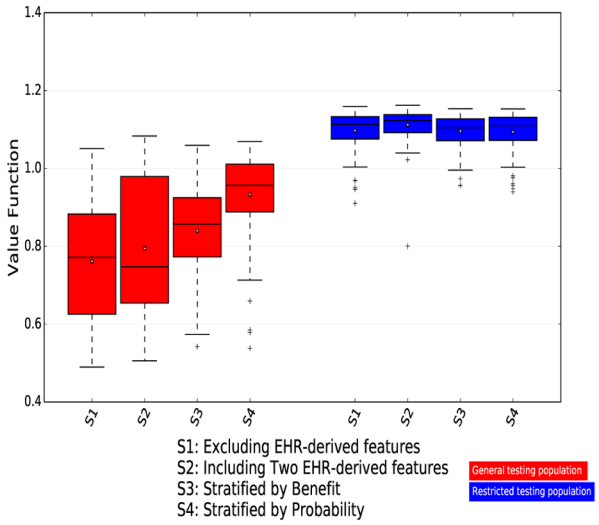


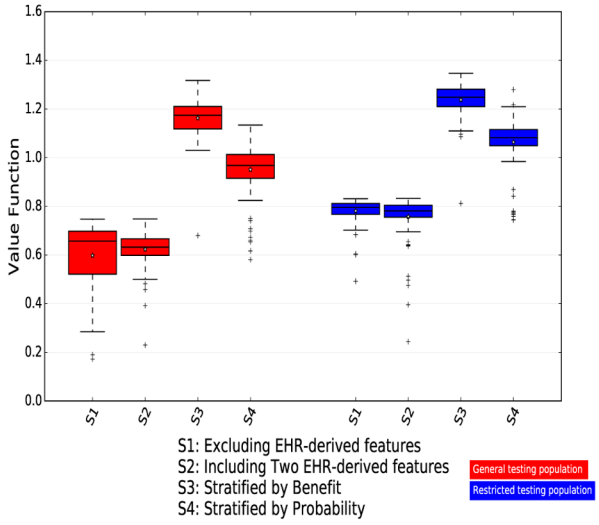
Figure 1.
Schematics of Proposed Domain Adaptation from EHR to RCT



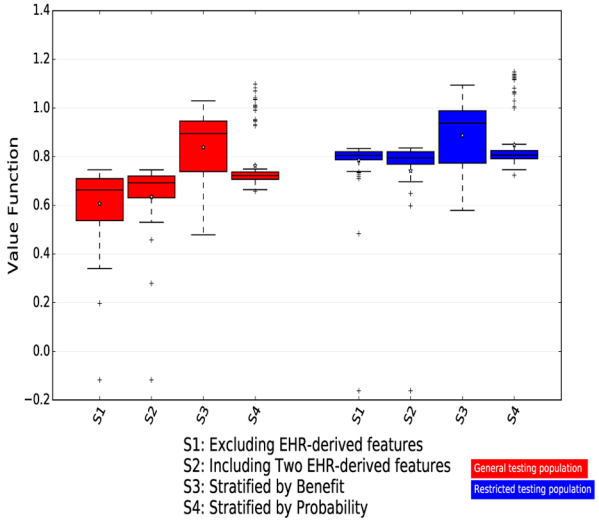
(a) Scenario (*i*), nonparametric rule



(b) Scenario (*ii*), nonparametric rule



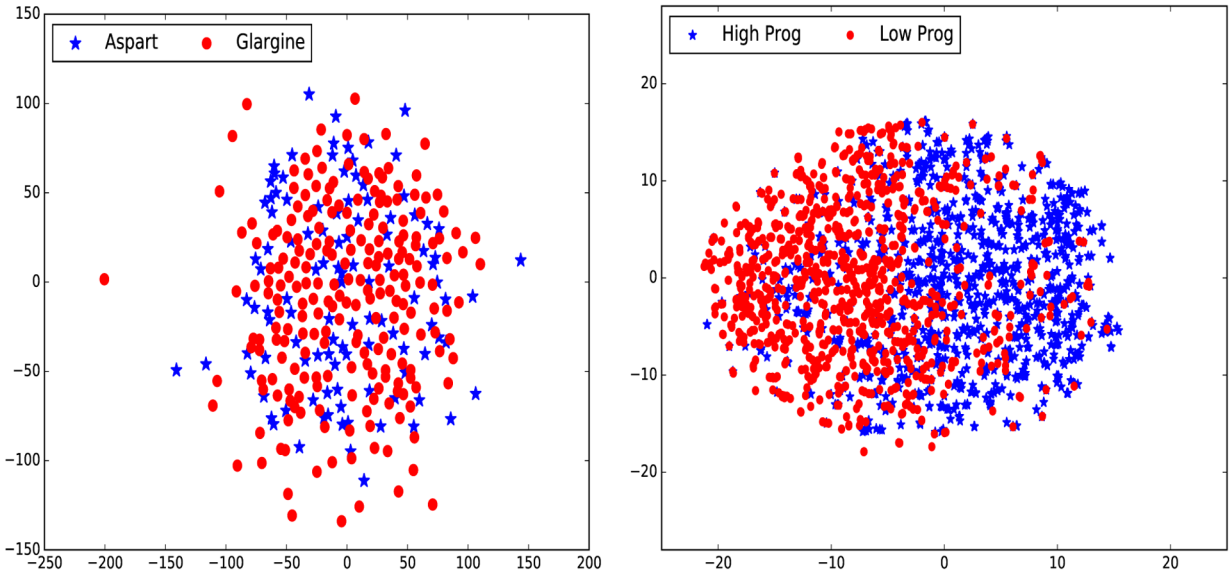
(c) Scenario (*i*), linear rule



(d) Scenario (*ii*), linear rule

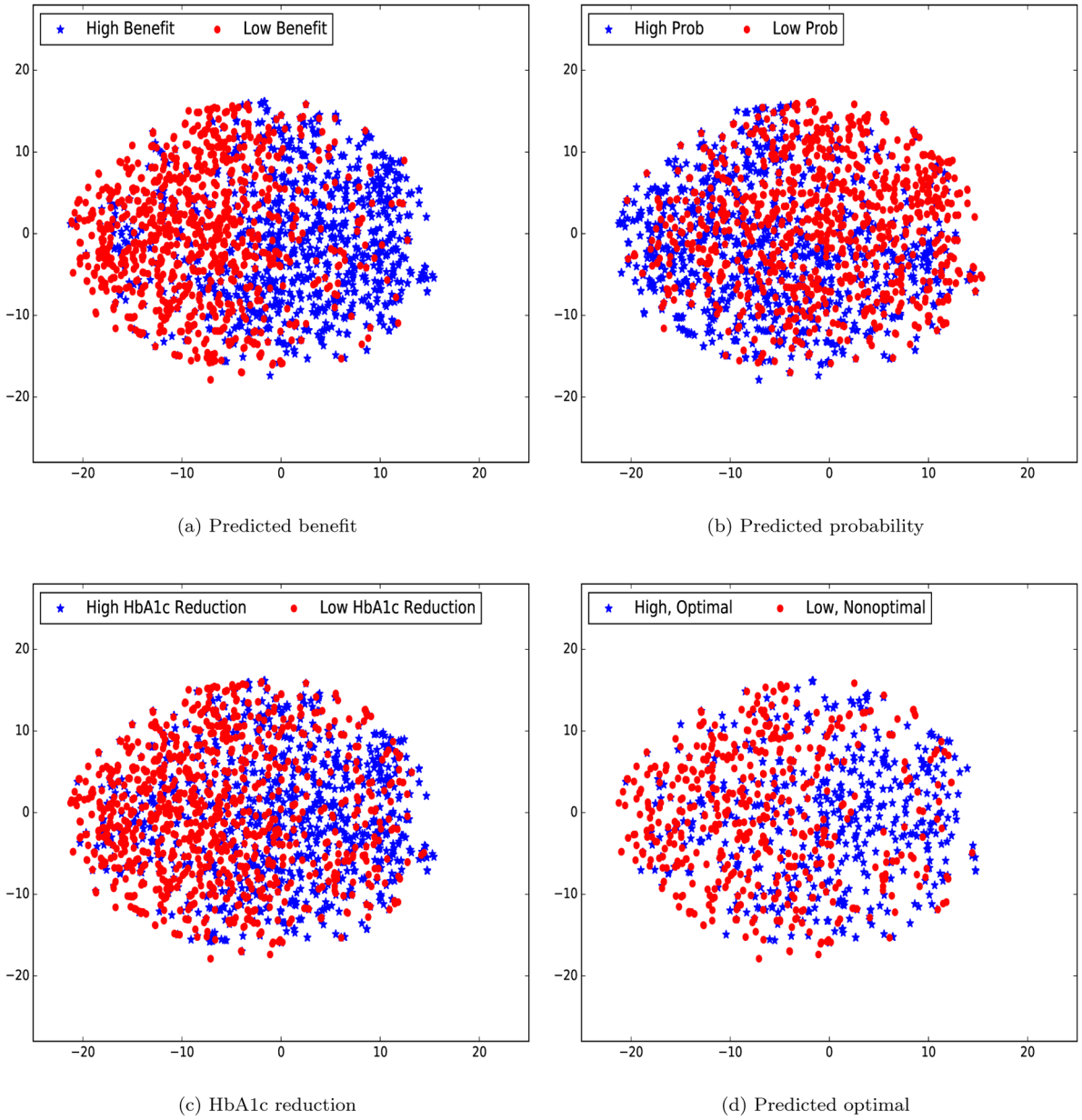
Figure 2.

Simulation comparisons for M-learning, evaluated on independent testing sets generated from general (EHR) or restricted (RCT) populations; scenario (*i*) has no latent tailoring variables, while scenario (*ii*) has a latent tailoring variable not used in learning



(a) EHR dataset (b) RCT dataset

Figure 3.
t-SNE plots for features extracted from CUMC EHRs and DURABLE trial



(a) Predicted benefit

(b) Predicted probability

(c) HbA1c reduction

(d) Predicted optimal

Figure 4.
t-SNE plots for features of a DURABLE trial

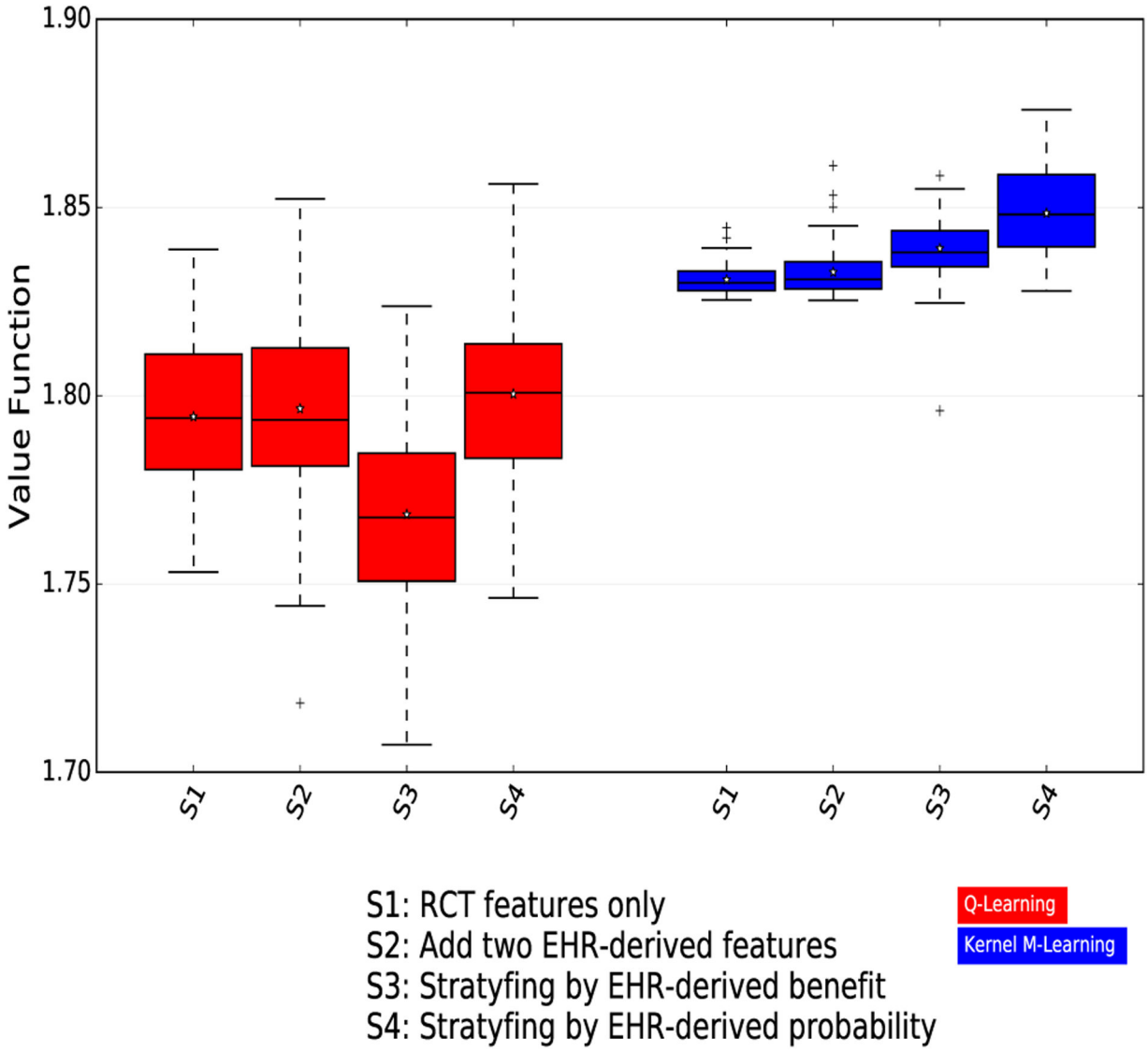


Figure 5.
Empirical value function of HbA1c reduction in the DURABLE trial with 100 repetitions of a three-fold cross-validation, nonparametric rule

Table 1

HbA1c reduction comparing domain adaptation learnings in the DURABLE trial (100 repetitions of a three-fold cross-validation, nonparametric rule)

One-Size-Fits-All: Lispro: 1.827, Glargine: 1.672				
	Q-Learning		**Kernel M-Learning**	
Strategy*	**Mean (Std)**	**Median**	**Mean (Std)**	**Median**
S1	1.794 (0.020)	1.794	1.831 (0.004)	1.830
S2	1.797 (0.024)	1.794	1.833 (0.006)	1.831
S3	1.769 (0.023)	1.768	1.839 (0.008)	1.838
S4	1.800 (0.021)	1.801	1.849 (0.011)	1.848

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