Statistical Learning Methods for Optimizing Personalized Treatments

Yuanjia Wang, Ph.D.

Department of Biostatistics, Mailman School of Public Health
Columbia University
& Division of Biostatistics, New York State Psychiatric Institute

Personalized Medicine

Personalized Medicine: **"the tailoring of medical treatment to the individual characteristics of each patient"** (President's Council of Advisors on Science and Technology).

- Motivations:
  - Heterogeneity in responses:
    1. Across patients: what works for one may not work for another (MDD response rate 47%, Trivedi et al., 2006).
    2. Within a patient: what works now may not work later (MDD relapse rate 50%, APA 2000).
  - Presence of co-morbidity and side effects severity (Colins et al. 2004)
- Current practice: largely based on "trial-and-error"
- Research goal: evidence-based tailored treatment to distinguish in advance which treatment will most likely to benefit a patient

Individualized (Dynamic) Treatment Rules

Individualized (dynamic) treatment rules (ITR/DTR, Lavori & Dawson 1998; Murphy 2005): decision rules prescribing medical treatment/therapy for patients in a given state.

Mathematically, it is a mapping from currently available information (e.g., biomarker, intermediate outcomes, a diagnostic test) into the space of possible decisions (e.g., intensify a treatment or augment).

DTR Example 1: Adaptive Pharmacological Behavioral Treatments for Children with Attention Deficit Hyperactive Disorder (Nahum-Shani 2012).

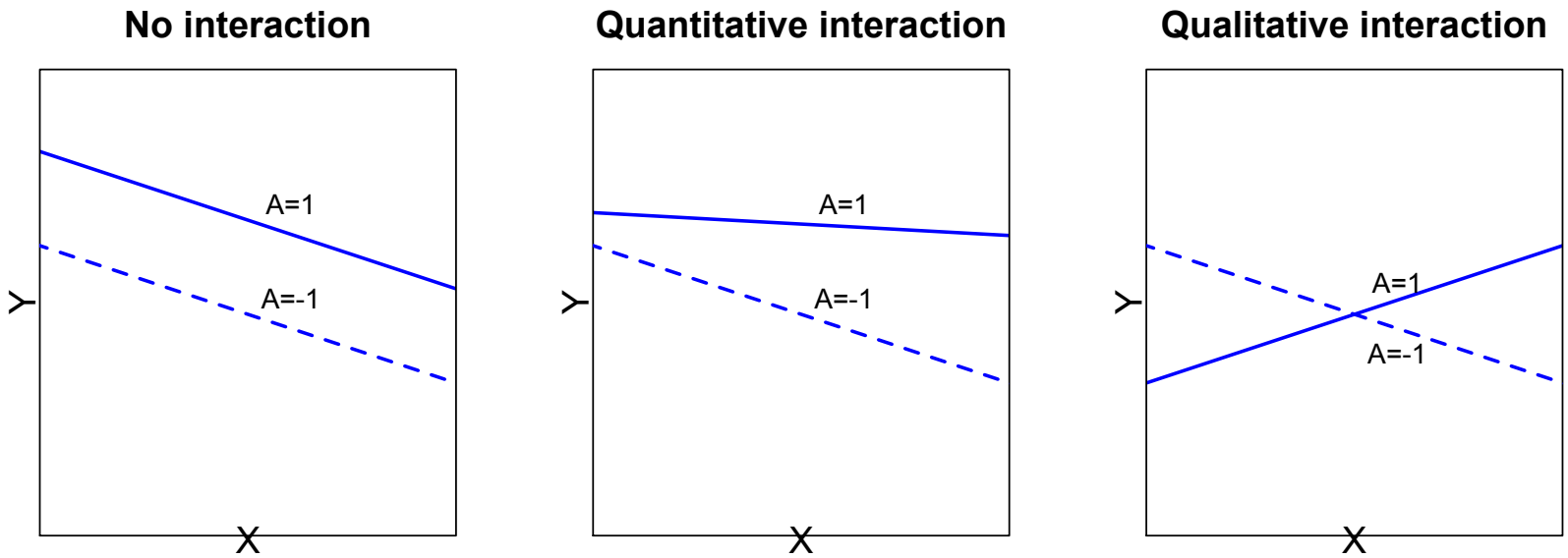
- Prescribe low-dose medication as initial treatment; if a child responds then continue; if a child does not respond then augment with behavioral modification.

Individualized (Dynamic) Treatment Rules

DTR Example 2: Healing Emotion After Loss (HEAL, Shear et al. 2016)

- Administer grief-informed clinical management as the initial treatment; if a patient responds then continue; if a patient does not respond then offer an anti-depressant (Citalopram).

Types of Variables for Tailoring Treatments



Three types of state variables (pre-treatment covariates):

- prognostic variables (associated with clinical outcomes, no interaction with treatment A)
- predictive variables (quantitative interaction)
- prescriptive variables (qualitative interaction)

Data-Driven Approaches fro Personalizing Treatment

Analytical challenges for discovering optimal ITR:

- Tailoring variables unknown
- Large number of candidate tailoring variables
- Structure among variables

Existing methods:

- Double robust regression (Zhang et al. 2012, 2014)
- Virtue twins (Foster et al. 2011); Interaction Trees (Su et al. 2009)
- Q-learning (Murphy 2005; Qian and Murphy 2011; Nahum-Shani et al. 2012); A-learning (Murphy 2003)
- O-learning (Zhao et al. 2012, 2014; Liu et al. 2014)

Q-learning and O-learning

Q-learning: Decompose expected outcome into two components and maximize over possible treatment options:

$$E(Y|X, A; \psi, \beta) = G(X; \psi) + H(X, A; \beta)$$

- Treatment-free model (effect of patient history on outcome without treatment): $G(X; \psi)$
- Blip model (effect of treatment on the outcome): $H(X, A; \beta)$
- Pose models to maximize value function
- Multi-stage problems use backwards induction

O-learning: Directly maximize the value function among class of treatment rules d

$$\max_{d \in \mathcal{D}} E^d(Y)$$

Prior work on ITR focus on maximizing efficacy outcomes

Introduction to Our Work

Part I: Consider Both Efficacy and Safety Outcome

Why considering safety outcomes when Estimating ITR?

Complete picture of treatment decision making involves both efficacy and safety

- Most efficacious treatment for a patient could also lead to a greater safety concern (escalating dosage of insulin may increase risk of hypoglycemia; Zhao et al. 2013)
- Patients with chronic disease and long duration treatment exposed to higher risk of adverse events (severe hypoglycemia, hospitalization; Wild et al., 2007)

Regulator/Industry: Important to characterize both the efficacy and risk profiles among patient populations (FDA guideline on evaluating cardiovascular risk for new antidiabetic therapies)

- In HEAL, dose of Citalopram had to be tapered due to FDA warning of cardiovascular risk

Consider Both Efficacy and Safety Outcomes

How to incorporate safety outcomes for estimating ITR?

- No treatment heterogeneity regarding safety outcomes
- Presence of heterogeneity:
  - Well known example: abundance of drug-metabolizing enzymes (cytochrome P45) varies across subjects, and thus adverse reactions to the same drug dosage
  - Risk of hypoglycemic events depends on patient characteristics and choice of treatment regimen (Sinclair et al., 2015)

Statistical Methodologies

Framework for ITR Under Risk Constraint

Notation:

- Y: efficacy outcome (e.g., symptom reduction; change in HbA1c)
- R: risk outcome (hypoglycemia episodes)
- Two treatment arms $A \in \{-1, 1\}$
- Patient health history $\boldsymbol{X}$
- ITR $\mathcal{D}(\boldsymbol{X})$: mapping from $\boldsymbol{X}$ to $\{-1, 1\}$.

Framework for ITR Under Risk Constraint

Notation:

- Y: efficacy outcome (e.g., symptom reduction; change in HbA1c)
- R: risk outcome (hypoglycemia episodes)
- Two treatment arms $A \in \{-1, 1\}$
- Patient health history $\boldsymbol{X}$
- ITR $\mathcal{D}(\boldsymbol{X})$: mapping from $\boldsymbol{X}$ to $\{-1, 1\}$.

Goal: estimate optimal ITR $\mathcal{D}^*$ while controlling for risk

$$\begin{cases} \max_{\mathcal{D}} & E^{\mathcal{D}}(Y), \\ s.t. & E^{\mathcal{D}}(R) \leq \tau, \end{cases}$$

- $E^{\mathcal{D}}(Y)$: expected efficacy outcome under ITR: $A = \mathcal{D}(\boldsymbol{X})$
- τ: pre-specified tolerance threshold of the risk

Theoretical Optimal ITR Under Risk Constraint

Using data $(Y, R, A, \boldsymbol{X})$ collected from RCT, equivalent to (Qian and Murphy 2011):

$$\begin{cases} \max_{\mathcal{D}} & E\left\{\frac{I(A=\mathcal{D}(\boldsymbol{X}))}{p(A|\boldsymbol{X})}Y\right\} \\ s.t. & E\left\{\frac{I(A=\mathcal{D}(\boldsymbol{X}))}{p(A|\boldsymbol{X})}R\right\} \leq \tau \end{cases}$$

Define $\mathcal{D}(\boldsymbol{X}) = \text{sign}(f(\boldsymbol{X}))$, the above is equivalent to

$$\begin{cases} \max_{f} & E\left\{\delta_Y(\boldsymbol{X})I(f(\boldsymbol{X}) > 0)\right\} \\ s.t. & E[\delta_R(\boldsymbol{X})I(f(\boldsymbol{X}) > 0)] \leq \alpha \end{cases}$$

where

$$\delta_Y(\boldsymbol{X}) = E[Y|\boldsymbol{X}, A = 1] - E[Y|\boldsymbol{X}, A = -1],$$

$$\delta_R(\boldsymbol{X}) = E[R|\boldsymbol{X}, A = 1] - E[R|\boldsymbol{X}, A = -1],$$

and $\alpha = \tau - E[R|A = -1]$.

Theoretical Optimal ITR Under Risk Constraint

Key theoretical result: *The optimal treatment rule under risk constraint is* $\mathcal{D}^*(\boldsymbol{X}) = sign(f^*(\boldsymbol{X}))$, *where*

$$f^*(\boldsymbol{X}) = \begin{cases} sign(\delta_Y(\boldsymbol{X})), & \boldsymbol{X} \in \mathcal{A} \\ sign\left(\delta_Y(\boldsymbol{X}) - \lambda^* \delta_R(\boldsymbol{X})\right), & \boldsymbol{X} \in \mathcal{A}^c \end{cases}$$

and $\mathcal{A} = \{\boldsymbol{X} : \delta_Y(\boldsymbol{X})\delta_R(\boldsymbol{X}) \leq 0\}$. *Here,* $\lambda^* = 0$ *if* $E\left[\delta_R^+(\boldsymbol{X})|\boldsymbol{X} \in \mathcal{A}^c\right] \leq \alpha^*$; *otherwise,* λ^* *solves equation*

$$\begin{aligned} &E\left[\delta_R(\boldsymbol{X})I\{\delta_R(\boldsymbol{X}) > 0, \delta_Y(\boldsymbol{X})/\delta_R(\boldsymbol{X}) > \lambda\}|\boldsymbol{X} \in \mathcal{A}^c\right] \\ +&E\left[\delta_R(\boldsymbol{X})I\{\delta_R(\boldsymbol{X}) < 0, \delta_Y(\boldsymbol{X})/\delta_R(\boldsymbol{X}) < \lambda\}|\boldsymbol{X} \in \mathcal{A}^c\right] = \alpha^* \end{aligned}$$

with $\alpha^* = \frac{\alpha - E[\delta_R(\boldsymbol{X})I(\delta_Y(\boldsymbol{X}) > 0, \boldsymbol{X} \in \mathcal{A})]}{P(\boldsymbol{X} \in \mathcal{A}^c)}$.

Remark 1. Solving for $\mathcal{D}^*$ is analogous to finding the optimal rejection region at a given type I error rate as in the Neyman-Pearson lemma.

Remark 2. When no treatment heterogeneity on safety outcomes, apply with $\delta_R(\boldsymbol{X}) = c$.

Estimating Optimal ITR Under Risk Constraint

Method 1: BR-Q learning

- Predictive modeling-based learning algorithm (reduces to Q-learning in the absence of R)
  - Step 1. Fit regression model for Y given $(A, \boldsymbol{X})$, obtain $\widehat{\delta}_Y(\boldsymbol{X}) = \widehat{E}[Y|\boldsymbol{X}, A = 1] - \widehat{E}[Y|\boldsymbol{X}, A = -1]$
  - Step 2. Fit regression model for R given $(A, \boldsymbol{X})$, obtain $\widehat{\delta}_R(\boldsymbol{X}) = \widehat{E}[R|\boldsymbol{X}, A = 1] - \widehat{E}[R|\boldsymbol{X}, A = -1]$
  - Step 3. Apply the theorem:

$$\widehat{f}(\boldsymbol{X}) = \begin{cases} \text{sign}(\widehat{\delta}_Y(\boldsymbol{X})), & \boldsymbol{X} \in \widehat{\mathcal{A}} \\ \text{sign}\left(\widehat{\delta}_Y(\boldsymbol{X}) - \widehat{\lambda}\widehat{\delta}_R(\boldsymbol{X})\right), & \boldsymbol{X} \in \widehat{\mathcal{A}}^c. \end{cases}$$

Estimating Optimal ITR Under Risk Constraint

Method 2: BR-O learning

- Directly estimate $\mathcal{D}^*$ under risk constraint without posing a regression model (reduces to O-learning in the absence of R):

$$\begin{cases} \max_{\mathcal{D}} & E\left\{\frac{I(A=\mathcal{D}(\boldsymbol{X}))}{p(A|\boldsymbol{X})}Y\right\}, \\ s.t. & E\left\{\frac{I(A=\mathcal{D}(\boldsymbol{X}))}{p(A|\boldsymbol{X})}R\right\} \leq \tau. \end{cases}$$

Maximizes empirical value function under constraint:

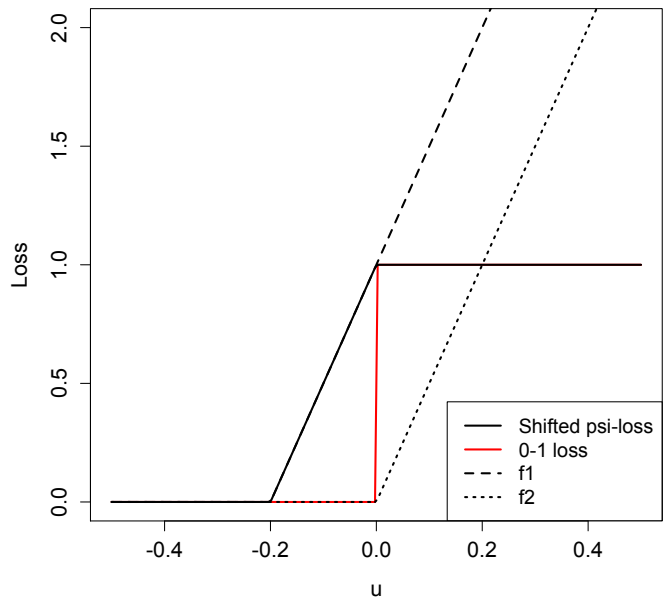
$$\begin{cases} \max_{f} & n^{-1}\sum_{i=1}^{n}\frac{Y_i}{P(A_i|\boldsymbol{X}_i)}I\left(A_i = \text{sign}(f(\boldsymbol{X}_i))\right), \\ s.t. & n^{-1}\sum_{i=1}^{n}\frac{R_i}{P(A_i|\boldsymbol{X}_i)}I\left(A_i = \text{sign}(f(\boldsymbol{X}_i))\right) \leq \tau. \end{cases}$$

Implementation of BR-O Learning

Challenges: constrained optimization with non-convex objective function and non-convex constraint.

Solution: approximate $I\left(A_i \neq \text{sign}(f(\mathbf{X}_i))\right)$ in objective function by a surrogate hinge loss, and approximate $I\left(A_i = \text{sign}(f(\mathbf{X}_i))\right)$ in the constraint by a shifted ramp loss (Huang et al. 2014) as upper bound

$$\begin{aligned} \psi_\delta(u) &= f_\delta^1(u) - f_\delta^0(u) \\ &= \delta^{-1}(u+\delta)_+ - \delta^{-1}(u)_+. \end{aligned}$$



Implementation of BR-O learning

The optimization solved by difference of convex functions algorithm (DCA) (Tao and An 1998) and quadratic programming:

$$\begin{cases} \min_f & C\sum_{i=1}^n \frac{Y^*}{p_i}\xi_i + \frac{1}{2}\beta_{(0)}^T \mathbf{K}\beta_{(0)}, \\ s.t. & \sum_{i=1}^n \frac{R_i}{p_i}\left[\delta^{-1}\{A_i f(X_i)+\delta\}_+ - \delta^{-1}\{A_i f(X_i)\}_+\right] \le n\tau, \\ & \xi_i \ge 1 - A_i^*\left\{\beta_0 + \sum_{j=1}^n \beta_j K(X_i, X_j)\right\}, \xi_i \ge 0 \quad \forall i. \end{cases}$$

Tuning parameters C and δ selected by cross validation.

Numeric Results

Simulation Design

- 20 covariates as $X_1, \ldots, X_{20}$ i.i.d. $U(0,1)$, $n = 300$
- Efficacy responses are normally distributed:

$$Y = 1 - 2X_1 + X_2 - X_3 + h_Y(X, A) + \epsilon_Y$$

$$h_Y = 2 * (1 - X_1 - X_2) * A.$$

- Safety responses are truncated normal (truncated at 1):

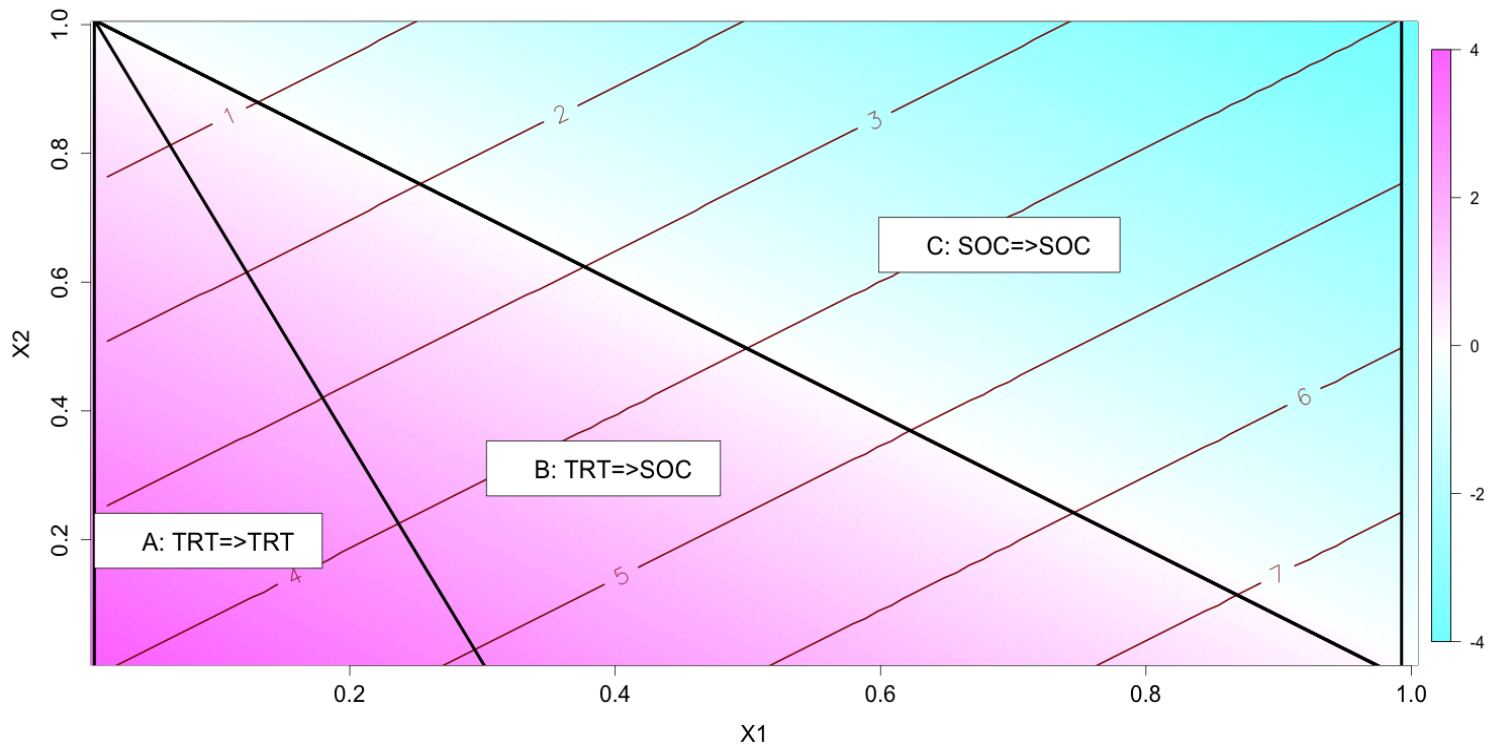
$$R = 2 + X_1 - 2X_2 - X_3 + h_R(X, A) + \epsilon_R$$

$$h_R = (1 + X_1 - X_2) * A.$$

- Prescriptive variables (X_1, X_2):
  - Optimal boundary for Y not considering R:
    $1 - X_1 - X_2 = 0$, positive indicates $A = 1$ is more efficacious

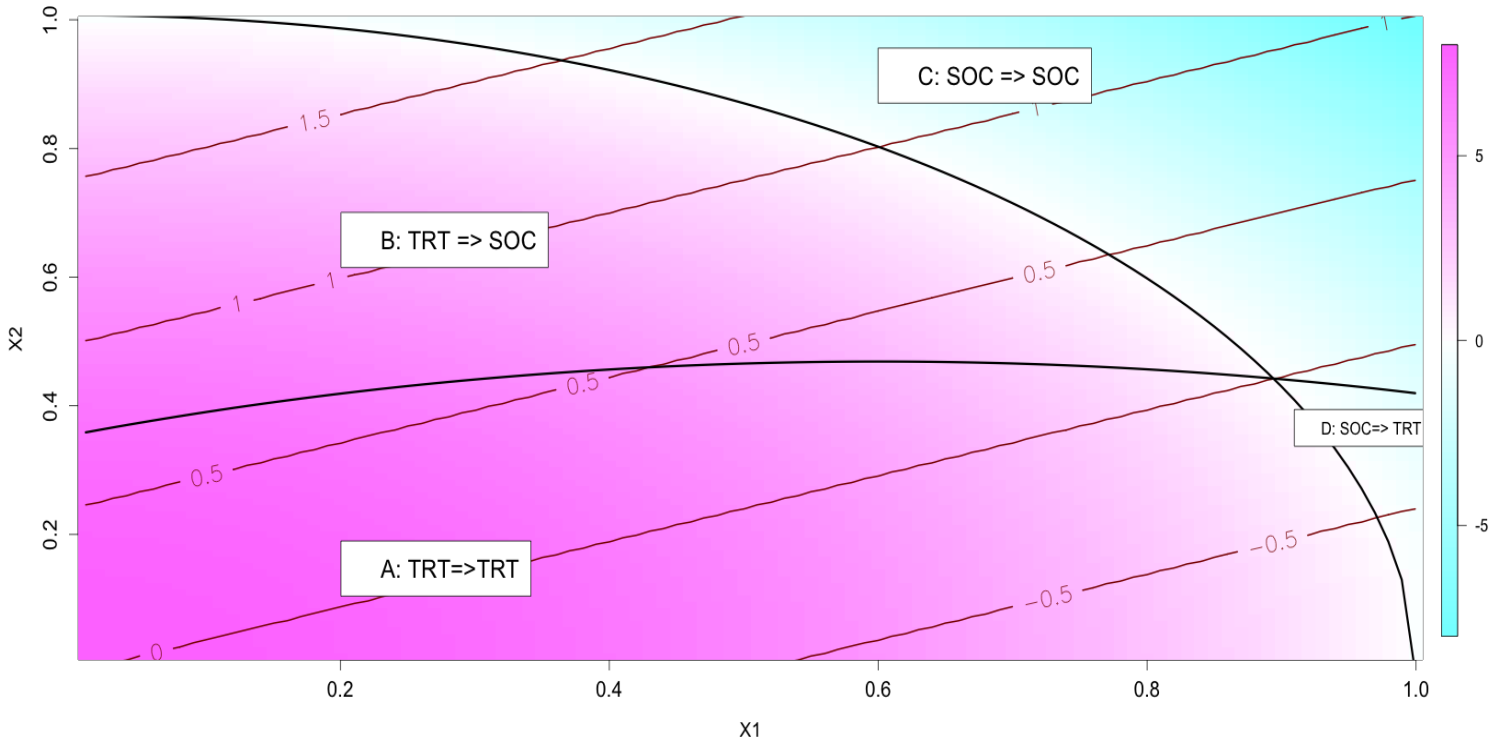
Theoretical Optimal Treatment Decision Boundaries

Figure: Regions of Optimal Treatments with and without Risk Constraint and Relationship with Average Benefit and Risk ($\tau = 1$): linear boundary



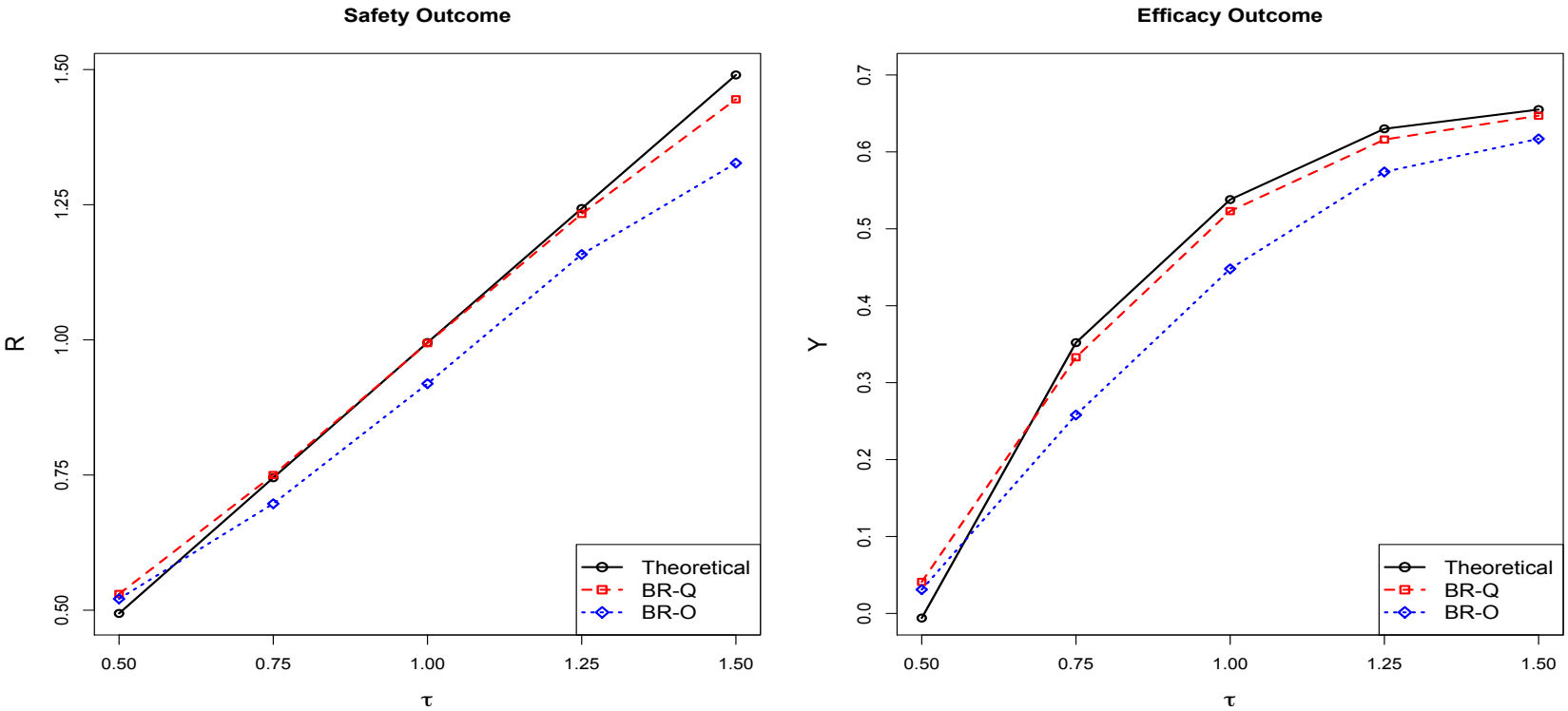
Theoretical Optimal Treatment Decision Boundaries

Figure: Regions of Optimal Treatments with and without Risk Constraint and Relationship with Average Benefit and Risk ($\tau = 1.75$): nonlinear boundary



Simulation Results

Figure: Average efficacy and safety outcome estimated by theoretical formula, BR-Q learning and BR-O learning as a function of pre-specified $\tau^{\dagger}$



†: Optimal average Y without safety constraint = 0.662

Simulation Results

Compare optimal rule computed using theoretical result, BR-Q and BR-O.

Table: Estimated average risk and optimal benefit[†].

	Safety outcome R			Efficacy outcome Y			% Correct	
τ	Theo	BR-Q	BR-O[‡]	Theo	BR-Q	BR-O	BR-Q	BR-O
0.50	0.494	0.530	0.521	-0.006	0.041	0.031	0.969	0.984
0.75	0.745	0.750	0.697	0.352	0.333	0.258	0.926	0.896
1.00	0.995	0.994	0.919	0.538	0.523	0.448	0.926	0.878
1.25	1.243	1.233	1.158	0.630	0.616	0.574	0.926	0.887
1.50	1.490	1.445	1.327	0.655	0.647	0.617	0.923	0.880

[†]: Average safety outcome is 1.503, and optimal value function without safety constraint is 0.662.
[‡]: "Theo": computed from theoretical formula, "BR-Q": Risk constrained Q-learning, and "BR-O": Risk constrained O-learning.

Application to DURABLE Trial

DURAbility of Basal Versus Lispro Mix 75/25 Insulin Efficacy (DURABLE) Trial (Buse et al., 2009):

- Randomized trial to compare the ability of two starter insulin regimens (once-daily basal insulin Glargin or twice-daily premixed insulin Lispro 75/25) to achieve glycemic control in patients with type 2 diabetes
- Insulin-naive patients with type 2 diabetes who did not achieve adequate control with oral antihyper-glycemic drugs
- Efficacy outcome: glycemic control (change in HbA1C from baseline to end point)
- Safety outcomes: hypoglycemia (a plasma glucose value$\leq$70 mg/dl or presence of typically associated symptoms)

Application to DURABLE Trial

Overall analyses results (Buse et al., 2009):

- Efficacy: Lispro 75/25 better control on glycemic than GL ($p = 0.005$)
- Safety: Lispro 75/25 had higher hypoglycemia rate compared to GL ($p = 0.007$)

Application Data Description:

- Sample size: 965 Lispro Mix and 980 insulin Glargin.
- Efficacy endpoint: A1c change from baseline after 24 weeks treatment.
- Safety endpoint: Hypoglycemic event rate per day.
- Candidate tailoring variables: 18 baseline covariates (weight, BMI, blood pressure, heart rate, 7 points blood glucose values, fasting blood glucose, fasting insulin etc.).

Application to DURABLE Trial

Table. Average benefit, risk over 100 repetitions (300 patients as training set, the rest 1,645 patients as a testing dataset).

Risk(τ)	Method	Risk-Training	Risk-Testing	Benefit-Training	Benefit-Testing
0.063	BR-Q	0.0639(0.005)	0.0689(0.004)	1.8668(0.142)	1.7201(0.049)
	BR-O	0.0626(0.003)	0.0640(0.006)	1.7824(0.142)	1.6980(0.042)
0.064	BR-Q	0.0644(0.006)	0.0690(0.004)	1.8682(0.141)	1.7209(0.050)
	BR-O	0.0632(0.003)	0.0650(0.006)	1.7905(0.135)	1.7004(0.050)
0.065	BR-Q	0.0652(0.006)	0.0692(0.004)	1.8736(0.142)	1.7228(0.050)
	BR-O	0.0638(0.003)	0.0650(0.006)	1.7983(0.135)	1.7030(0.051)
0.066	BR-Q	0.0657(0.006)	0.0694(0.004)	1.8780(0.146)	1.7241(0.051)
	BR-O	0.0644(0.003)	0.0655(0.006)	1.8048(0.135)	1.7021(0.046)
0.067	BR-Q	0.0667(0.006)	0.0696(0.004)	1.8827(0.148)	1.7250(0.052)
	BR-O	0.0654(0.003)	0.0660(0.006)	1.8273(0.131)	1.7093(0.048)
∞	BR-Q	0.0756(0.010)	0.0712(0.003)	1.9392(0.153)	1.7378(0.048)
	BR-O	0.0769(0.010)	0.0714(0.005)	1.9895(0.146)	1.7360(0.052)

Conclusion: BR-O controls risk below τ with similar benefit as BR-Q

Ranks of Important Biomarkers

Table: Ranking of Baseline Biomarkers Based on Average Standardized Effects over 100 Repetitions.

$\tau =$	0.063	0.064	0.065	0.066	0.067	∞
Baseline A1C	1	1	1	1	1	1
BMI	2	2	2	2	2	2
Fasting Blood Glucose	3	3	3	3	3	3
Height	4	4	4	4	4	4
Adiponectin	5	5	5	5	5	5
Duration of diabetes	6	6	6	6	6	6
Body Weight	7	7	7	7	7	7
Diastolic blood pressure	8	8	8	8	8	8
Fasting Insulin	9	9	9	9	9	10
Heart rate	10	10	10	10	10	9
Systolic blood pressure	11	11	11	11	11	11
Glucose:Morning before meal	12	12	12	12	12	12
Glucose: 3am at night	13	13	13	13	13	14
Glucose:Evening before meal	14	14	14	14	14	13
Glucose:Morning 2 hours after meal	15	15	16	15	15	16
Glucose:Evening after meal	16	16	15	16	16	15
Glucose:Noon before meal	17	17	18	17	17	18
Glucose:Noon 2 hours after meal	18	18	17	18	18	17

Estimated Optimal Treatment Decision Boundaries Stratified by Baseline A1c

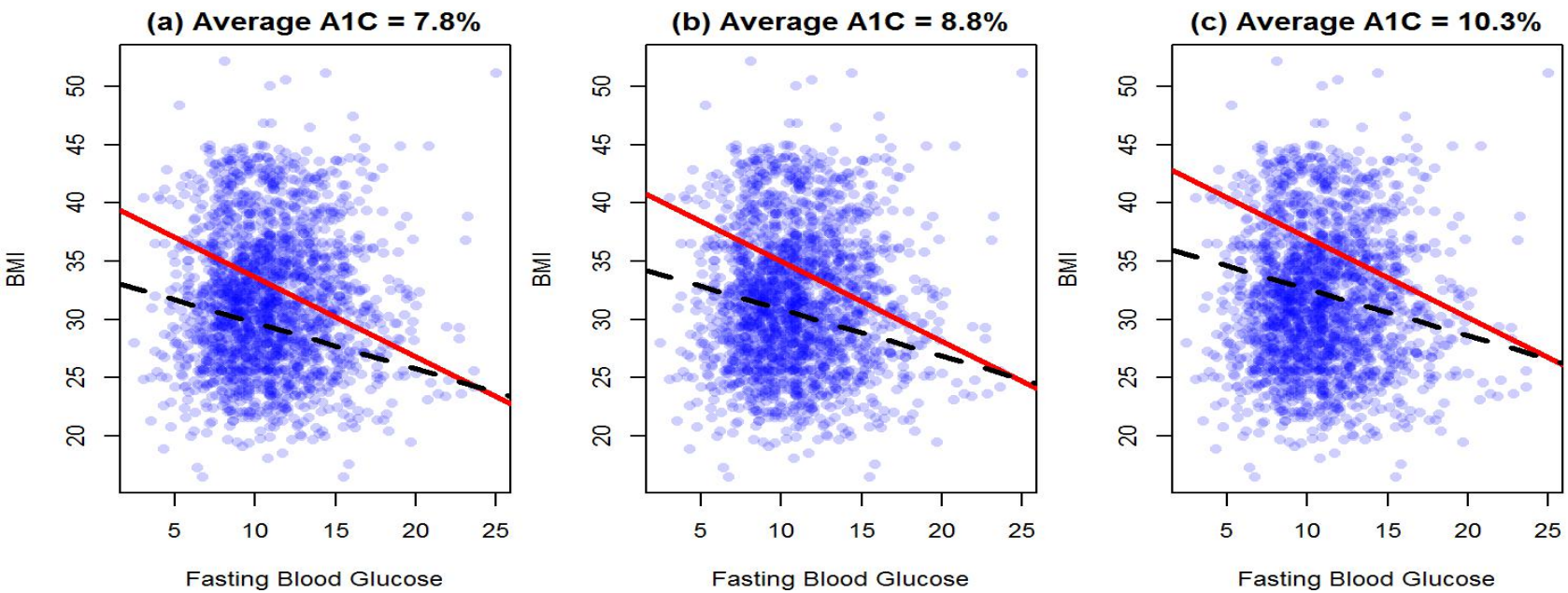


Figure: **Black** dahsed line: O-learning without risk constraint. Red solid line: BR-O ($\tau = 0.065$). Patients above the lines recommended to take mix 75/25 and patients below recommended to take GL.

Part I: Discussion

Two methods (BR-Q and BR-O) for estimating optimal ITR while controlling for average risk.

- Both control theoretical risk adequately and approach theoretical optimal efficacy level.
- In the application, BR-O slightly conservative on the training, but controls risk better on testing.
- BR-O more computationally intensive

Extensions:

- Multiple efficacy and safety outcomes
- Multiple group-dependent thresholds
- Multi-stage trials (SMART, Lavori & Dawson 2000, 2004; Murphy 2005)
- Real world observational studies

Part II: Shifting Towards Observational Studies

Real World Setting: Electronic Health Records (EHRs)

- Clinical guidelines derived from RCTs may be of limited use in real world clinical practice due to stringent inclusion/exclusion criteria, lack of long term outcomes or adverse events etc.
- Electronic health records (EHR) adoption has increased more than nine-fold from 2008 to 2012 (Charles et al. 2013), and the trend continues
- EHR data resources contain massive information: CUMC clinical data warehouse (CDW) contains 20 years of health information for about 4.5 million patients with diverse ethnicity
- However, EHR data are not collected for research oriented studies in the first place, hence might face potential challenges (e.g, confounding; selection bias).

Methods for Observational Studies

Existing O-learning methods mostly focus on inverse probability weighting (IPW) of propensity scores to handle confounding bias.

Matching methods provide a flexible alternative.

- Important tool to achieve covariates balance.
- Flexible matching methods: nearest neighborhood matching, subclassification, full matching etc.
- For example, nearest neighbor matching with replacement is useful when treatment assignment is imbalanced and matching with replacement can reduce bias and avoid the order issue in matching the treated units.

Comparison between IPW and Matching

- Matching approaches require less model specification and can be nonparametric.
- IPW-based methods ensure different treatment groups have similar distribution of confounders at the population level.
- Some matching methods provide flexible tools to control confounding in subgroups or even on individual subject.
- Feature selection, distance metric and measure of covariates balance can be combined to optimize matching.

Matched Learning (M-learning)

Motivation: Two subjects are matched in confounders or propensity scores but are observed to receive opposite treatments, and the observed treatment leading to a larger clinical outcome should be more likely to be the optimal treatment for this subject.

Identify a matched set $\mathcal{M}_i$, let

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

where $d(\cdot, \cdot)$ is a metric defined in the feature space and δ_i is a pre-specified threshold to determine the size of $\mathcal{M}_i$.

Matched Learning (M-learning)

For example, if we choose $\mathcal{M}_i$ to be the nearest neighbor, then δ_i is the minimal distance between subject i and any subjects with the opposite treatment.

Objective function (matching-based preference value function) for maximization:

$$V_n(\mathcal{D}; g) = \sum_{i=1}^{n} |\mathcal{M}_i|^{-1} \sum_{j \in \mathcal{M}_i} \left\{ I(R_j \geq R_i, \mathcal{D}(H_i) = -A_i) + I(R_j \leq R_i, \mathcal{D}(H_i) = A_i) \right\} g(|R_j - R_i|),$$

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

Matched Learning (M-learning)

The value function is equivalent to

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

where $\mathcal{D}(H) = \text{sign}(f(H))$.

Note: Let $R_j = \widehat{\beta}^T X_j$ and $g(x) = 1$, M-learning reduces to residualized O-learning (Liu et al. 2014)!

In order to find the optimal ITR, we aim to maximize $V_n(f;g)$ using quadratic programming.

Matched Learning Properties

- Fisher consistency can be established under conditions
- Double robust matching can be used to improve efficiency
- Residualized M-learning further improves efficiency

Simulation Designs

We considered four simulation designs:

1. RCT where treatment assignment does not depend on H;
2. Observational studies where treatment assignment depends on H and propensity score model is correctly specified;
3. Observational studies where treatment assignment depends on H but propensity score model is misspecified;
4. Observational studies in the presence of unmeasured confounders where treatment assignment depends on H and some components of H are not observed.

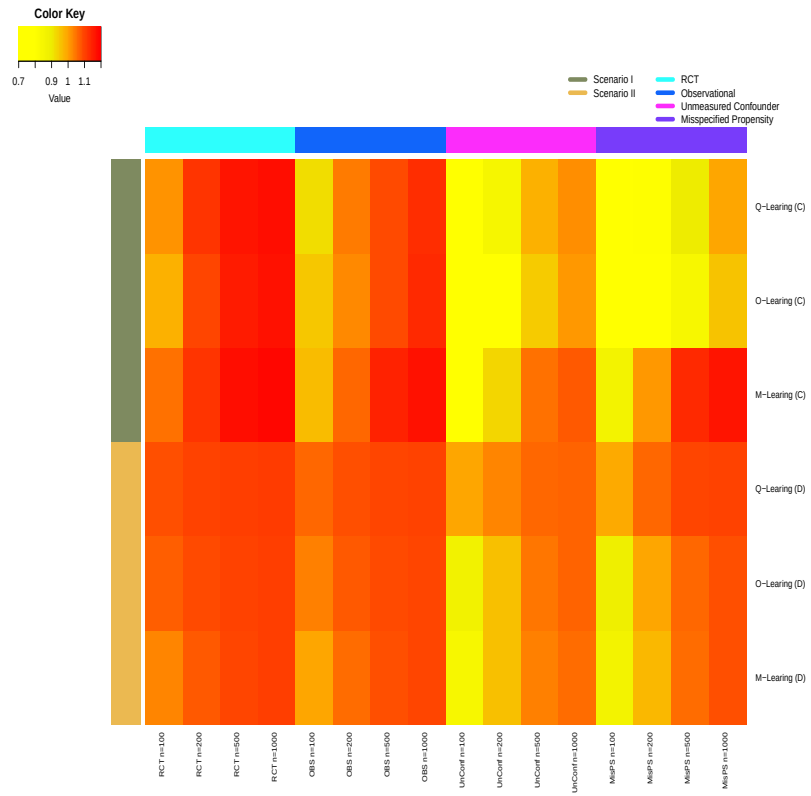
In the following simulations, we considered $1:1$ nearest neighbor matching (NNM) with replacement using Euclidean distance. The procedures were repeated 100 times using 3-fold CV tuning and tested on a large independent testing set.

Numeric Results

Simulation Study Results

- M-learning has a better performance than O-learning for continuous outcome especially when propensity score model is misspecified and in the presence of unmeasured confounders.
- For discrete outcome, there is no large difference between performance of O-learning and M-learning in randomized experiment and observational study.

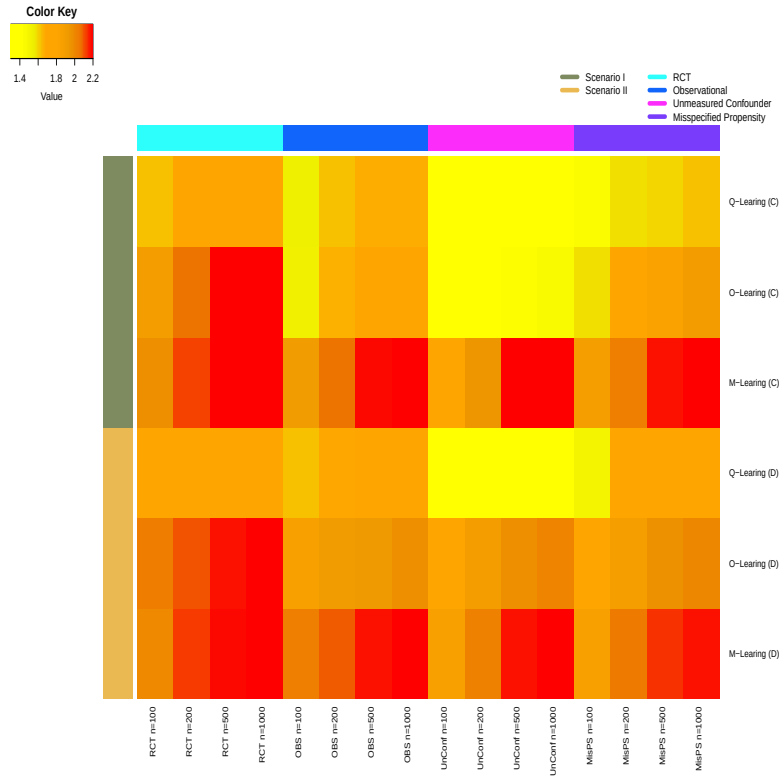
Heatmap of Value Function in Linear Setting



Simulation Study Results

- M-learning ranks the best comparing to other two methods in both scenarios.
- Matching by both propensity score and prognostic score significantly enhances the performance of M-learning.

Heatmap of Value Function in Noninear Setting



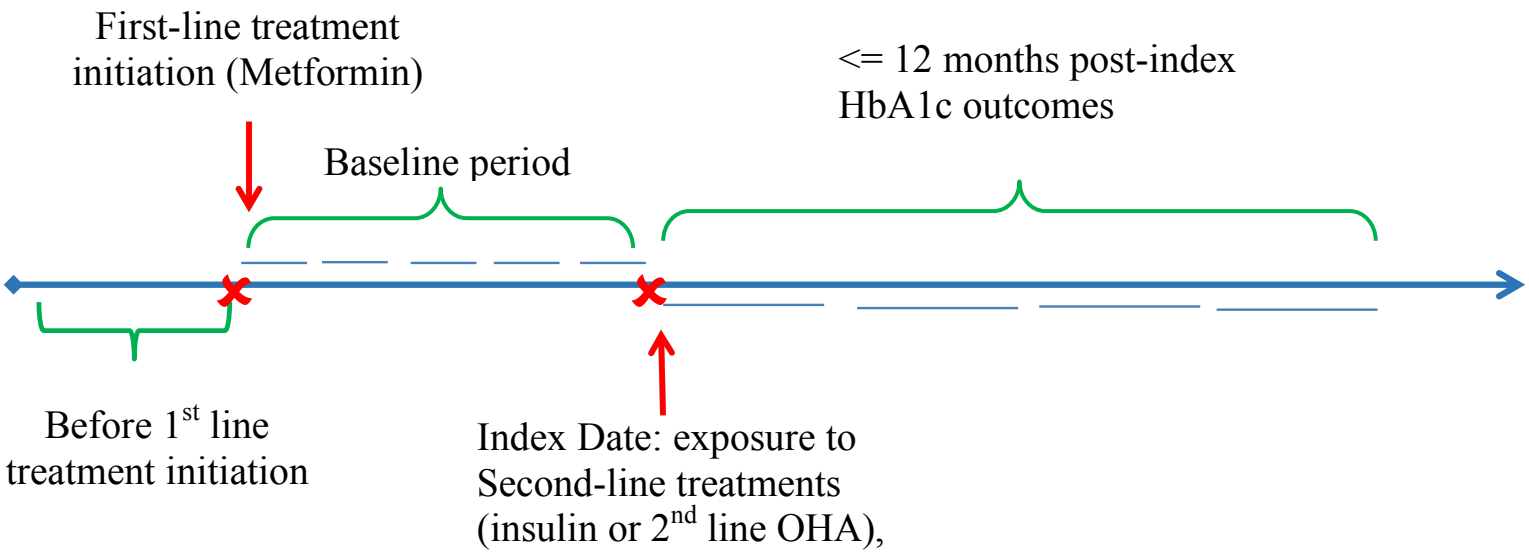
Selecting Second Line Treatment for Type II Diabetes (T2D) using EHRs

Research Aim: Estimate optimal ITR for treating T2D

- Recommended first-line treatment: Metformin (Met).
- Compare ITR for second line therapy: Met + insulin vs Met + Sulf (Sulfonylureas as a class of second line oral agents includes Glipizide and Glyburide).
- Outcome: The primary outcome of interest is average HbA1c level (%) within 1 year post second-line treatment initiation and a lower value corresponds to a favorable outcome.
- Samples: CUMC Clinical Data Warehourse (CDW) adults age 18 or older and having at least one T2D diagnosis ICD-9 codes between 1/1/2008 and 12/31/2012.

Cohort Identification from EHR

Figure: New User Design of Ascertaining T2D Patients' EHR



Matches time-varying confounding and captures early treatment responses.

Feature Extraction: Measurement Patterns Informative

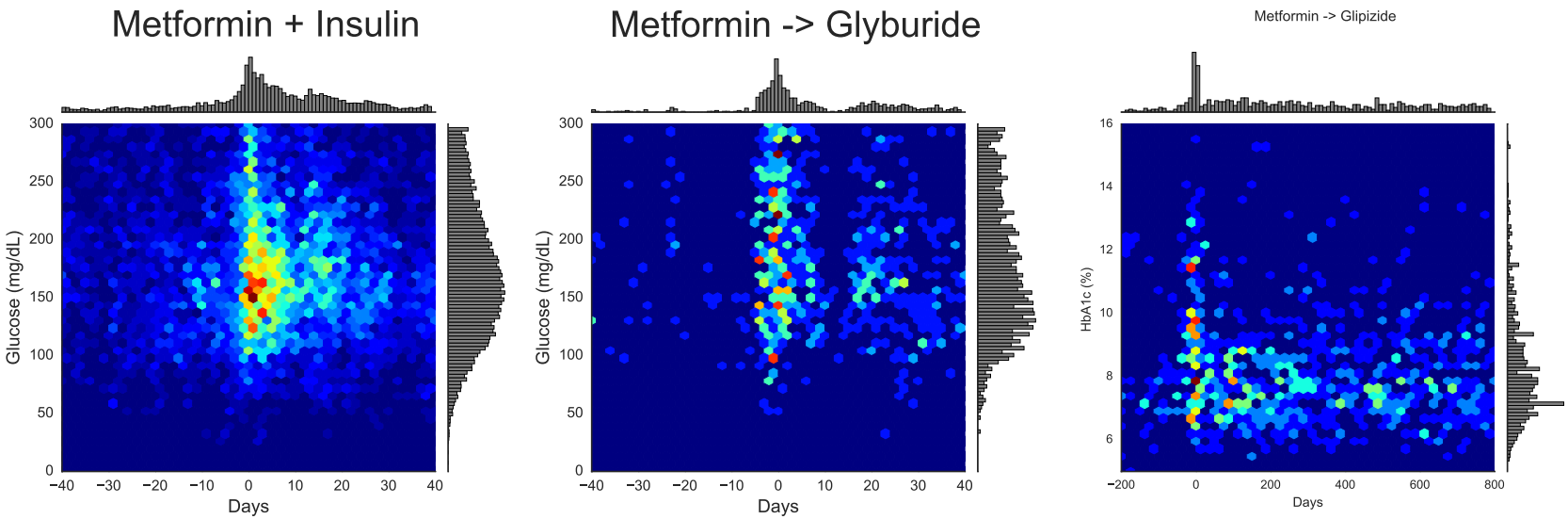
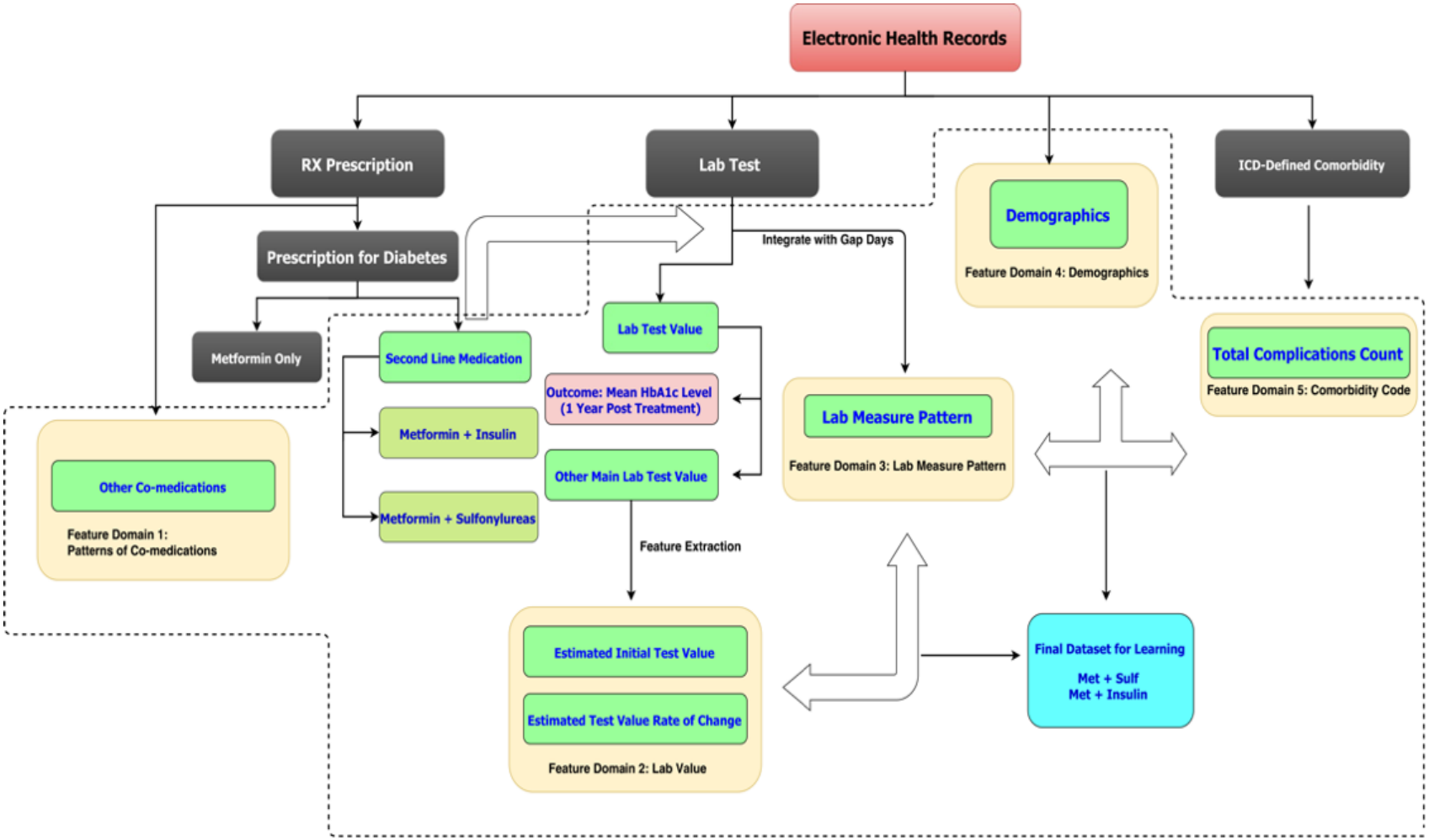


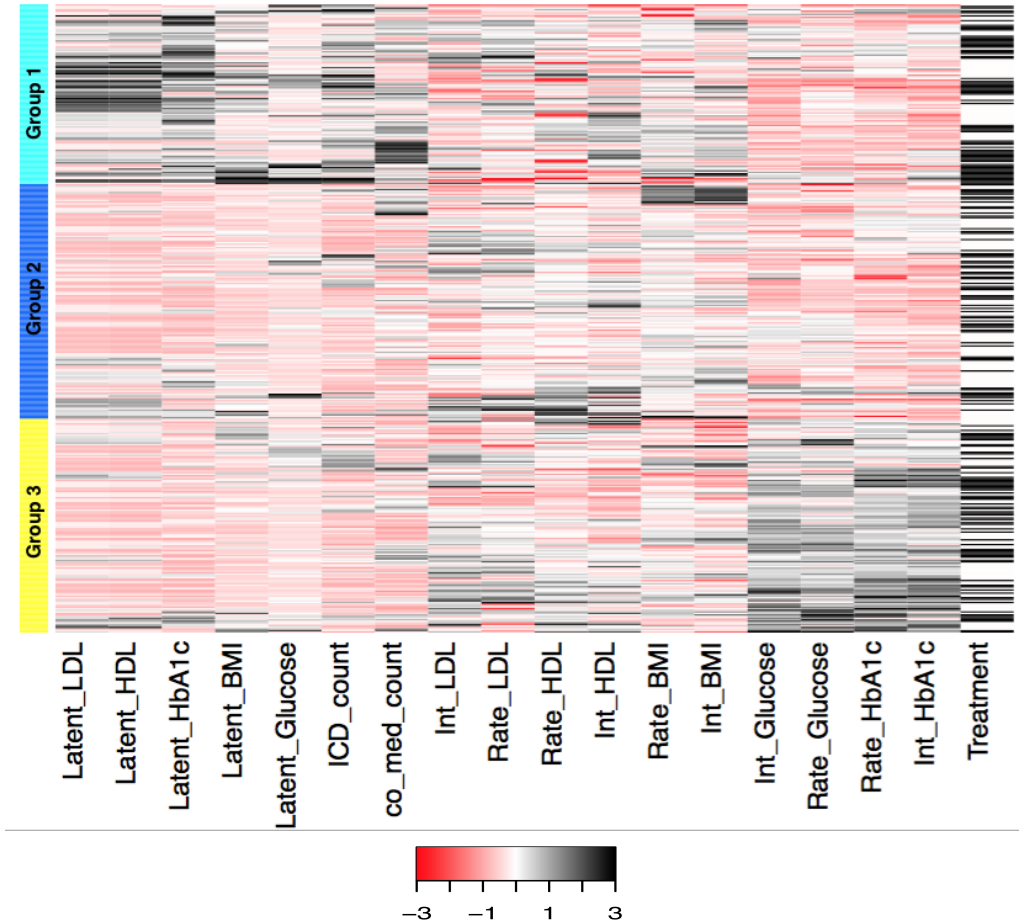
Figure: Glucose Test Values and Measurement Patterns

Main Information Captured in CDW EHRs

Figure: Schematics of EHR Data Processing Procedures



Heatmap of Patient Records and Extracted Features



Group 1: moderately ill; regular, frequent documentation pattern; more co-morbidity and non-diabetic medications; more patients received a OHA; (2) a moderately ill, less-frequent measurements, less co-morbidity and medications; more patients received OHA; (3) a fast progression; less measurements, a higher LDL, HDL; more patients received insulin.

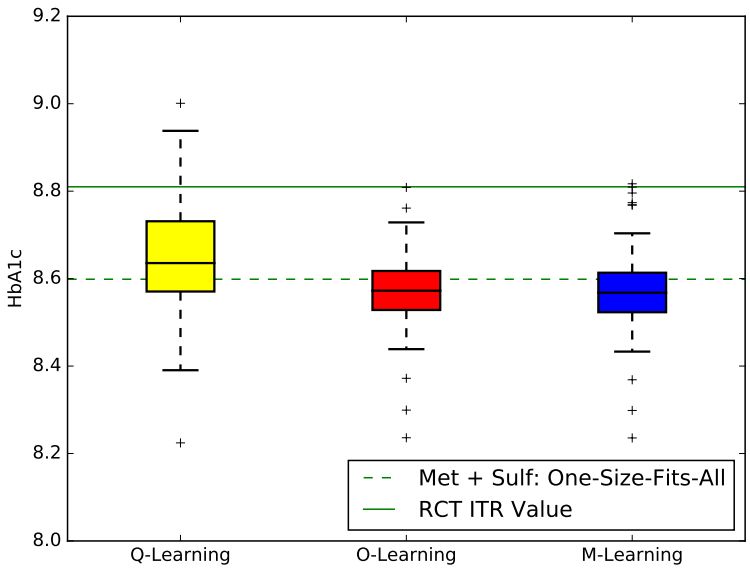
Addressing major statistical challenges

- Confounding bias: matching based on propensity scores constructed from different feature domains, a set of features and prognostic score; use Mahalanobis distance as matching similarity measure.
- Selection bias: use IPW and constructed two models to compute weights
  1. logistic regression model (whether a subject has any post index measure)
  2. proportional hazards model (time to first lab measurement post index)
- Incompleteness in features: Multiple Imputation by Chained Equations (MICE) under assumption of missing at random (MAR).

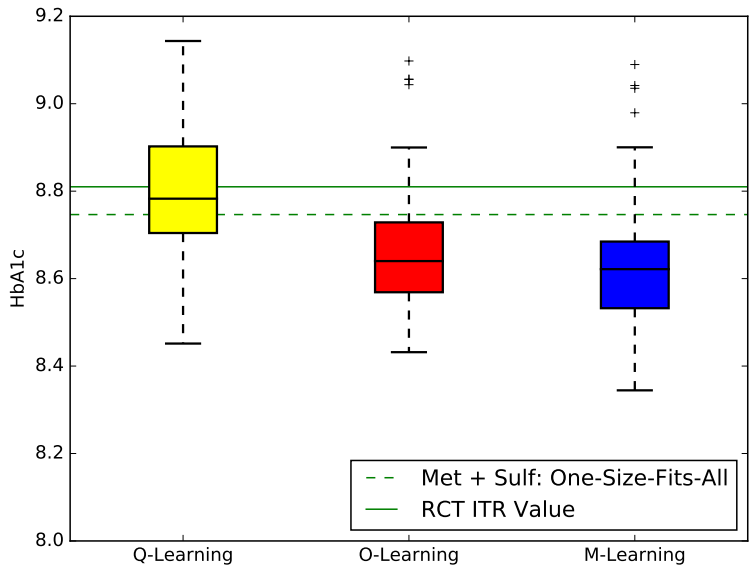
Selecting Second Line Treatment for T2D

Figure: Empirical Value Function of HbA1c in EHR Data (100 Times 2-fold Cross-Validations)

Whole Sample



Subgroup



Summary of Analyses Results

- Most influential features: pre-index rate of change of glucose, LDL and BMI, initial value of HDL before treatment, and race.
- 645 of the 787 patients are recommended to "Metformin + Sulfonylureas".
- Subgroup analysis demonstrates that under the circumstances where the distribution of propensity scores has poor overlap, M-learning will outperform O-learning

Part II Discussion

Demonstrates M-learning is useful with observational studies

- Extension to both efficacy outcomes and adverse events
- Derive multi-level random effects models to handle multiple sources of selection bias and perform sensitivity analysis.
- Consider latent confounders and model to alleviate confounding bias.
- Validate results with RCT or population-based observational data.

Acknowledgments

- Collaborators and Students:
  - Donglin Zeng (UNC Chapel Hill)
  - Haoda Fu (Eli Lilly)
  - Peng Wu (Biostatistics, Columbia University)
  - Ying Liu (Wisconsin Medical College)
  - Christine Mauro (Columbia University)
  - Xin Qiu (Columbia University)
  - Katherine Shear (Columbia University)
- Research Support
  - NIH: NS073671, NS082062
  - NYSPI/RFMH

THANK YOU!