Fairness-Oriented Learning for Optimal Individualized Treatment Rules

Ethan X. Fang* Zhaoran Wang† Lan Wang‡

Abstract

There has recently been a surge on the methodological development for optimal individualized treatment rule (ITR) estimation. The standard methods in the literature are designed to maximize the potential average performance (assuming larger outcomes are desirable). A notable drawback of the standard approach, due to heterogeneity in treatment response, is that the estimated optimal ITR may be suboptimal or even detrimental to certain disadvantaged subpopulations. Motivated by the importance of incorporating an appropriate fairness constraint in optimal decision making (e.g., assign treatment with protection to those with shorter survival time, or assign a job training program with protection to those with lower wages), we propose a new framework that aims to estimate an optimal ITR to maximize the average value with the guarantee that its tail performance exceeds a prespecified threshold. The optimal fairness-oriented ITR corresponds to a solution of a nonconvex optimization problem. To handle the computational challenge, we develop a new efficient first-order algorithm. We establish theoretical guarantees for the proposed estimator. Furthermore, we extend the proposed method to dynamic optimal ITRs. The advantages of the proposed approach over existing methods are demonstrated via extensive numerical studies and real data analysis.

*Department of Statistics, Pennsylvania State University, University Park, PA 16802; email: xxf13@psu.edu.

†Department of Industrial al Engineering and Management Science, Northwestern University, Evanston, IL 60208; email: zhaoranwang@gmail.com.

‡Department of Management Science, Miami Herbert Business School, University of Miami, Coral Gables, FL 33146; email: lanwang@mbs.miami.edu.

Keywords: Quantile, Individualized Treatment Regime, Nonconvex Problem

1 Introduction

One of the primary goals of precision medicine is to estimate the optimal individualized treatment rule (ITR), tailoring the treatment recommendation to patients according to their individual characteristics, such as age, gender, and clinical history. The last decade has witnessed a prodigious surge in research on optimal ITR estimation. Popular existing approaches for estimating optimal ITRs include model-based methods such as Q-learning (Watkins and Dayan, 1992; Murphy, 2003; Moodie et al., 2007; Chakraborty et al., 2010; Goldberg and Kosorok, 2012; Song et al., 2015), A-learning (Robins et al., 2000; Murphy, 2005), model-free policy search methods (Robins et al., 2008; Orellana et al., 2010a,b; Zhang et al., 2012; Zhao et al., 2012, 2014), among others (Robins, 2004; Moodie et al., 2009; Cai et al., 2010; Henderson et al., 2010; Thall et al., 2002; Imai et al., 2013; Huang et al., 2015; Tao and Wang, 2017). See Chakraborty et al. (2010); Chakraborty and Moodie (2013); Laber et al. (2014); Kosorok and Moodie (2015) and references therein for a thorough review. The problem of optimal ITR estimation has also received considerable attention in other fields, such as recommending the most effective training program for the unemployed (Frölich, 2008; Behncke et al., 2009; Staghoj et al., 2010; Wunsch, 2013) and finding the best approach to encouraging voter turnout (Gerber and Green, 2000; Imai et al., 2013). Several econometrics researchers studied optimal ITR estimation in a decision theory framework (Hirano and Porter, 2009; Bhattacharya, 2009; Bhattacharya and Dupas, 2012; Tetenov, 2012).

Typically, an optimal ITR is estimated by maximizing the potential average performance (assuming a larger outcome is desirable) if all patients in the population were to receive the treatment recommended by the decision rule. Due to the patients' diversity in their responsiveness to treatment and vulnerability to adverse effects, the treatment effects are often heterogeneous. A notable limitation of the previous work in this area is that the estimated optimal ITR may be suboptimal or even detrimental to a certain disadvantaged subpopulation. To provide a concrete illustration of this consequence, we consider a simple yet illustrative example in Section 2. In this example, applying the standard optimal ITR is

actually harmful to a significant portion of the population. In the setting of recommending a medical treatment, this severe consequence demands careful attention in order to protect the vulnerable. Motivated by this concern, we propose a new fairness-aware framework that aims to estimate a mean-optimal ITR under the constraint that its induced potential outcome distribution has a lower quantile above a given threshold. For example, we may maximize the average treatment benefit for the whole population while requiring that 95% of the patients benefit from the treatment (say, by requiring the 5th percentile of the potential outcome distribution to be above a prespecified threshold).

Although the proposed fairness-aware optimal ITR (F-ITR) is conceptually intuitive, its computational and statistical theories are highly nontrivial as the optimal F-ITR corresponds to a solution to a nonconvex optimization problem. We consider estimating the optimal F-ITR within a class of stochastic decision rules indexed by a Euclidean parameter. However, we do not require to specify an outcome regression model. Hence, our approach belongs to the category of model-free policy search methods. We study both the static and dynamic ITRs. We derive the asymptotic convergence theory of the proposed estimator using empirical process techniques. Considering the class of stochastic treatment rules alleviates some aspect of the computational challenge. Moreover, we show that doing so will lead to an optimal decision rule as good as the optimal rule within the corresponding class of deterministic decision rules. We prove that the estimated optimal decision rule satisfies the quantile constraint asymptotically, and that its value function converges at a $\mathcal{O}_P(n^{-1/2})$ rate, where n is the sample size. We further develop a new first-order dual algorithm to efficiently compute the estimator. The new algorithm and theory are of independent interest and can be useful for other optimality criteria, such as the composite criterion in Luckett et al. (2017) for balancing multiple and possibly competing outcomes and the robust criterion in Xiao et al. (2019) for achieving robustness against skewed, heterogeneous, heavy-tailed errors or outliers in data.

We point out that the proposed F-ITR framework is also closely related to robust methods for deriving the optimal ITRs. There are some robust methods for deriving optimal ITRs. In particular, Wang et al. (2018a) study the quantile optimal treatment regimes. Linn et al. (2017); Qi et al. (2019a) propose quantile regression approaches to indirectly and

approximately maximize the quantile of outcomes over some classes of decision rules. Wang et al. (2018b) study estimating the mean-optimal treatment regime under the constraint on the mean objectives of risk outcome. Qi et al. (2019b) propose a general decision-rule based risk measure for individualized decision making. However, none of the work above considers both mean and quantile objectives simultaneously, though both objectives can be important in practice.

Paper Organization. The rest of this paper is organized as follows. In Section 2, we present the fairness-oriented individualized treatment regime. In Section 3, we present our estimator. In Section 4, we derive the asymptotic result. We extend our framework to dynamic treatment regimes in Section 5. We conduct extensive numerical studies in Section 6, and we conclude the paper and discuss future directions in Section 7.

2 Fairness-Oriented Optimality Criterion

2.1 Notation and Setup

In this paper, we consider the setting of a binary treatment. We denote the treatment for patient i as $A_i \in \{0, 1\}$. Let Y_i be the corresponding outcome. Without loss of generality, we assume that a larger outcome is preferable. To define the optimal ITR, we adopt the counterfactual (or potential) outcome framework (Rubin, 1978; Splawa-Neyman et al., 1990) in causal inference. Specifically, let $Y_i^*(0)$ be the outcome of patient i had this patient receive treatment 0, and let $Y_i^*(1)$ be defined similarly. Since each patient can only be assigned to one treatment, for patient i, we observe either $Y_i(0)$ or $Y_i(1)$, but not both.

We assume that the observed outcome for patient i is $Y_i = Y_i^*(0)(1 - A_i) + Y_i^*(1)A_i$. In other words, the observed outcome is the outcome corresponding to the treatment the patient actually receives. In causal inference, this is referred as the consistency assumption. Also, we assume that the potential outcome of one patient should not be affected by treatments assigned to the other patients, or the stable unit treatment value assumption (Rubin, 1986).

2.2 A Motivating Example

For illustration, we consider a heteroscedastic outcome regression model $Y_i = 1 + 3A_i + X_i - 5A_iX_i + (1 + A_i + 2A_iX_i)\varepsilon_i$, where the covariate $X_i \sim \text{Unif}[0,1]$, the noise $\varepsilon_i \sim N(0,1)$, and the treatment $A_i = 1$ if patient i receives the treatment, and $A_i = 0$ if patient i is in the control group. We consider the following seven decision rules: (1) $A_i = 0$, for all i; (2) $A_i = \mathbb{1}(X_i \leqslant 3/5)$; (3) $A_i = \mathbb{1}(X_i \leqslant 1/2)$; (4) $A_i = \mathbb{1}(X_i \leqslant 1/5)$; (5) $A_i = \mathbb{1}(X_i \leqslant 1/10)$; (6) $A_i = 1$ for all i, and (7) random assignment $\mathbb{P}(A_i = 1) = 0.5$. We evaluate the performance of the seven decision rules on a large independent sample of size 10^6.

Table 1: Mean and the 0.10-th quantile of the outcomes of the seven different treatment regimes estimated using 10^6 simulated samples

Regime	(1)	(2)	(3)	(4)	(5)	(6)	(7)
mean	1.50	**2.40**	**2.37**	2.00	1.78	2.00	1.74
$Q_{0.10}$	0.16	−0.03	**0.20**	**0.33**	0.26	−2.29	−0.81

Table 1 summarizes the 0.10-th quantile ($Q_{0.10}$) and the mean of potential outcome distribution of each of the seven decision rules. This example was initially considered in Wang et al. (2018a) to illustrate the quantile-optimal ITR, which maximizes $Q_\tau\{Y^*(f)\}$. In this example, the mean optimal treatment regime is Regime 2, and the optimal 0.10-th quantile treatment regime is Regime 4. However, Regime 4 only achieves a mean outcome 2.00, which is about 20% lower than the optimal mean outcome achieved by Regime 2. As an example of the F-ITR, we consider (1) with the 0.10-th quantile being constrained to be non-negative. This leads to Regime 3 as the desired choice among the seven. We observe that the mean of Regime 3 is 2.37 and is very close to the unconstrained optimal mean value 2.40 given by Regime 2. This example demonstrates that in some applications, it is sensible and beneficial to go beyond optimizing a single criterion such as mean or quantile. We further plot the empirical treatment effect distributions of Regime 2 and Regime 3 in Figure 1, and we observe that Regime 3 substantially enlightens the left tail of the distribution. This shows that by slightly reducing the mean, we can potentially achieve much better fairness.

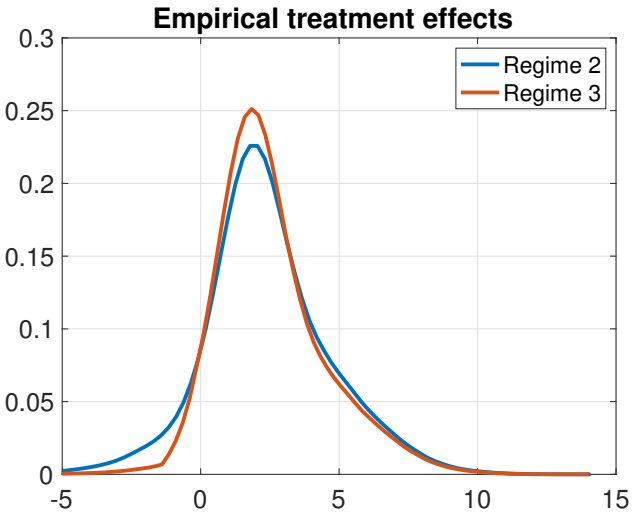

Figure 1: Empirical distributions of treatment Regimes 2 and 3

2.3 Fairness-Oriented Optimality

Our goal is to tailor the treatment recommendation to patient i by considering the patient's individual characteristics, summarized in the covariate vector $\boldsymbol{X}_i \in \mathbb{R}^d$, to achieve certain optimal performance regarding the treatment benefit in the population. An individualized treatment rule (ITR) is a mapping that takes the covariate vector x_i as input and outputs a binary variable in $\{0, 1\}$. For computational efficiencies as discussed in the next section, we consider stochastic ITRs, which output a probability for assigning the treatment. A stochastic ITR can be represented by $f(\cdot, \cdot) : \mathbb{R}^d \times U \to \{0, 1\}$, where $U \sim \text{Uniform}[0, 1]$. Given an ITR $f(\cdot, \cdot)$, the corresponding potential outcome is $Y_i^*(f) = Y_i^*(1)f(\boldsymbol{X}_i, U) + Y_i^*(0)\{1 - f(\boldsymbol{X}_i, U)\}$. Note that from here onward, for ease of presentation, we omit the term U. In particular, $Y_i^*(f)$ is the outcome following treatment regime $f(\cdot, \cdot)$, which assigns patient i to treatment 1 or 0. In this paper, we focus on randomized trials for which A_i and $\big(Y_i^*(1), Y_i^*(0)\big)$ are independent. Our results can be extended to observational studies, as discussed in Remark 7.

To evaluate a treatment regime, existing work has focused on the population mean of the potential outcome distribution, i.e., $\mathbb{E}\{Y^*(f)\}$. We consider a refinement of this metric by enforcing certain fairness constraints. Intuitively, to protect the vulnerable, we would wish the lower tail of the potential outcome distribution should not be inferior. Specifically, the τ-th quantile of $Y^*(f)$ is defined as $Q_\tau\{Y^*(f)\} = \inf\{t : F^*(t) \geqslant \tau\}$, where F^* denotes

the cumulative distribution function of $Y^*(f)$, and $\tau \in (0,1)$ is the quantile level of interest (e.g., the 0.10-quantile). Formally, given a collection $\mathbb{D}$ of treatment regimes, we propose to estimate the following fairness-oriented optimal ITR (F-ITR), which is defined as the solution to the optimization problem

$$\underset{f\in\mathbb{D}}{\text{maximize}}\ \mathbb{E}\{Y^*(f)\},\ \text{subject to}\ \mathcal{Q}_\tau\{Y^*(f)\} \geqslant q, \tag{1}$$

where $q \in \mathbb{R}$ is a pre-specified threshold.

3 Proposed Estimator and Algorithm

3.1 Estimation

We first discuss how to estimate the F-ITR defined in Section 2.2 from a randomized trial. Denote the observed sample set as $\{(\mathbf{x}_i, y_i, a_i)\}_{i=1}^n$, $i = 1, \ldots, n$, where $\mathbf{x}_i \in \mathbb{R}^d$ denotes the covariates, $y_i \in \mathbb{R}$ denotes the response, and $a_i \in \{0,1\}$ denotes the treatment. The optimization problem in (1) is challenging as it involves a nonsmooth nonconvex objective function subjecting to a nonsmooth nonconvex constraint. We show below how the computational challenges can be partly alleviated by considering stochastic ITRs.

Specifically, the stochastic ITR assigns treatment 1 to patient i with probability $f(\mathbf{x}_i, \boldsymbol{\beta})$, where $f(\cdot, \boldsymbol{\beta}) \in \mathbb{D}$, a parametric class of functions. For example, we may adopt the logistic function $f(\mathbf{x}, \boldsymbol{\beta}) = \{1+\exp(-\mathbf{x}^\top\boldsymbol{\beta})\}^{-1}$. There has been substantial recent interest in stochastic ITRs, see Luedtke and van der Laan (2016); Díaz and van der Laan (2018); Kennedy (2019); Díaz and Hejazi (2020); Qiu et al. (2020), among others. Luedtke and van der Laan (2016) overcomes the challenges of an NP-hard knapsack problem by focusing on stochastic ITRs. Qiu et al. (2020) showed that in a nonparametric setting with instrumental variables, the optimal ITR among all stochastic rules is in fact deterministic whenever there is heterogeneity in the average treatment effect across subgroups defined by measured covariates in the population. In Section 4, we also establish a link between the stochastic ITR and the deterministic ITR for the current problem.

For a stochastic ITR induced by $f(\mathbf{x}_i, \boldsymbol{\beta})$, we denote the mean and the τ-th quantile of

the corresponding potential outcome distribution by $\mathcal{M}(\boldsymbol{\beta})$ and $\mathcal{Q}_\tau(\boldsymbol{\beta})$, respectively. The optimal F-ITR is indexed by the parameter $\boldsymbol{\beta}^*$:

$$\boldsymbol{\beta}^* \in \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, \mathcal{M}(\boldsymbol{\beta}), \text{ subject to } \mathcal{Q}_\tau(\boldsymbol{\beta}) \geqslant q, \tag{2}$$

where $\mathcal{M}(\boldsymbol{\beta})$ is the mean of the potential outcome distribution induced by the ITR index by $\boldsymbol{\beta}$. Let $\mu(1, \boldsymbol{X}) = \mathbb{E}[Y_i|a_i = 1, \boldsymbol{X}_i] = \mathbb{E}[Y_i^*(1)|\boldsymbol{X}_i]$ and $\mu(0, \boldsymbol{X}_i) = \mathbb{E}[Y_i|a = 0, \boldsymbol{X}] = \mathbb{E}[Y_i^*(0)|\boldsymbol{X}_i]$. Recalling $f_i(\boldsymbol{\beta}) = \mathbb{P}(a_i = 1|\boldsymbol{X}_i, \boldsymbol{\beta})$, we have

$$\begin{aligned}\mathcal{M}(\boldsymbol{\beta}) &= \mathbb{E}\big[Y_i^*(1)\mathbb{1}(a_i = 1) + Y_i^*(0)\mathbb{1}(a_i = 0)\big] \\ &= \mathbb{E}_{\boldsymbol{X}_i}\big[\mu(1, \boldsymbol{X}_i)f_i(\boldsymbol{\beta}) + \mu(0, \boldsymbol{X}_i)\big(1 - f_i(\boldsymbol{\beta})\big)\big].\end{aligned} \tag{3}$$

Similarly as in Zhang et al. (2012) and Wang et al. (2018a), we can consistently estimate $\mathcal{M}(\boldsymbol{\beta})$ and $\mathcal{Q}_\tau(\boldsymbol{\beta})$ without specifying an outcome regression model by

$$\widehat{\mathcal{M}}(\boldsymbol{\beta}) = \operatorname{argmin}_u n^{-1} \sum_{i=1}^{n} c_i(\boldsymbol{\beta})(y_i - u)^2, \tag{4}$$

and

$$\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}) = \operatorname{argmin}_q n^{-1} \sum_{i=1}^{n} c_i(\boldsymbol{\beta})\rho_\tau(y_i - q) \tag{5}$$

respectively, where $\rho_\tau(u) = u\{\tau - \mathbb{1}(u < 0)\}$ is the quantile loss function, and $c_i(\boldsymbol{\beta}) = a_i f(\mathbf{x}_i, \boldsymbol{\beta}) + (1 - a_i)\{1 - f(\mathbf{x}_i, \boldsymbol{\beta})\}$. We estimate $\boldsymbol{\beta}^*$ by

$$\widehat{\boldsymbol{\beta}} \in \underset{\beta}{\operatorname{argmax}}\, \widehat{\mathcal{M}}(\boldsymbol{\beta}), \text{ subject to } \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}) \geqslant q - C/\sqrt{n}, \tag{6}$$

where C is a positive constant. Note that we introduce the term $C/\sqrt{n}$ to ensure the feasibility of $\mathcal{Q}_\tau(\widehat{\boldsymbol{\beta}})$ for ease of presentation in the theoretical results. However, this term can be dropped in practice.

3.2 Lagrangian Dual Problem and its Properties

Solving problem (6) is computationally challenging as the constraint is nonconvex in $\boldsymbol{\beta}$. To generate a feasible high-quality solution, we consider its Lagrangian dual problem:

$$\underset{\lambda \geqslant 0}{\text{minimize}}\ \mathcal{L}(\lambda) := \max_{\boldsymbol{\beta}} \big\{ \widehat{\mathcal{M}}(\boldsymbol{\beta}) + \lambda \big\{ q - \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}) \big\} \big\}. \tag{7}$$

As this dual problem is convex in λ, which is a scalar, classical methods such as golden section search can be applied to efficiently find the optimal λ^*. Then, the corresponding $\widetilde{\boldsymbol{\beta}}$, defined as

$$\widetilde{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmax}} \big\{ \widehat{\mathcal{M}}(\boldsymbol{\beta}) + \lambda^* \big\{ q - \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}) \big\} \big\}, \tag{8}$$

is the dual optimal solution and satisfies the quantile constraint.

In what follows, we present the key properties of the dual solution $\widetilde{\boldsymbol{\beta}}$ and compare it with the primal solution $\widehat{\boldsymbol{\beta}}$. We first show that the dual solution in (8) indeed exists. We provide the proof in Appendix Section A

Proposition 1. *Assume that there exists a feasible primal optimal solution as in (6). A dual optimal solution $\widetilde{\boldsymbol{\beta}}$ in (8) also exits.*

The next theorem quantifies the duality gap between the primal and dual optimal solutions.

Theorem 2. *Let $\widehat{\boldsymbol{\beta}}$ be the solution to problem (6), and λ^* and $\widetilde{\boldsymbol{\beta}}$ be some optimal Lagrangian multiplier and dual solution to problems (7) and (8), respectively. We have that there exists a $\widetilde{\boldsymbol{\beta}}$ such that the duality gap is bounded by*

$$\big| \widehat{\mathcal{M}}(\widetilde{\boldsymbol{\beta}}) - \widehat{\mathcal{M}}(\widehat{\boldsymbol{\beta}}) \big| \leqslant \lambda^* \big| \widehat{\mathcal{Q}}_\tau(\widehat{\boldsymbol{\beta}}) - \widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) \big|.$$

Proof. See Appendix B for the detailed proof. ■

Remark 3. *As $\widehat{\boldsymbol{\beta}}$ is a feasible solution to (5), $\mathcal{Q}_\tau(\widehat{\boldsymbol{\beta}}) \geqslant q$ and $\mathcal{Q}_\tau(\widehat{\boldsymbol{\beta}}) \in \{y_1, ..., y_n\}$. Without loss of generality, assuming $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_k \leqslant q \leqslant \cdots \leqslant y_n$, we have that the duality gap is upper bounded by $\lambda^* |y_k - \widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}})|$. In the simulation studies in Section 4, we observe that*

our algorithm achieves small duality gaps in different settings under consideration, which demonstrates that the proposed algorithm generates high-quality solutions in practice. We emphasize that although in this paper we adopt $f_i(\boldsymbol{\beta}) = \{1+\exp(\mathbf{x}_i^\top \boldsymbol{\beta})\}^{-1}$*, Theorem 2 actually holds for general choices of continuous* f_i*.*

3.3 Algorithm

We summarize the algorithm below. For the Lagrangian dual problem, during each iteration, given some λ, we solve the follwing problem

$$\underset{\boldsymbol{\beta}}{\text{maximize}}\ \mathcal{D}(\boldsymbol{\beta}) := \widehat{\mathcal{M}}(\boldsymbol{\beta}) + \lambda\big\{\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}) - q\big\} \tag{9}$$

to evaluate the value of the Lagrangian dual function $\mathcal{L}(\lambda)$. In the current literature of precision medicine, this type of problem is commonly solved by genetic algorithms (Whitley, 1994), which are known to be inefficient and lack stability. We propose to solve the problem by an efficient first-order method. In particular, at the t-th iteration, given the current solution $\boldsymbol{\beta}^t$, we compute a corresponding sub(super)-gradient $\boldsymbol{v}^t \in \hat{\partial}\mathcal{D}_s(\boldsymbol{\beta}^t)$. In particular, we first consider $\widehat{\mathcal{M}}(\boldsymbol{\beta})$ in (4). By straightforward calculation, we obtain

$$\widehat{\mathcal{M}}(\beta) = n^{-1}\sum_{i=1}^{n} c_i(\boldsymbol{\beta}) y_i.$$

Note that $c_i(\boldsymbol{\beta})$ is smooth due to the use of stochastic ITR. We can thus compute the gradient of $\widehat{\mathcal{M}}(\boldsymbol{\beta}^t)$ efficiently, which we denote as $\boldsymbol{v}_M^t$. Next, we consider $\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta})$. Due to the discontinuity of the sample quantile function, we replace the indicator function in in the quantile loss function $\rho_\tau(u) = u\{\tau - \mathbb{1}(u<0)\}$ by a sigmoid function.

Taking $\boldsymbol{v}^t = \boldsymbol{v}_M^t + \lambda \boldsymbol{v}_Q^t$, we update the solution by the following gradient step

$$\boldsymbol{\beta}^{t+1} = \boldsymbol{\beta}^t + \alpha_t \boldsymbol{v}^t,$$

where α_t is a prespecified stepsize. This algorithm can be implemented efficiently and displays satisfactory performance in our simulation experiments.

4 Statistical Theory

In this section, we present the statistical theory of the proposed estimator. In particular, we show that the estimator $\widehat{\boldsymbol{\beta}}$ achieves the optimal risk asymptotically.

Note that we consider a class of stochastic ITRs, a more general class of decision rules than deterministic ITRs. It is not difficult to see that the expected risk achieved by optimal stochastic decision rule is always lower bounded by the risk achieved by the optimal deterministic decision rule. In addition, consider linear deterministic ITRs of the form $\widetilde{f}_i(\boldsymbol{\beta}) = \mathbb{1}(\mathbf{x}_i^\top \boldsymbol{\beta} > 0)$ and the corresponding stochastic ITRs $f_i(\boldsymbol{\beta}) = \mathbb{P}(a_i = 1|\mathbf{x}_i, \boldsymbol{\beta}) = \{1 + \exp(-\mathbf{x}_i^\top \boldsymbol{\beta})\}^{-1}$, the latter of which are our primary focus as discussed earlier. We show in Proposition 4 below that if some linear deterministic decision rule achieves optimal risk and satisfies the quantile constraint, there exists a linear stochastic decision rule that approximates the risk and constraint up to arbitrary precision. Throughout our discussions, we assume that the response y_i's and the covariates $\mathbf{x}_i$'s are bounded.

Proposition 4. *Suppose there exists a $\check{\boldsymbol{\beta}}$, which is an optimal solution to the problem that*

$$\check{\boldsymbol{\beta}} = \operatorname*{argmax}_{\boldsymbol{\beta}} \mathcal{M}(\boldsymbol{\beta}), \quad \text{subject to } \mathcal{Q}_\tau(\boldsymbol{\beta}) \geqslant q, \tag{10}$$

where the deterministic ITR $f_i = \mathbb{1}(\mathbf{x}_i^\top \boldsymbol{\beta} > 0)$ is adopted. Considering problem (6), where stochastic ITR $f(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbb{P}(f_i = 1) = \{1 + \exp(-\mathbf{x}_i^\top \boldsymbol{\beta})\}^{-1}$ is adopted, denote by $\widehat{\boldsymbol{\beta}}$ an optimal solution to problem (6). We have that for any given $\varepsilon > 0$, as $n \to +\infty$, $\widehat{\boldsymbol{\beta}}$ satisfies the constraint of problem (2), and $\mathbb{P}(\mathcal{M}(\widehat{\boldsymbol{\beta}}) \geqslant \mathcal{M}(\check{\boldsymbol{\beta}}) - \varepsilon) \to 1$.

Proof. See Appendix C for the detailed proof. ■

The following theorem proves that the risk incurred by $\widehat{\boldsymbol{\beta}}$ converges to the optimal risk $\mathcal{R}^*$. To derive the consistency, we employ the empirical process techniques. The detailed proof is presented in Appendix D.

Theorem 5. *Suppose $\boldsymbol{\beta}^* \in \mathcal{B}$, where $\mathcal{B}$ is a compact set. Then*

$$\lim_{n\to\infty} \left\{ \mathcal{M}(\widehat{\boldsymbol{\beta}}) - \sup_{\boldsymbol{\beta}\in\widetilde{\mathcal{Q}}} \mathcal{M}(\boldsymbol{\beta}) \right\} = 0$$

in probability, where $\widetilde{\mathcal{Q}}$ *denotes the closure of* $\mathcal{Q} = \{\boldsymbol{\beta} : \mathcal{Q}_\tau(\boldsymbol{\beta}) \geqslant q\}$. *Thus, if* $\boldsymbol{\beta}^*$ *belongs to the closure of* $\widetilde{\mathcal{Q}}$, *then* $\lim_{n\to\infty} \widehat{\mathcal{M}} = \mathcal{M}^*$ *in probability, where* $\widehat{\mathcal{M}}$ *and* $\mathcal{M}^*$ *denotes the estimator's corresponding population mean treatment results and optimal mean treatment results under the quantile constraint, respectively.*

The next theorem derives the convergence rate of $\mathcal{M}(\widehat{\boldsymbol{\beta}}) - \mathcal{M}(\boldsymbol{\beta}^*)$.

Theorem 6. *Suppose that* $\boldsymbol{\beta}^*$ *belongs to a compact set* $\mathcal{B}(M)$, *where* $M > 0$ *is a constant. Then we have that* $\forall\ \tau \geqslant 1$ *we have*

$$\mathbb{P}^*\big(\mathcal{M}(\widehat{\boldsymbol{\beta}}) \geqslant \mathcal{M}(\boldsymbol{\beta}^*) - \varepsilon\big) \geqslant 1 - e^{-\tau},$$

where $\mathbb{P}^*$ *denotes the outer probability for possibly nonmeasureable sets, and* $\varepsilon = \mathcal{O}(n^{-1/2})$. *Or, equivalently,*

$$|\mathcal{M}(\widehat{\boldsymbol{\beta}}) - \mathcal{M}(\boldsymbol{\beta}^*)| = \mathcal{O}_P(n^{-1/2}).$$

Proof. See Appendix E for the detailed proof. ■

Remark 7. *The proposed method and theoretical results can be extended to observational studies using the propensity score weighting approach. Assume the popular no unmeasured confounder assumption* $\{Y^*(1), Y^*(0)\} \perp A|\boldsymbol{X}$ *holds. Leting* $\pi(\boldsymbol{X}) = \mathbb{P}(A = 1|\boldsymbol{X})$, *the propensity score* $\mathbb{P}(C(\boldsymbol{\beta}) = 1|\boldsymbol{X}) = \pi(\boldsymbol{X})f(\boldsymbol{X}, \boldsymbol{\beta}) + (1 - \pi(X))(1 - f(\boldsymbol{X}, \boldsymbol{\beta}))$. *Denoting the propensity score by* $\pi_c(\boldsymbol{X}, \boldsymbol{\beta})$, *we estimate the mean and the* τ*-th quantile of the treatment effect by* $\operatorname{argmin}_u n^{-1} \sum_{i=1}^n \frac{c(\mathbf{x}_i, \boldsymbol{\beta})}{\widehat{\pi}_c(\mathbf{x}_i, \boldsymbol{\beta})}(\boldsymbol{\beta})(y_i - u)^2$ *and* $\operatorname{argmin}_q n^{-1} \sum_{i=1}^n \frac{c(\mathbf{x}_i, \boldsymbol{\beta})}{\widehat{\pi}_c(\mathbf{x}_i, \boldsymbol{\beta})} c(\mathbf{x}_i, \boldsymbol{\beta})\rho_\tau(y_i - q)$, *respectively, where* $\widehat{\pi}_c(\mathbf{x}_i, \boldsymbol{\beta})$ *is an estimator of the propensity score* $\pi_c(\mathbf{x}_i, \boldsymbol{\beta})$. *In our simulation, we use the logistic regression to estimate* $\pi_c(\mathbf{x}_i, \boldsymbol{\beta})$, *where we model* $\pi(\boldsymbol{X})$ *as* $\pi(\boldsymbol{X}, \boldsymbol{\gamma}) = \exp(\boldsymbol{X}^\top\boldsymbol{\gamma})/(1+\exp(\boldsymbol{X}^\top\boldsymbol{\gamma}))$. *In practice, we may use other semiparametric or nonparametric models for the estimation.*

Remark 8. *In practice, the future patient population may not be exactly the same as the training samples, for example, a slight shift of age or other covariates. To solve this challenge, we refer the readers to Mo et al. (2020), which addresses the covariate shift problem.*

5 Extension to Dynamic Treatment Regime

We extend the proposed F-ITR to the dynamic setting, which involves a sequence of decision rules. For example, in treating a chronic disease, the patient's condition often needs to be re-evaluated over time. Depending on the patient's clinical information and how he/she responds to the previous treatment, the doctor may need to adapt the treatment decision.

For ease of presentation, we consider a two-stage dynamic setting, but our methods and results can be extended to the general T-stage case by induction. Assume that patient i receives treatment $a_i^{(1)} \in \{0,1\}$ at stage 1 and treatment $a_i^{(2)} \in \{0,1\}$ at stage 2. At the end of stage 2, we observe the outcome Y_i, based on which the overall treatment effect will be evaluated. A dynamic ITR has the form $f = \{f^{(1)}, f^{(2)}\}$, such that $f^{(j)}$ is a function of all information available before making the j-th decision. We denote the baseline covariate vector as $\boldsymbol{X}_i^{(1)}$, and denote the covariate at the second stage as $\boldsymbol{X}_i^{(2)}$, which may depend on $\boldsymbol{X}_i^{(1)}$ and $a_i^{(1)}$ and may include intermediate outcomes. Let $\boldsymbol{H}_i^{(1)} = \{\boldsymbol{X}_i^{(i)}\}$ and $\boldsymbol{H}_i^{(2)} = \{\boldsymbol{X}_i^{(1)}, a_i^{(1)}, \boldsymbol{X}_i^{(2)}\}$. Throughout our discussion, we adopt the no unmeasured confounder or sequential ignorability assumption, that is, conditioning on the history, the treatment is independent of any future information, see Robins (1997) for details. In addition, we adopt the positivity assumption that there exist positive constants $c_1 < c_2$ such that $c_1 \leqslant \mathbb{P}(a_j = a|H_j) \leqslant c_2$ for $a \in \{0,1\}$ and $j = 1, 2$.

Here our goal is to estimate the optimal dynamic ITR, which is defined as the one that maximized the average final outcome under the constraint that a lower quantile of the potential outcome distribution of the ITR exceeds a given threshold. Consider the class of candidate ITRs index by $\boldsymbol{\beta} = \{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}\}$, and $f^{(j)}(\boldsymbol{H}_i^{(j)}|\boldsymbol{\beta}^{(j)}) = \mathbb{P}(a_i^{(j)} = 1|\boldsymbol{H}_i^{(j)}) = \{1 + \exp(-\boldsymbol{H}_i^{(j)\top}\boldsymbol{\beta}^{(j)})\}^{-1}$, $j = 1, 2$. Given a $\boldsymbol{\beta}$, we denote the corresponding stochastic sequential decision rule by $d(\boldsymbol{\beta}) = \big(d_1(\boldsymbol{H}^{(1)}|\boldsymbol{\beta}^{(1)}), d_2(\boldsymbol{H}^{(2)}|\boldsymbol{\beta}^{(2)})\big)$, where $d_1(\boldsymbol{H}^{(1)}|\boldsymbol{\beta}^{(1)}) = \text{Bernoulli}\big(f^{(1)}(\boldsymbol{H}_i^{(1)}|\boldsymbol{\beta}^{(1)})\big)$ and $d_2(\boldsymbol{H}^{(2)}|\boldsymbol{\beta}^{(2)}) = \text{Bernoulli}\big(f^{(2)}(\boldsymbol{H}_i^{(2)}|\boldsymbol{\beta}^{(2)})\big)$. We sometimes write $d(\boldsymbol{\beta}) = (d_1(\boldsymbol{\beta}^{(1)}), d_2(\boldsymbol{\beta}^{(2)}))$ for brevity.

For sample i, the potential final outcomes are denoted as $\{Y_i^*(1,1), Y_i^*(1,0), Y_i^*(0,1), Y_i^*(0,0)\}$, corresponding to the four possible treatment sequences. Given a dynamic ITR $d(\boldsymbol{\beta})$, the potential intermediate information is denoted by $\boldsymbol{X}_i^{(2)*}(d_1(\boldsymbol{\beta}^{(1)}))$ and the potential final out-

come is denoted by $Y^*(d(\boldsymbol{\beta}))$. We have

$$\begin{aligned} Y^*\big(d(\boldsymbol{\beta})\big) = {} & Y^*(1,1)\mathbb{1}\big(d_1(\boldsymbol{\beta}^{(1)}) = 1, d_2(\boldsymbol{\beta}^{(2)}) = 1\big) \\ & + Y^*(1,0)\mathbb{1}\big(d_1(\boldsymbol{\beta}^{(1)}) = 1, d_2(\boldsymbol{\beta}^{(2)}) = 0\big) \\ & + Y^*(0,1)\mathbb{1}\big(d_1(\boldsymbol{\beta}^{(1)}) = 0, d_2(\boldsymbol{\beta}^{(2)}) = 1\big) \\ & + Y^*(0,0)\mathbb{1}\big(d_1(\boldsymbol{\beta}^{(1)}) = 0, d_2(\boldsymbol{\beta}^{(2)}) = 0\big). \end{aligned}$$

Thus, we have that the population mean of the potential outcome distribution induced by $d(\boldsymbol{\beta})$ is:

$$\begin{aligned} \mathcal{M}(\boldsymbol{\beta}) &= \mathbb{E}\big[Y^*\big(d(\boldsymbol{\beta})\big)\big] \\ &= \mathbb{E}\big[f^{(1)}(\boldsymbol{H}_i^{(1)}, \boldsymbol{\beta}^{(1)}) f^{(2)}(\boldsymbol{H}_i^{(2)*}, \boldsymbol{\beta}^{(2)}) \mu(1, 1, \boldsymbol{H}^{(2)*}) \\ &\quad + f^{(1)}(\boldsymbol{H}_i^{(1)}, \boldsymbol{\beta}^{(1)})(1 - f^{(2)}(\boldsymbol{H}_i^{(2)*}, \boldsymbol{\beta}^{(2)})) \mu(1, 0, \boldsymbol{H}^{(2)*}) \\ &\quad + (1 - f^{(1)}(\boldsymbol{H}_i^{(1)}, \boldsymbol{\beta}^{(1)})) f^{(2)}(\boldsymbol{H}_i^{(2)*}, \boldsymbol{\beta}^{(2)}) \mu(0, 1, \boldsymbol{H}^{(2)*}) \\ &\quad + (1 - f^{(1)}(\boldsymbol{H}_i^{(1)}, \boldsymbol{\beta}^{(1)}))(1 - f^{(2)}(\boldsymbol{H}_i^{(2)*}, \boldsymbol{\beta}^{(2)})) \mu(0, 0, \boldsymbol{H}^{(2)*})\big], \end{aligned}$$

where $\mu(j_1, j_2, \boldsymbol{H}^{(2)*}) = \mathbb{E}\big[Y^*(j_1, j_2) \big| \boldsymbol{X}_i^{(1)}, A_i^{(1)} = j_1, \boldsymbol{X}_i^{(2)*}(j_1)\big]$, $j_1, j_2 \in \{0, 1\}$.

To estimate $\mathcal{M}(\boldsymbol{\beta})$, suppose we have a sample $\{\mathbf{x}_i^{(1)}, a_i^{(1)}, \mathbf{x}_i^{(2)}, a_i^{(2)}, y_i\}_{i \in [n]}$, and let $\mathbf{h}_i^{(1)} = \mathbf{x}_i^{(1)}$ and $\mathbf{h}_i^{(2)} = (\mathbf{x}_i^{(1)\top}, a_i^{(1)}, \mathbf{x}_i^{(2)})^\top$. We further assume that the sample is from the sequential multiple assignment randomized trial (SMART) (Murphy, 2008; Lavori and Dawson, 2000) with

$$\pi_1(\boldsymbol{h}_i^{(1)}) = \mathbb{P}(a^{(1)} | \boldsymbol{h}_i^{(1)}) = \pi_1, \text{ and } \pi_2(\boldsymbol{h}_i^{(2)}) = \mathbb{P}(a^{(2)} | \boldsymbol{h}_i^{(2)}) = \pi_2,$$

where $\pi_1, \pi_2 \in (0, 1)$ are two known constants. Let

$$\begin{aligned} c_i(\boldsymbol{\beta}) = {} & \frac{a_i^{(1)} a_i^{(2)}}{\pi_1 \pi_2} \cdot f_i^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)}) f_i^{(2)}(\mathbf{h}_i^{(2)}, \boldsymbol{\beta}^{(2)}) \\ & + \frac{a_i^{(1)}(1 - a_i^{(2)})}{\pi_1(1 - \pi_2)} \cdot f_i^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)}) \big(1 - f_i^{(2)}(\mathbf{h}_i^{(2)}, \boldsymbol{\beta}^{(2)})\big) \\ & + \frac{(1 - a_i^{(1)}) a_i^{(2)}}{(1 - \pi_1)\pi_2} \cdot \big(1 - f_i^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)})\big) f_i^{(2)}(\mathbf{h}_i^{(2)}, \boldsymbol{\beta}^{(2)}) \\ & + \frac{(1 - a_i^{(1)})(1 - a_i^{(2)})}{(1 - \pi_1)(1 - \pi_2)} \cdot \big(1 - f_i^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)})\big) \big(1 - f_i^{(2)}(\mathbf{h}_i^{(2)}, \boldsymbol{\beta}^{(2)})\big). \end{aligned} \tag{11}$$

Then, we estimate $\mathcal{M}(\boldsymbol{\beta})$ by

$$\widehat{\mathcal{M}}(\boldsymbol{\beta}) = \text{argmin}_{\mu} n^{-1} \sum_{i=1}^{n} c_i(\boldsymbol{\beta})(y_i - \mu)^2.$$

Similarly, we estimate the τ-th quantile of the outcome by

$$\widehat{\mathcal{Q}}_{\tau}(\boldsymbol{\beta}) = \text{argmin}_{q} n^{-1} \sum_{i=1}^{n} c_i(\boldsymbol{\beta}) \rho_{\tau}(y_i - q).$$

Then, our estimator for the F-ITR is given by

$$\widehat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{\beta}} \widehat{\mathcal{M}}(\boldsymbol{\beta}), \text{ subject to } \widehat{\mathcal{Q}}_{\tau}(\boldsymbol{\beta}) \geqslant q - C/\sqrt{n}. \tag{12}$$

Let $\mathcal{M}(\boldsymbol{\beta}^*)$ be the optimal risk satisfying the quantile constraint that $\mathcal{Q}_{\tau}(\boldsymbol{\beta}^*) \geqslant q$. We derive the risk consistency by showing that $\mathcal{M}(\widehat{\boldsymbol{\beta}})$ converges to $\mathcal{M}(\boldsymbol{\beta}^*)$ as sample size n goes to infinity. The results hold for general T-stage problems by an induction argument.

Proposition 9. *Suppose there exists $\breve{\boldsymbol{\beta}} = \{\breve{\boldsymbol{\beta}}^{(1)}, \breve{\boldsymbol{\beta}}^{(2)}\}$, which is the optimal solution to the problem dynamic treatment regime problem at the population level, and the deterministic decision rule is adopted. Considering problem (12), where stochastic decision rule is adopted, denote by $\widehat{\boldsymbol{\beta}} = \{\widehat{\boldsymbol{\beta}}^{(1)}, \widehat{\boldsymbol{\beta}}^{(2)}\}$ an optimal solution to (12). We have that for any given $\varepsilon > 0$, as $n \to \infty$, $\widehat{\boldsymbol{\beta}}$ satisfies the constraint of problem (12), and $\mathbb{P}(\mathcal{M}(\widehat{\boldsymbol{\beta}}) \geqslant \mathcal{M}(\breve{\boldsymbol{\beta}}) - \varepsilon) \to 1$.*

Proof. The proof is similar to the proof of Proposition 4, and we omit it here. ■

Next, we present the convergence rate. The proof is similar to the proof of Theorem 6..

Theorem 10. *Suppose that the optimal regime $\boldsymbol{\beta}^* = \{\boldsymbol{\beta}^{(1)*}, \boldsymbol{\beta}^{(2)*}\}$ belongs to a compact set $\mathcal{B}(M)$, where $M > 0$ is a constant. Then we have that for all $\xi \geqslant 1$ we have*

$$\mathbb{P}^*\big(\mathcal{M}(\widehat{\boldsymbol{\beta}}) \geqslant \mathcal{M}(\boldsymbol{\beta}^*) - \varepsilon\big) \geqslant 1 - e^{-\xi},$$

where $\mathbb{P}^*$ denotes the outer probability for possibly nonmeasureable sets, and $\varepsilon = \mathcal{O}(n^{-1/2})$. Or, equivalently,

$$|\widehat{\mathcal{M}}(\widehat{\boldsymbol{\beta}}) - \mathcal{M}(\boldsymbol{\beta}^*)| = \mathcal{O}_P(n^{-1/2}).$$

Remark 11. *In this section, we focus on the scenario where the final outcome is of interest. In Appendix F, we extend the analysis to a different scenario where a reward is observed at each stage and the goal is to optimize the total reward while constraining the quantile of the potential outcome distribution at each stage to exceed some threshold. Meanwhile, suppose the goal is maximizing the sum of outcome of stages 1 and 2. Denote the potential outcome for sample i at stages 1 and 2 with actions $a_i^{(1)}$ and $a_i^{(2)}$ by $Y_i^{(1)}(a_i^{(1)})$ and $Y_i^{(2)}(a_i^{(1)}, a_i^{(2)})$. We may just replace the original potential outcome $Y_i^*(a_i^{(1)}, a_i^{(2)})$ by $Y_i^{(1)}(a_i^{(1)}) + Y_i^{(2)}(a_i^{(1)}, a_i^{(2)})$. Then the proposed method still applies.*

6 Numerical Results

In this section, we conduct extensive numerical studies using both synthetic and real datasets to investigate the empirical performance of our proposed approach. Our studies demonstrate that the proposed methods achieve desired quantile constraints and obtain desirable mean outcomes.

6.1 Monte Carlo Studies

Example 1 (Static ITR) We generate the random outcome y_i from the following heteroscedastic outcome regression model

$$y_i = 1 + x_{i1} - x_{i2} + x_{i3}^3 + e^{x_{i4}} + a_i(3 - 5x_{i1} + 2x_{i2} - 3x_{i3} + x_{i4}) + \big\{1 + a_i(1 + x_{i1} + x_{i2} + x_{i3} + x_{i4})\big\}\varepsilon_i,$$

$i = 1, \ldots, n$, where the x_{ij}'s are independently generated from the Uniform$(0, 1)$ distribution, and the treatment indicator a_i satisfies $\log\big\{\mathbb{P}(a_i = 1|\mathbf{x}_i)/\mathbb{P}(a_i = 0|\mathbf{x}_i)\big\} = -0.5 - 0.5(x_{i1} + x_{i2} + x_{i3} + x_{i4})$. We consider two different distributions for the random error ε_i: the standard normal distribution, and a highly non-symmetric distribution $\chi_5^2 - 5$. We consider two different sample sizes $n = 500$ or 1000.

We consider the class of treatment regimes $\mathbb{P}(a_i = 1|\mathbf{x}_i) = 1/\{1 + \exp(-\mathbf{x}_i^T\eta)\}$. For each combination of error distribution and sample size n, we consider two different choices of τ (0.1 and 0.25). For each τ, we consider two choices of q. We aim to answer three

Table 2: Simulation studies of F-ITR. Under different quantile constraints, where we require the τ-th quantile of the treatment effect is at least q, we report the averaged treatment effects of sample mean $\mathcal{M}_{\text{mean}}$, sample quantile $\mathcal{Q}_\tau$, sample duality gap (Dual), the corresponding population mean treatment $\mathbb{E}(\mathcal{M}_{\text{mean}})$, the population quantile $\mathbb{E}(\mathcal{Q}_\tau)$, and the percentage of infeasible cases (IF) among the total 1000 simulations.

Error	n	τ	q	$\mathcal{M}_{\text{mean}}$	$\mathcal{Q}_\tau$	Dual	$\mathbb{E}(\mathcal{M}_{\text{mean}})$	$\mathbb{E}(\mathcal{Q}_\tau)$	IF
$N(0,1)$	500	0.10	1.00	4.03	1.03	0.04	3.86	1.04	0.8%
				(0.25)	(0.07)	(0.08)	(0.15)	(0.18)	NA
			1.20	3.83	1.23	0.05	3.74	1.22	6.7%
				(0.25)	(0.03)	(0.16)	(0.14)	(0.13)	NA
		0.25	1.90	4.07	1.95	0.04	4.05	2.00	0
				(0.17)	(0.26)	(0.02)	(0.09)	(0.10)	NA
			2.00	4.04	2.05	0.04	4.01	2.08	0
				(0.12)	(0.15)	(0.02)	(0.03)	(0.06)	NA
	1000	0.10	1.00	3.93	1.03	0.03	3.88	1.06	0
				(0.15)	(0.02)	(0.08)	(0.07)	(0.15)	NA
			1.20	3.76	1.23	0.03	3.74	1.26	3.2%
				(0.18)	(0.01)	(0.03)	(0.16)	(0.10)	NA
		0.25	1.90	4.01	1.94	0.03	4.04	1.95	0
				(0.08)	(0.16)	(0.03)	(0.02)	(0.04)	NA
			2.00	4.06	2.04	0.04	3.97	2.06	0
				(0.11)	(0.17)	(0.02)	(0.02)	(0.05)	NA
χ^2_5	500	0.10	-0.75	3.30	-0.68	0.06	3.14	-0.75	6.6%
				(0.59)	(0.14)	(0.08)	(0.36)	(0.33)	NA
			-0.80	3.61	-0.74	0.05	3.53	-0.85	4.2%
				(0.50)	(0.06)	(0.15)	(0.24)	(0.17)	NA
		0.25	0.45	3.65	0.52	0.07	3.60	0.43	9.6%
				(0.45)	(0.05)	(0.22)	(0.30)	(0.22)	NA
			0.50	3.52	0.58	0.07	3.44	0.48	10.7%
				(0.39)	(0.12)	(0.16)	(0.31)	(0.15)	NA
	1000	0.10	-0.75	3.39	-0.68	0.05	3.28	-0.71	3.4%
				(0.27)	(0.03)	(0.05)	(0.11)	(0.12)	NA
			-0.80	3.46	-0.75	0.06	3.36	-0.82	1.3%
				(0.32)	(0.03)	(0.09)	(0.22)	(0.11)	NA
		0.25	0.45	3.71	0.50	0.04	3.64	0.43	6.7%
				(0.28)	(0.09)	(0.07)	(0.15)	(0.10)	NA
			0.50	3.62	0.54	0.07	3.54	0.47	8.3%
				(0.35)	(0.03)	(0.07)	(0.16)	(0.16)	NA

essential questions through this Monte Carlo study: (1) How well does the proposed F-ITR meet the quantile constraint for fairness protection? (2) How does the F-ITR compare with

the traditional mean-optimal ITR without fairness constraint? (3) How does the proposed algorithm work comparing with the traditional genetic algorithm?

Table 2 summarizes the results based on 1000 simulations runs, including the average of the estimated mean value ($\mathcal{M}_{\text{mean}}$) and the average of the estimated τ-th quantile ($\mathcal{Q}_\tau$) based on the sample. We also report the averaged duality gap (Dual) for each setting, which is an estimate of the optimality of the achieved objectives. Furthermore, using a large independent Monte Carlo sample of size one million, we evaluate the expected mean value ($\mathbb{E}(\mathcal{M}_{\text{mean}})$) and the τ-th quantile ($\mathbb{E}(\mathcal{Q}_\tau)$) when the estimated ITR is used to assign treatment for each individual in the sample. We also report the number of infeasible (IF) cases in the last column, where our algorithm fails to find a feasible solution among 1000 runs. We point out that the infeasibility of the problem may due to the random samples that no such regime satisfying the quantile constraint exists.

Table 3: Quantitative comparisons of the proposed algorithm (Alg) and the genetic algorithm. We report the averaged objective value achieved by different algorithms, and the averaged running times in seconds after repeating the simulation 1000 times, the values in parentheses correspond to the sample standard deviations.

			Mean		$\tau = 0.10$		$\tau = 0.25$	
Error	n	Method	Obj.	Time(s)	Obj.	Time(s)	Obj.	Time(s)
$N(0,1)$	500	Alg.	4.19	3.01	1.39	3.08	2.29	3.07
			(0.24)	(0.18)	(0.15)	(0.20)	(0.04)	(0.25)
		Genetic	4.14	5.97	1.24	23.03	2.04	22.14
			(0.18)	(1.61)	(0.12)	(2.79)	(0.05)	(3.16)
	1000	Alg.	4.08	4.54	1.42	4.46	2.29	4.66
			(0.13)	(0.38)	(0.05)	(0.38)	(0.07)	(0.16)
		Genetic	4.20	6.95	1.22	33.18	2.04	31.08
			(0.13)	(1.64)	(0.07)	(4.97)	(0.05)	(2.92)
$\chi^2_5 - 5$	500	Alg.	3.70	2.98	-0.56	3.08	0.68	2.89
			(0.45)	(0.24)	(0.12)	(0.23)	(0.15)	(0.17)
		Genetic	3.78	6.42	-0.91	23.94	0.51	20.35
			(0.60)	(1.24)	(0.15)	(4.76)	(0.16)	(3.72)
	1000	Alg.	3.87	3.15	-0.47	4.03	0.58	4.08
			(0.42)	(0.20)	(0.12)	(0.28)	(0.13)	(0.26)
		Genetic	3.91	7.10	-0.83	39.10	0.53	31.85
			(0.37)	(1.61)	(0.12)	(3.75)	(0.11)	(8.03)

First, the values $\mathcal{Q}_\tau$ and $\mathbb{E}(\mathcal{Q}_\tau)$ reported in Table 2 confirm that the fairness constraints

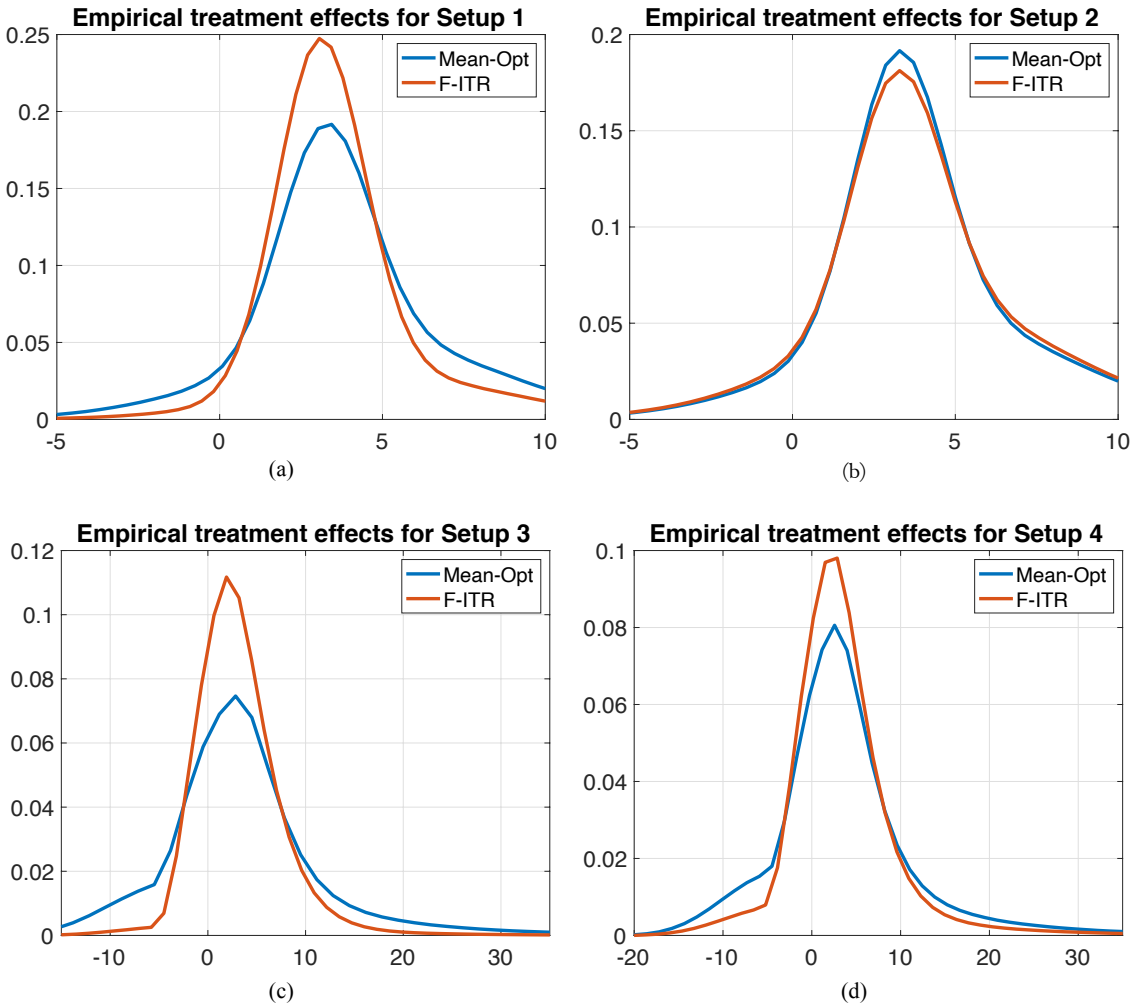

Figure 2: Empirical treatment effect distribution of mean optimal treatment regime (Mean-Opt) and F-ITR under four setups, where we estimate the regimes using $n = 500$ samples, and test on 1 million samples. In setups 1 and 2, the errors are from $N(0, 1)$. In setup 1 we set $\tau = 0.1$, $q = 1$, and in setup 2, we set $\tau = 0.25$, $q = 2$. In setups 3 and 4, the errors are from $\chi^2_5 - 5$. In setup 3, we set $\tau = 0.1$, $q = -0.75$, and in setup 4, we set $\tau = 0.25$, $q = 0.5$.

are satisfied. Second, we compare the performance of F-ITR with M-ITR (mean-optimal ITR without fairness constraint) based on the testing samples. Figure 2 displays the density plots of the estimated potential outcome distribution of F-ITR and that of M-ITR under four different setups considered in the Monte Carlo experiment. It is observed that in setups 1, 3, and 4, F-ITR achieves substantially lower left-tail densities comparing with M-ITR while does not reduce the mean performance significantly. In setup 2, we observe that the two distributions do not differ significantly. This is because that the mean optimal treatment regime also satisfies the quantile constraint. In this case, we observe that F-ITR almost has the same distribution as M-ITR. These figures demonstrate that the proposed F-ITR leads to improved performance at the left tail, with little sacrifice the overall average benefits. Furthermore, we observe that when the quantile constraint is relatively relaxed, the achieved mean value is very close to that of M-ITR. These observations demonstrate strong evidence supporting the benefits of the proposed F-ITR.

Finally, we evaluate the performance and computational speed of the proposed new algorithm. We apply the proposed algorithm and the genetic algorithm to estimate the mean-optimal ITR (M-ITR) and quantile-optimal ITR (maximizing the τ-th quantile of the potential outcome distribution). Note that the genetic algorithm is not applicable to F-ITR. Table 3 compares the average computational time and estimated values based on 100 simulation runs. For M-ITR, the new algorithm achieves similar values as the genetic algorithm does, while reducing the computational time by about one third. For Q-ITR, the new algorithm achieves significantly better values and only requires a small fraction of the computational time of the genetic algorithm.

Example 2 (Dynamic ITR) We consider a two-stage example and generate the data from the model

$$y_i = 1+x_{i1}+a_{i1}\{1-3(x_{i1}-0.2)^2\}+x_{i2}+a_{i2}\{1-5(x_{i2}-0.4)^2\}+\big(1+0.5a_{i1}-a_{i1}x_{i1}+0.5a_{i2}-a_{i2}x_{i2}\big)\varepsilon_i,$$

where x_{i1} and $x_{i2}|\{x_{i1}, a_{i1}\}$ are generated from uniform distributions on [0,1] and $[x_{i1}, x_{i1}+1]$, respectively, and $a_{i1}|x_{i1}$ and $a_{i2}|\{x_{i1}, a_{i1}, x_{i2}\}$ are generated from Bernoulli(expit$(-0.5+x_{i1})$)

and Bernoulli(expit$(-1+x_{i2})$), respectively. Similarly as in the previous example, we consider two different distributions for the random error ε_i: the standard normal distribution $N(0,1)$ and the asymmetric $0.5 \cdot (\chi^2_5 - 5)$ distribution. We consider sequential ITRs of the form (A_1, A_2), where $A_1 = \mathbf{1}\{c_1 X_1 + b_1 < 0\}$, and $A_2 = \mathbf{1}\{c_2 X_2 + b_2 < 0\}$. We consider the dynamic F-ITR in Appendix F. We aim to optimize the mean value under different quantile constraints $\mathcal{Q}_\tau \geqslant q$ for different τ and q. We consider sample size $n = 1000$ or 2000

The simulation results are summarized in Table 4. For both normal and chi-square errors, we provide in the first line the mean of the M-ITR (without constraint) together with its 10% and 25%-th quantiles as benchmarks. It is observed that for both error distributions, the F-ITR has only slightly smaller mean than M-ITR but satisfies the conditional quantile constraints well. In contrast, for normal error the M-ITR has negative 0.10 quantile, for the chi-square error, the M-ITR has negative 0.10 and 0.25 quantiles, suggesting that the M-ITR may have undesirable effects for fragile individuals. Our proposed F-ITR achieve the desired constraints, and achieve near-optimal mean treatment effects as the duality gap is small.

6.2 Application

We apply the proposed method to analyze the ACTG175 dataset from the R package `speff2trial`. This dataset contains 2,139 HIV-infected patients. These patients are randomly assigned to one of the four treatments including zidovudine (AZT) monotherapy, AZT+didanosine(ddI), AZT+zalcitabine(ddC), and ddI monotherapy. The goal of the trial is to determine if the treatment with one drug (monotherapy) is better than the treatment with two drugs (combination therapy) in patients with CD4-T cells between 200 and $500/mm^3$. See Hammer et al. (1996) for more details.

In exploratory analysis, we observe heteroscedastic treatment effects and high asymmetry in the outcome variable distribution. It was known in the medical literature that the patients who had taken AZT before entering the trial, treated with ddI or AZT+ddI are better than continuing to take AZT alone. We thus consider the problem of how to assign treatment to the patients who had taken AZT before the trial, to either continued treatment with AZT+ddI combination or the ddI monotherapy. Denote $A_i = 1$ if patient i is assigned to the AZT+ddI therapy, and $A_i = 0$ if the patient is assigned to the ddI monotherapy. The

Table 4: Simulation studies of F-ITR in dynamic settings. Under different quantile constraints, where we require the τ-th quantile of the treatment effect is at least q, we report the averaged treatment effects of sample mean $\mathcal{M}_{\text{mean}}$, sample quantile $\mathcal{Q}_\tau$, sample duality gap (Dual), the corresponding population mean treatment $\mathbb{E}(\mathcal{M}_{\text{mean}})$, the population quantile $\mathbb{E}(\mathcal{Q}_\tau)$, and the percentage of infeasible cases (IF) among the total 1000 simulations.

Error	n	τ	q	$\mathcal{M}_{\text{mean}}$	$\mathcal{Q}_{10\%}$	$\mathcal{Q}_{25\%}$	Dual	$\mathbb{E}(\mathcal{M}_{\text{mean}})$	$\mathbb{E}(\mathcal{Q}_{10\%})$	$\mathbb{E}(\mathcal{Q}_{25\%})$	IF
$N(0,1)$	1000	/	/	3.41	−0.80	1.37	/	3.18	−0.71	1.41	/
		0.10	0.00	3.29	0.41	1.88	0.15	3.12	0.16	1.86	0
				(0.17)	(0.33)	(0.21)	(0.09)	(0.11)	(0.41)	(0.16)	/
			0.20	3.14	0.60	1.98	0.34	3.06	0.42	1.93	0
				(0.18)	(0.25)	(0.19)	(0.14)	(0.11)	(0.32)	(0.08)	/
		0.25	1.60	3.14	0.36	1.73	0.19	3.18	0.31	1.84	3.2%
				(0.26)	(0.52)	(0.21)	(0.11)	(0.15)	(0.32)	(0.16)	/
			1.70	3.09	0.49	1.86	0.22	3.09	0.40	1.83	16.3%
				(0.33)	(0.26)	(0.18)	(0.16)	(0.12)	(0.47)	(0.15)	/
	2000	0.10	0.00	3.15	0.28	1.82	0.20	3.16	0.19	1.91	0
				(0.14)	(0.24)	(0.16)	(0.18)	(0.04)	(0.30)	(0.14)	/
			0.20	3.14	0.35	1.88	0.15	3.18	0.27	1.86	0
				(0.12)	(0.16)	(0.08)	(0.11)	(0.04)	(0.27)	(0.07)	/
		0.25	1.60	3.20	0.13	1.81	0.18	3.16	0.07	1.79	8.3%
				(0.27)	(0.21)	(0.15)	(0.18)	(0.11)	(0.61)	(0.13)	/
			1.70	3.12	0.34	1.83	0.13	3.18	0.25	1.89	9.0%
				(0.10)	(0.29)	(0.12)	(0.22)	(0.16)	(0.24)	(0.11)	/
χ^2_5	1000	/	/	3.19	−3.21	−0.78	/	3.24	−2.80	−0.43	/
				(0.38)	(0.61)	(0.57)	/	(0.16)	(1.05)	(0.77)	/
		0.10	0.00	2.78	0.25	1.23	0.19	2.91	0.21	1.36	0
				(0.26)	(0.16)	(0.11)	(0.26)	(0.13)	(0.26)	(0.03)	/
			0.20	2.69	0.42	1.39	0.18	2.81	0.36	1.43	1.3%
				(0.24)	(0.13)	(0.09)	(0.38)	(0.15)	(0.12)	(0.03)	/
			1.00	3.03	0.05	1.17	0.27	2.74	−0.08	1.18	3.8%
				(0.45)	(0.07)	(0.16)	(0.27)	(0.32)	NA		
			1.10	2.87	0.14	1.23	0.20	2.94	0.01	1.34	21.2%
				(0.43)	(0.41)	(0.19)	(0.16)	(0.14)	(0.38)	(0.14)	/
	2000	0.10	0.00	2.79	0.25	1.30	0.11	2.93	0.18	1.41	0
				(0.20)	(0.18)	(0.13)	(0.10)	(0.09)	(0.31)	(0.02)	/
			0.20	2.70	0.34	1.24	0.19	2.82	0.38	1.38	0
				(0.23)	(0.12)	(0.14)	(0.10)	(0.13)	(0.17)	(0.03)	/
		0.25	1.00	2.99	−0.18	1.30	0.14	−0.33	1.22	3.1%	
				(0.33)	(0.40)	(0.11)	(0.15)	(0.15)	(0.55)	(0.23)	/
			1.10	2.89	−0.04	1.26	0.23	2.98	−0.07	1.18	23.8%
				(0.24)	(0.56)	(0.07)	(0.23)	(0.13)	(0.49)	0.14	/

outcome is the CD4 count at 96±5 weeks from baseline (denoted as CD496) as it is a crucial measure of the progression for HIV-infected patients.

We consider two covariates for estimating the treatment regimes, which are the baseline weights of the patients, and the baseline CD4-T cell counts. We then estimate the M-ITR, Q-ITR (maximizing the 0.25-th quantile), and the F-ITR (under the constraint that

Table 5: Estimated quantiles and means of different treatment regimes for ACTG175 data analysis.

Method	$\widehat{Q}_{.25}$	$\widehat{M}$
0.25-th Quantile-optimal	263.0	346.5
Mean-optimal	219.3	403.9
F-ITR	232.4	398.7

the 0.25-th quantile is lower bounded by 230). (It has been observed that when CD4 is below 200 cells/mm^3, the risk of serious health problems increases. For example, the risk of PCP (fungal pneumonia) and chest infections rise steeply when the CD4 falls below 200 cells/mm^3). The results are summarized in Table 5. We observe that the 0.25-quantiles and the means are significantly different for the Q-ITR and the M-ITR. Meanwhile, with a quantile constraint, the estimated mean of the F-ITR is close to the mean of M-ITR.

7 Discussion

To conclude, we propose a new framework for fairness-aware optimal ITR estimation under a quantile constraint. We show that the proposed estimator satisfies the quantile constraint, and achieves the optimal mean treatment effects asymptotically. Our extensive simulation studies demonstrate that though the estimator is derived from a highly nonconvex problem, our proposed algorithm achieves high-quality solutions in practice.

In practice, it is important to properly choose the quantile level τ and the threshold q in our proposed model (2). In one of the motivating examples, we aim to control the tail behavior of the treatment results. Thus, in such applications, we suggest that we let τ be 0.05 or 0.10, and q be 0 or a small positive number, where we assume that a positive result means that the patient benefits from the treatment. In future work, we will discuss with practitioners and make better recommendations.

Unlike the unconstrained mean-optimal ITR approach, the proposed method does not achieve the Fisher consistency in general even if the decision space increases. We argue here that by imposing a *practically meaningful* quantile constraint, the Fisher's consistency holds asymptotically. In particular, for the unconstrained approach, as the functional space of the

decision rule f increases, it is reasonable to assume that in expectation, the treatment effects are positive for all individuals, i.e., $\mathbb{E}\{Y_i^*(f)|x_i\} \geqslant 0$ for all x_i. Meanwhile, as mentioned above, the main motivation of imposing the quantile/fairness constraint is to ensure vast majority of the patients would benefit from the treatment. That is, some lower tail of the distribution of treatment effects is positive. Then, if we impose a constraint that

$$\mathcal{Q}_{0.05}\{Y^*(f)\} \geqslant 0,$$

this constraint is satisfied by the optimal solution to the unconstrained ITR problem by our assumption. Thus, the two solutions coincide, and the Fisher's consistency is satisfied.

For future work, our proposed method can be potentially generalized to achieve *group fairness*. In particular, suppose that we have a small number of K groups of patients. The groups can be defined by gender, age, or income status. We can then require the quantile constraint to be satisfied for each group by imposing multiple constraints. For example, suppose that we have two groups of patients. Denote the two groups as $\mathcal{G}_1$ and $\mathcal{G}_2$. With a slight abuse of notation, to ensure that the two groups get fair results, we may impose the constraints

$$\mathcal{Q}_\tau(Y_i^*(f)) \geqslant q \text{ for } i \in \mathcal{G}_k,\ k = 1, 2.$$

In this case, we have multiple constraints, and by a similar Lagrangian dual approach as we propose to tackle our original problem with a single constraint, we can potentially solve the problem. We will study this problem from both algorithmic and statistical perspectives in the future.

Acknowledgement

The authors sincerely thank the Editor, Associate Editor, and two anonymous reviewers for their invaluable comments, which lead to a significant improvement of this paper. Ethan X. Fang was partially supported by NSF Grants DMS-1820702, DMS-1953196, and DMS-2015539. Zhaoran Wang was partially supported by NSF Grants ECCS-2048075, CCF-2008827, DMS-2015568, and CCF-1934931, Simons Institute (Theory of Reinforcement Learning), and gifts from Amazon, Two Sigma and J.P. Morgan. Lan Wang was partially

supported by NSF Grants DMS-1952373 and OAC-1940160.

References

Behncke, S., Frölich, M. and Lechner, M. (2009). Targeting labour market programmes'results from a randomized experiment. *Swiss Journal of Economics and Statistics*, **145** 221–268.

Bhattacharya, D. (2009). Inferring optimal peer assignment from experimental data. *Journal of the American Statistical Association*, **104** 486–500.

Bhattacharya, D. and Dupas, P. (2012). Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics*, **167** 168–196.

Cai, T., Tian, L., Wong, P. H. and Wei, L. (2010). Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics*, **12** 270–282.

Chakraborty, B. and Moodie, E. (2013). *Statistical Methods for Dynamic Treatment Regimes*. Springer.

Chakraborty, B., Murphy, S. and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, **19** 317–343.

Díaz, I. and Hejazi, N. S. (2020). Causal mediation analysis for stochastic interventions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Díaz, I. and van der Laan, M. J. (2018). Stochastic treatment regimes. *Targeted Learning in Data Science,* 219–232.

Frölich, M. (2008). Statistical treatment choice: an application to active labor market programs. *Journal of the American Statistical Association*, **103** 547–558.

Gerber, A. S. and Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, **94** 653–663.

Goldberg, Y. and Kosorok, M. R. (2012). Q-learning with censored data. *Annals of Statistics*, **40** 529.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M. et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, **335** 1081–1090.

Henderson, R., Ansell, P. and Alshibani, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics*, **66** 1192–1201.

Hirano, K. and Porter, J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica*, **77** 1683–1701.

Huang, X., Choi, S., Wang, L. and Thall, P. F. (2015). Optimization of multi-stage dynamic treatment regimes utilizing accumulated data. *Statistics in Medicine*, **34** 3424–3443.

Imai, K., Ratkovic, M. et al. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, **7** 443–470.

Kennedy, E. H. (2019). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, **114** 645–656.

Kosorok, M. R. and Moodie, E. E. (2015). *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*, vol. 21. SIAM.

Laber, E. B., Lizotte, D. J., Qian, M., Pelham, W. E. and Murphy, S. A. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electronic Journal of Statistics*, **8** 1225–1272.

Lavori, P. W. and Dawson, R. (2000). A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society: Series A*, **163** 29–38.

Linn, K. A., Laber, E. B. and Stefanski, L. A. (2017). Interactive Q-learning for quantiles. *Journal of the American Statistical Association*, **112** 638–649.

Luckett, D. J., Laber, E. B. and Kosorok, M. R. (2017). Estimation and optimization of composite outcomes. *arXiv preprint arXiv:1711.10581*.

Luedtke, A. R. and van der Laan, M. J. (2016). Optimal individualized treatments in resource-limited settings. *The international journal of biostatistics*, **12** 283–303.

Mo, W., Qi, Z. and Liu, Y. (2020). Learning optimal distributionally robust individualized treatment rules. *Journal of the American Statistical Association* 1–16.

Moodie, E. E., Platt, R. W. and Kramer, M. S. (2009). Estimating response-maximized decision rules with applications to breastfeeding. *Journal of the American Statistical Association*, **104** 155–165.

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

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

Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, **24** 1455–1481.

Murphy, S. A. (2008). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, **24** 1455–1481.

Orellana, L., Rotnitzky, A. and Robins, J. M. (2010a). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: main content. *The International Journal of Biostatistics*, **6**.

Orellana, L., Rotnitzky, A. and Robins, J. M. (2010b). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part II: proofs of results. *The International Journal of Biostatistics*, **6**.

Qi, Z., Cui, Y., Liu, Y. and Pang, J.-S. (2019a). Estimation of individualized decision rules based on an optimized covariate-dependent equivalent of random outcomes. *SIAM Journal on Optimization*, **29** 2337–2362.

Qi, Z., Pang, J.-S. and Liu, Y. (2019b). Estimating individualized decision rules with tail controls. *arXiv preprint arXiv:1903.04367*.

Qiu, H., Carone, M., Sadikova, E., Petukhova, M., Kessler, R. C. and Luedtke, A. (2020). Optimal individualized decision rules using instrumental variable methods. *Journal of the American Statistical Association*.

Robins, J., Orellana, L. and Rotnitzky, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine*, **27** 4678–4721.

Robins, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality*. Springer, 69–117.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*. Springer, 189–326.

Robins, J. M., Hernan, M. A. and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics* 34–58.

Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, **81** 961–962.

Song, R., Wang, W., Zeng, D. and Kosorok, M. R. (2015). Penalized Q-learning for dynamic treatment regimens. *Statistica Sinica*, **25** 901.

Splawa-Neyman, J., Dabrowska, D. M. and Speed, T. (1990). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science* 465–472.

Staghoj, J., Svarer, M. and Rosholm, M. (2010). Choosing the best training programme: Is there a case for statistical treatment rules? *Oxford Bulletin of Economics and Statistics*, **72** 172–201.

Steinwart, I., Scovel, C. et al. (2007). Fast rates for support vector machines using gaussian kernels. *The Annals of Statistics*, **35** 575–607.

Tao, Y. and Wang, L. (2017). Adaptive contrast weighted learning for multi-stage multi-treatment decision-making. *Biometrics*, **73** 145–155.

Tetenov, A. (2012). Statistical treatment choice based on asymmetric minimax regret criteria. *Journal of Econometrics*, **166** 157–165.

Thall, P. F., Sung, H.-G. and Estey, E. H. (2002). Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association*, **97** 29–39.

Wang, L., Zhou, Y., Song, R. and Sherwood, B. (2018a). Quantile-optimal treatment regimes. *Journal of the American Statistical Association* 1–12.

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

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

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, **4** 65–85.

Wunsch, C. (2013). Optimal use of labor market policies: the role of job search assistance. *Review of Economics and Statistics*, **95** 1030–1045.

Xiao, W., Zhang, H. H. and Lu, W. (2019). Robust regression for optimal individualized treatment rules. *Statistics in Medicine*, **38** 2059–2073.

Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, **68** 1010–1018.

Zhao, Y., Zeng, D., Rush, A. J. and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, **107** 1106–1118.

Zhao, Y.-Q., Zeng, D., Laber, E. B., Song, R., Yuan, M. and Kosorok, M. R. (2014). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, **102** 151–168.

Appendix

A Proof of Proposition 1

Proof. First, we show that an optimal λ^* for problem (7) exits. By the definition of the Lagrangian dual function, we have that $\mathcal{L}(\lambda)$ in (7) is an infimum of a collection of linear functions. Thus, it holds that $\mathcal{L}(\lambda)$ is a convex function. Also, it is not difficult to see that as $\lambda \to +\infty$, we have $\mathcal{L}(\lambda) \to +\infty$. Thus, together with the convexity of $\mathcal{L}(\lambda)$, we have that $\mathcal{L}(\lambda)$ has compact level sets. That is, for any $\alpha \in \mathbb{R}$, the set $\{\lambda : \mathcal{L}(\lambda) \leqslant \alpha\}$ is compact. By the Bolzano-Weistrass Theorem, there exists an optimal Lagrangian multiplier λ^* that minimizes $\mathcal{L}(\lambda)$.

Then, given the optimal Lagrangian multiplier λ^*. Since, by assumption, the primal solution $\widehat{\boldsymbol{\beta}}$ exists, we have that the function $\widehat{\mathcal{M}}(\boldsymbol{\beta})$ is bounded above. We have that a dual optimal solution $\widetilde{\boldsymbol{\beta}}$ exists. ■

B Proof of Theorem 2

Proof. In the proof, for ease of presentation, we let the constraint for the primal problem be $\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}) \geqslant q$.

We consider the case where the optimal Lagrangian multiplier $\lambda^* = 0$ or $\lambda^* > 0$. We first have that, if $\lambda^* = 0$, we have that $\boldsymbol{\beta}^* = \operatorname{argmax} \widehat{\mathcal{M}}(\boldsymbol{\beta})$ by (8). Since $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) \geqslant q$ by the feasibility of $\widetilde{\boldsymbol{\beta}}$, we have $\widetilde{\boldsymbol{\beta}}$ is also a primal optimal solution, and our clam holds.

If $\lambda^* > 0$, we show in Lemma 12 that one of the two cases hold

i. There exists a dual optimal solution such that $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) = q$.

ii. There exist at least two solutions achieve the dual optimal objective, denoted as $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}'$, such that $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) < q$ and $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}') > q$.

Considering the two cases separately, for case (i), there exists a dual optimal solution $\widetilde{\boldsymbol{\beta}}$ such

that $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) = q$. By the weak duality, we have

$$\widehat{\mathcal{M}}(\widehat{\boldsymbol{\beta}}) \geqslant \widehat{\mathcal{M}}(\widetilde{\boldsymbol{\beta}}) + \lambda^*\{q - \widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}})\} = \widehat{\mathcal{M}}(\widetilde{\boldsymbol{\beta}}),$$

and our claim holds as desired. Note that in this case, the dual optimal solution actually also achieves the primal optimality.

We then focus on case (ii). Given the multiplier λ^*, there exist multiple solutions achieve the dual optimality. Suppose that there are m of them. Let these solutions be $\boldsymbol{\beta}_{(1)}$,...,$\boldsymbol{\beta}_{(m)}$ be the sequence of solutions ranked by their corresponding primal objective values that

$$\widehat{\mathcal{M}}(\boldsymbol{\beta}_{(1)}) \leqslant \widehat{\mathcal{M}}(\boldsymbol{\beta}_{(2)}) \leqslant \cdots \leqslant \widehat{\mathcal{M}}(\boldsymbol{\beta}_{(m)}).$$

Meanwhile, by the dual optimality, we have that

$$\lambda^*\widehat{\mathcal{M}}(\boldsymbol{\beta}_{(1)}) + \lambda^*\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(1)}) = \widehat{\mathcal{M}}(\boldsymbol{\beta}_{(2)}) + \lambda^*\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(2)}) = \cdots = \widehat{\mathcal{M}}(\boldsymbol{\beta}_{(m)}) + \lambda^*\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(m)}).$$

Since $\lambda^* > 0$, we have

$$\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(1)}) \geqslant \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(2)}) \geqslant \cdots \geqslant \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(m)}).$$

Meanwhile, by our assumption, we have that there exists some $k \in [m]$ such that

$$\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(k)}) \geqslant q \geqslant \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(k+1)}).$$

This shows that there exists a dual solution, $\boldsymbol{\beta}_{(k+1)}$ in this case, that satisfies the primal constraint, and the duality gap is upper bounded by $\widetilde{\lambda}(\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(k+1)} - q)$. Note that by the discrete nature of the sample quantile function $\widehat{\mathcal{Q}}_\tau(\cdot)$, the primal solution's corresponding sample quantile value is $\widehat{\mathcal{Q}}_\tau(\widehat{\boldsymbol{\beta}})$, which might be different from q. We thus have, the duality bound can be bounded by $\widetilde{\lambda}\big\{\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}_{(k+1)}) - \widehat{\mathcal{Q}}_\tau(\widehat{\boldsymbol{\beta}})\big\}$, which concludes our proof. ■

Lemma 12. *For the dual problem (8), suppose that the optimal Lagrangian multiplier $\lambda^* > 0$. One of the following two cases must hold that*

i. There exists a dual optimal solution such that $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) = q$.

ii. There exist at least two solutions achieve the dual optimal objective, denoted as $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}'$, such that $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) < q$ and $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}') > q$.

Proof. We prove the lemma by contradiction. We assume the contrary that $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) < q$ for all dual optimal solutions $\widetilde{\boldsymbol{\beta}}$ that achieve the dual optimal objective. (Note that the other case $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) < q$ follows by similar arguments.) We have that

$$\begin{aligned}\mathcal{L}(\lambda^*) &= \underset{\boldsymbol{\beta}}{\text{maximize}}\ \widehat{\mathcal{M}}(\boldsymbol{\beta}) + \lambda^*\{q - \widehat{\mathcal{Q}}_\lambda(\boldsymbol{\beta})\} \\ &= \underset{\ell\in[n]}{\text{maximize}}\ \widehat{\mathcal{M}}(\boldsymbol{\beta}^{(\ell)}) + \lambda^*\{q - \widehat{\mathcal{Q}}_\lambda(\boldsymbol{\beta}^{(\ell)})\},\end{aligned}$$

where $\boldsymbol{\beta}^{(\ell)} = \text{argmax}_{\boldsymbol{\beta}:\mathcal{Q}_\tau(\boldsymbol{\beta})=y_\ell}\, \widehat{\mathcal{M}}(\boldsymbol{\beta})(\boldsymbol{\beta})$, by the fact that $\mathcal{Q}_\tau(\boldsymbol{\beta}) = y_i$ for some $i \in [n]$.

By our assumption that $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}})$ is strictly less than q. As shown in Lemma 13, we have that for small $\varepsilon > 0$, we have

$$\begin{aligned}\mathcal{L}(\lambda^* + \varepsilon) &= \underset{\ell\in[n]}{\text{maximize}}\ \widehat{\mathcal{M}}(\boldsymbol{\beta}) + (\lambda^* + \varepsilon)\{q - \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta})\} \\ &< \underset{\ell\in[n]}{\text{maximize}}\ \widehat{\mathcal{M}}(\boldsymbol{\beta}) + \lambda^*\{q - \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta})\} \\ &= \mathcal{L}(\lambda^*).\end{aligned}$$

However, since λ^* is the optimal Lagrangian multiplier by our assumption, it minimizes the function $\mathcal{L}(\lambda^*)$. The above result gives a contradiction, and our result holds as desired. ■

Lemma 13. *Suppose that the dual optimal Lagrangian multiplier $\lambda^* > 0$. Let $\boldsymbol{\beta}^{(\ell)} = \text{argmax}\{\widehat{\mathcal{M}}(\boldsymbol{\beta}) : \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}) = y_\ell\}$ for $\ell = 1, ..., n$. With loss of generality, assume $y_1 < y_2 < \cdots < y_n$. It holds that if $0 < \varepsilon < \min_{\ell\in\{2,...,n\}} \{\mathcal{M}(\boldsymbol{\beta}^{(\ell-1)}) - \mathcal{M}(\boldsymbol{\beta}^{(\ell)}\}$,*

$$\mathcal{L}(\lambda^* + \varepsilon) = \underset{\ell\in[n]}{minimize}\widehat{\mathcal{M}}(\boldsymbol{\beta}) + (\lambda^* + \varepsilon)\{q - \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta})\}.$$

Proof. When we perturb the optimal λ^* to $\lambda^* + \varepsilon$, the corresponding dual solution becomes

$$\widetilde{\boldsymbol{\beta}}' = \underset{\boldsymbol{\beta}}{\text{argmax}}\, \widehat{\mathcal{M}}(\boldsymbol{\beta}) + (\lambda^* + \varepsilon)\{q - \widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta})\}.$$

By our choice of ε, it is not difficult to see that our claim holds as desired. ■

C Proof of Proposition 4

Proof. For ease of presentation, we denote by $\breve{M}(\breve{\boldsymbol{\beta}})$ and $\breve{Q}_\tau(\breve{\boldsymbol{\beta}})$ the sample mean and τ-th quantile of the treatment effects following deterministic decision rule $f_i = \mathbb{1}(\mathbf{x}_i^\top \breve{\boldsymbol{\beta}} > 0)$.

We first prove that there exists a $\widetilde{\boldsymbol{\beta}}$ that if we follow the stochastic ITR $f(\mathbf{x}_i, \widetilde{\boldsymbol{\beta}}) = \mathbb{P}(f_i = 1) = \{1 + \exp(-\mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}})\}^{-1}$, the corresponding objective can be arbitrarily close to the objective achieved by the deterministic ITR $f_i = \mathbb{1}(\mathbf{x}_i^\top \breve{\boldsymbol{\beta}} > 0)$, and the quantile constraint is approximately satisfied by the stochastic ITR. We have that for any $\breve{\boldsymbol{\beta}}$ and $\delta > 0$, there exists some $\widetilde{\boldsymbol{\beta}}$ such that $\left|\mathbb{1}(\mathbf{x}_i^\top \breve{\boldsymbol{\beta}} > 0) - \left\{1 + \exp(-\mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}})\right\}^{-1}\right| \leqslant \delta$ for all $\mathbf{x}_i$. (Note that here we implicitly assume that $\mathbf{x}_i \neq 0$. If we indeed have some $\mathbf{x}_i = \mathbf{0}$, we may perturb the data by letting all $\mathbf{x}_i' = \mathbf{x}_i + \delta$ for some δ such that all $\mathbf{x}_i' \neq 0$.) This implies that by considering stochastic ITRs that $f(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbb{P}(f_i = 1) = \{1 + \exp(-\mathbf{x}_i^\top \boldsymbol{\beta})\}^{-1}$, we have for any given $\varepsilon_1 > 0$, there exists some $\widetilde{\boldsymbol{\beta}}$, such that the corresponding objective satisfies $\widehat{\mathcal{M}}(\widetilde{\boldsymbol{\beta}}) > \breve{\mathcal{M}}(\breve{\boldsymbol{\beta}}) - \varepsilon_1$ and the corresponding quantile constraint satisfies $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) > \breve{\mathcal{Q}}_\tau(\breve{\boldsymbol{\beta}}) - \varepsilon_1$.

In addition, we have that by our assumptions that all outcomes are bounded, and $\breve{\boldsymbol{\beta}}$ achieves the quantile constraint in population. Also, as shown above, for any $\varepsilon_1 > 0$, there exists some $\widetilde{\boldsymbol{\beta}}$ such that $\widehat{\mathcal{Q}}_\tau(\widetilde{\boldsymbol{\beta}}) > \breve{\mathcal{Q}}_\tau(\breve{\boldsymbol{\beta}}) - \varepsilon_1$. We thus have that if n is large enough, problem (6) is feasible. Note that as $n \to \infty$ both $\widehat{\mathcal{M}}(\boldsymbol{\beta})$ and $\breve{\mathcal{M}}(\boldsymbol{\beta})$ converge to $\mathcal{M}(\boldsymbol{\beta}) = E(Y^*(\boldsymbol{\beta}))$. Meanwhile, we have $\widehat{\mathcal{M}}(\widetilde{\boldsymbol{\beta}}) > \breve{\mathcal{M}}(\breve{\boldsymbol{\beta}}) - \varepsilon_1$. We then have for any $\varepsilon_2 > 0$, the solution to our problem (6), $\widehat{\boldsymbol{\beta}}$, satisfies that $\mathcal{M}(\widehat{\boldsymbol{\beta}}) \geqslant \mathcal{M}(\breve{\boldsymbol{\beta}}) - \varepsilon_1 - \varepsilon_2$ with probability approaching one, and satisfies the quantile constraint in (6). Since ε_1 and ε_2 are arbitrary, our claim follows as desired. ■

D Proof of Theorem 5

Proof. Denote by $\mathbb{P}_n$ the empirical measure of the observed samples. Let $\boldsymbol{\beta}^*$ be a minimizer to the loss function under the quantile constraint in expectation that

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta} \in \mathcal{B}}{\operatorname{argmax}}\, \mathcal{M}(\boldsymbol{\beta}), \text{ subject to } \mathcal{Q}_\tau(\boldsymbol{\beta}) \geqslant q.$$

First, we have that as $\mathcal{Q}_\tau(\boldsymbol{\beta}^*) \geqslant q$, by Theorem 1 of Wang et al. (2018a), it is not difficult to see that, as n increases, $\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}^*) \geqslant q - C \cdot n^{-1/2}$ for some constant C with probability goes to 1. Thus, we have that as n increases, $\boldsymbol{\beta}^*$ is a feasible point for problem (6) with probability goes to 1.

Meanwhile, by the definition that $\widehat{\boldsymbol{\beta}}$ is the maximizer for the empirical mean function under the constraint, we have that for any n large enough, $\widehat{\mathcal{M}}(\widehat{\boldsymbol{\beta}}) \geqslant \widehat{\mathcal{M}}(\boldsymbol{\beta}^*)$ for all $\boldsymbol{\beta}^* \in \mathcal{B}$ and satisfies $\widehat{\mathcal{Q}}_\tau(\boldsymbol{\beta}^*) \geqslant q - C \cdot n^{-1/2}$ with high probability. Thus, we only need to prove that $\widehat{\mathcal{M}}(\widehat{\boldsymbol{\beta}}) \to \mathcal{M}(\widehat{\boldsymbol{\beta}})$ in probability.

By our assumption that $\boldsymbol{\beta} \in \mathcal{B}$, and $\mathcal{B}$ is compact, we have that $\widehat{\boldsymbol{\beta}}$ is bounded. This implies that $\{\widehat{\mathcal{M}}(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \mathcal{B}\}$ belongs to a Donsker class because it is not difficult to see $\widehat{\mathcal{M}}(\boldsymbol{\beta})$ is Lipschitz continuous with respect to $\boldsymbol{\beta}$. Consequently, we have

$$\sqrt{n}\{\widehat{\mathcal{M}}(\widehat{\boldsymbol{\beta}}) - \mathcal{M}(\widehat{\boldsymbol{\beta}})\} = \mathcal{O}_P(1).$$

Our claim holds as desired. ■

E Proof of Theorem 6

Proof. The proof is based on an application of Theorem 5.6 of Steinwart et al. (2007). Specifically, let $\mathcal{G}$ be the function class

$$\mathcal{G} = \{\widehat{\mathcal{M}}(\boldsymbol{\beta}) - \widehat{\mathcal{M}}(\boldsymbol{\beta}^*) : \boldsymbol{\beta} \in \mathcal{Q}_\tau(q)\},$$

where $\boldsymbol{\beta}^* \in \operatorname{argmax}_{\boldsymbol{\beta} \in \mathcal{Q}_\tau(q)} \mathcal{M}(\boldsymbol{\beta})$, and $\mathcal{Q}_\tau(q) = \{\boldsymbol{\beta} : \mathcal{Q}_\tau(\boldsymbol{\beta}^*) \geqslant q\}$. We first have that $\mathbb{E}(g) \leqslant 0$ for any $g \in \mathcal{G}$ as $\boldsymbol{\beta}^*$ is a maximizer in expectation. Note that our loss function is Lipschitz conitnuous with respect to $\boldsymbol{\beta}$. Denote that Lipschitz constant as C_L, we have $|g| \leqslant C_L \|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|$. As we assume that $\boldsymbol{\beta} \in \mathcal{B}(M)$, we have $|g| \leqslant B = 2MC_L$. Consequently, squaring both sides and taking expectations, we have $\mathbb{E}(g^2) \leqslant \mathbb{E}(g) + 4B^2$.

Next, for the covering number $N\big(B^{-1}\mathcal{G}, \varepsilon, L_2(\mathbb{P}_n)\big)$, we have

$$\begin{aligned}\log N\big(B^{-1}\mathcal{G}, \varepsilon, L_2(\mathbb{P}_n)\big) &\leqslant \log N\big(B^{-1}\{\widehat{\mathcal{M}}(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \mathcal{B}(M)\}, \varepsilon, L_2(\mathbb{P}_n)\big) \\ &\leqslant \log N\big(\mathcal{B}(M), B\varepsilon/C_L, , L_2(\mathbb{P}_n)\big) \\ &\leqslant \log N\big(\mathcal{B}(1), 2\varepsilon, L_2(\mathbb{P}_n)\big).\end{aligned}$$

Thus, by Theorem 2.1 of Steinwart et al. (2007), we have that for some constant C,

$$\sup_{\mathbb{P}_n} \log N\big(B^{-1}\mathcal{G}, \varepsilon, L_2(\mathbb{P}_n)\big) \leqslant C\varepsilon^{-2}.$$

Consequently, by Theorem 5.6 of Steinwart et al. (2007), there exists a constant C_S such that for all $n \geqslant 1$ and $\tau \geqslant 1$, we have that

$$\mathbb{P}^*\big(\mathcal{M}(\widehat{\boldsymbol{\beta}}) < \mathcal{M}(\boldsymbol{\beta}^*) - C_S\varepsilon(n, C_1, B, \tau)\big) \leqslant e^{-\tau},$$

where

$$\varepsilon(n, C_1, B, \tau) = B \cdot \Big(\frac{1}{n} + \frac{4}{\sqrt{n}}\Big) + (B + C_1)\frac{\tau}{n}.$$

Our claim holds as desired. ■

F Dynamic Treatment Regime with Intermediate Outcome

In this section, we extend the dynamic treatment regime discussed in Section 5 to the more general case where we observe intermediate outcome at each stage. Similar to Section 5, we consider 2-stage dynamic treatment regime for ease of presentation, and the methods and results for the general T-stage case can be easily generalized. We also assume that the data are from some SMART trial.

The main difference between the setup with intermediate outcome is that after the first stage, we observe an intermediate outcome $Y_i^{(1)}$ for sample i, and after the second stage, we observe an outcome $Y_i^{(2)}$. Let $\boldsymbol{H}_i^{(1)} = \boldsymbol{X}_i^{(1)}$ and $\boldsymbol{H}_i^{(2)} = (\boldsymbol{X}_i^{(1)\top}, A_i^{(1)}, Y_i^{(1)}, \boldsymbol{X}_i^{(2)\top})^\top$. We

consider candidate stochastic F-ITR indexed by $\boldsymbol{\beta} = \{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}\}$ such that $f_j(\boldsymbol{H}_i^{(j)}, \boldsymbol{\beta}^{(j)}) = \mathbb{P}(A_i^{(j)} = 1 | \boldsymbol{H}_i^{(j)}) = \{1 + \exp(-\boldsymbol{H}_i^{(j)\top} \boldsymbol{\beta}^{(j)})\}$ for $j = 1, 2$.

Suppose we have random samples $\{\mathbf{x}_i^{(1)}, a_i^{(1)}, y_i^{(1)}, \mathbf{x}_i^{(2)}, a_i^{(2)}, y_i^{(2)}\}_{i \in [n]}$, and we let $\mathbf{h}_i^{(1)} = \mathbf{x}_i^{(1)}$, and $\mathbf{h}_i^{(2)} = (\mathbf{x}_i^{(1)\top}, a_i^{(1)}, y_i^{(1)}, \mathbf{x}_i^{(2)\top})^\top$. We consider a backward fitting approach to estimating the optimal F-ITR. Specifically, letting $c_i^{(2)}(\boldsymbol{\beta}^{(2)}) = a_i^{(2)} f^{(2)}(\mathbf{h}_i^{(2)}, \boldsymbol{\beta}^{(2)}) + (1 - a_i^{(2)})\{1 - f^{(2)}(\mathbf{h}_i^{(2)}, \boldsymbol{\beta}^{(2)})\}$, we estimate the regime for stage 2 by

$$\widehat{\boldsymbol{\beta}}^{(2)} \in \operatorname{argmax} \widehat{\mathcal{M}}^{(2)}(\boldsymbol{\beta}^{(2)}), \text{ subject to } \widehat{\mathcal{Q}}_{\tau_2}^{(2)}(\boldsymbol{\beta}^{(2)}) \geqslant q - C_2/\sqrt{n}, \tag{13}$$

where $\widehat{\mathcal{M}}^{(2)}(\boldsymbol{\beta}^{(2)})$ and $\widehat{\mathcal{Q}}_{\tau_2}^{(2)}(\boldsymbol{\beta}^{(2)})$ are the estimators for the mean and τ_2-th quantile of outcome in stage 2 that

$$\widehat{\mathcal{M}}^{(2)}(\boldsymbol{\beta}^{(2)}) = \operatorname{argmin}_\mu n^{-1} \sum_{i=1}^n c_i^{(2)}(\boldsymbol{\beta}^{(2)})(y_i^{(2)} - \mu)^2,$$

and

$$\widehat{\mathcal{Q}}^{(2)}(\boldsymbol{\beta}^{(2)}) = \operatorname{argmin}_q n^{-1} \sum_{i=1}^n c_i^{(2)}(\boldsymbol{\beta}^{(2)}) \rho_{\tau_2}(y_i^{(2)} - q),$$

and C_2 is a constant.

After getting $\widehat{\boldsymbol{\beta}}^{(2)}$, we estimate the regime for stage 1. First, similar to (11), we let

$$\begin{aligned} c_i^{(1)}(\boldsymbol{\beta}^{(1)}) =& \frac{a_i^{(1)} a_i^{(2)}}{\pi_1 \pi_2} \cdot f_i^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)}) f_i^{(2)}(\mathbf{h}_i^{(2)}, \widehat{\boldsymbol{\beta}}^{(2)}) \\ &+ \frac{a_i^{(1)}(1 - a_i^{(2)})}{\pi_1(1 - \pi_2)} \cdot f_i^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)}) \big(1 - f_i^{(2)}(\mathbf{h}_i^{(2)}, \widehat{\boldsymbol{\beta}}^{(2)})\big) \\ &+ \frac{(1 - a_i^{(1)}) a_i^{(2)}}{(1 - \pi_1)\pi_2} \cdot \big(1 - f_i^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)})\big) f_i^{(2)}(\mathbf{h}_i^{(2)}, \widehat{\boldsymbol{\beta}}^{(2)}) \\ &+ \frac{(1 - a_i^{(1)})(1 - a_i^{(2)})}{(1 - \pi_1)(1 - \pi_2)} \cdot \big(1 - f_i^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)})\big) \big(1 - f_i^{(2)}(\mathbf{h}_i^{(2)}, \widehat{\boldsymbol{\beta}}^{(2)})\big). \end{aligned}$$

Then, we estimate the F-ITR at stage 1 by

$$\widehat{\boldsymbol{\beta}}^{(1)} = \operatorname*{argmax}_{\boldsymbol{\beta}^{(1)}} \widehat{\mathcal{M}}(\boldsymbol{\beta}^{(1)}), \text{ subject to } \widehat{\mathcal{Q}}_{\tau_1}^{(1)}(\boldsymbol{\beta}^{(1)}) \geqslant q_1 - C_1/\sqrt{n}, \text{ and } \widehat{\mathcal{Q}}_{\tau}(\boldsymbol{\beta}^{(1)}) \geqslant q - C/\sqrt{n}, \tag{14}$$

where

$$\widehat{\mathcal{M}}(\boldsymbol{\beta}^{(1)}) = \text{argmin}_{\mu} n^{-1} \sum_{i=1}^{n} c_i^{(1)}(\boldsymbol{\beta}^{(1)})(y_i^{(1)} + y_i^{(2)} - \mu)^2$$

is the estimator of the mean of total outcome, and

$$\widehat{\mathcal{Q}}_{\tau}(\boldsymbol{\beta}^{(1)}) = \text{argmin}_{q} n^{-1} \sum_{i=1}^{n} c_i^{(1)}(\boldsymbol{\beta}^{(1)})\rho_{\tau}(y_i^{(1)} + y_i^{(2)} - q)$$

is the estimator for the τ-th quantile of the total outcome, and

$$\widehat{\mathcal{Q}}_{\tau_1}^{(1)}(\boldsymbol{\beta}^{(1)}) = \text{argmin}_{q} n^{-1} \sum_{i=1}^{n} c_i(\boldsymbol{\beta}^{(1)})\rho_{\tau_1}(y_i^{(1)} - q),$$

where $c_i(\boldsymbol{\beta}^{(1)}) = a_i^{(1)} f^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)}) + (1 - a_i^{(1)})\{1 - f^{(1)}(\mathbf{h}_i^{(1)}, \boldsymbol{\beta}^{(1)})\}$, is the estimator for the τ_1-th quantile of the stage 1 intermediate outcome.

For the estimator $\widehat{\boldsymbol{\beta}} = \{\widehat{\boldsymbol{\beta}}^{(1)}, \widehat{\boldsymbol{\beta}}^{(2)}\}$ derived above, we can get similar $\mathcal{O}_P(n^{-1/2})$ rate of convergence to the optimal risk $\mathcal{M}(\boldsymbol{\beta}^*)$, while satisfying the quantile constraints by backward induction and similar arguments in the proof of Theorem 6.

Theorem 14. *Suppose that $\boldsymbol{\beta}^* = \{\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*\}$ belongs to a compact set $\mathcal{B}(M)$, where $M > 0$ is a constant. Then we have that for all $\tau \geqslant 1$ we have*

$$\mathbb{P}^*\big(\mathcal{M}(\widehat{\boldsymbol{\beta}}) \geqslant \mathcal{M}(\boldsymbol{\beta}^*) - \varepsilon\big) \geqslant 1 - e^{-\tau},$$

where $\mathbb{P}^*$ denotes the outer probability for possibly nonmeasureable sets, and $\varepsilon = \mathcal{O}(n^{-1/2})$. Or, equivalently,

$$|\mathcal{M}(\widehat{\boldsymbol{\beta}}) - \mathcal{M}(\boldsymbol{\beta}^*)| = \mathcal{O}_P(n^{-1/2}).$$

In addition, we have that, with probability goes to 1,

$$\mathcal{Q}_{\tau}^{(1)}(\widehat{\boldsymbol{\beta}}^{(1)}) \geqslant q_1, \;\; \mathcal{Q}_{\tau}^{(2)}(\widehat{\boldsymbol{\beta}}^{(2)}) \geqslant q_2, \;\; \text{and } \mathcal{Q}_{\tau}(\widehat{\boldsymbol{\beta}}) \geqslant q,$$

where $\mathcal{Q}_{\tau_1}^{(1)}(\boldsymbol{\beta}^{(1)})$ denotes the τ_1-th quantile of the stage 1 intermediate outcome, $\mathcal{Q}_{\tau_2}^{(2)}(\boldsymbol{\beta}^{(2)})$ denotes the τ_2-th quantile of the stage 2 intermediate outcome, and $\mathcal{Q}_{\tau}(\widehat{\boldsymbol{\beta}})$ denotes the τ-th quantile of the total outcome.