

Research Article

Hongxiang Qiu*, Marco Carone, and Alex Luedtke

Individualized treatment rules under stochastic treatment cost constraints

<https://doi.org/10.1515/jci-2022-0005>

received January 17, 2022; accepted November 25, 2022

Abstract: Estimation and evaluation of individualized treatment rules have been studied extensively, but real-world treatment resource constraints have received limited attention in existing methods. We investigate a setting in which treatment is intervened upon based on covariates to optimize the mean counterfactual outcome under treatment cost constraints when the treatment cost is random. In a particularly interesting special case, an instrumental variable corresponding to encouragement to treatment is intervened upon with constraints on the proportion receiving treatment. For such settings, we first develop a method to estimate optimal individualized treatment rules. We further construct an asymptotically efficient plug-in estimator of the corresponding average treatment effect relative to a given reference rule.

Keywords: nonparametric inference, average treatment effect, dynamic treatment regime

MSC 2020: 62D20, 62G20

1 Introduction

The effect of a treatment often varies across subgroups of the population [1,2]. When such differences are clinically meaningful, it may be beneficial to assign treatments strategically depending on subgroup membership. Such treatment assignment mechanisms are called individualized treatment rules (ITRs). A treatment rule is commonly evaluated on the basis of the mean counterfactual outcome value it generates – what is often referred to as the treatment rule’s value – and an ITR with an optimal value is called an optimal ITR. There is an extensive literature on estimation of optimal ITRs and their corresponding values using data from randomized trials or observational studies [3–7].

Most existing approaches for estimating ITRs do not incorporate real-world resource constraints. Without such constraints, an optimal ITR would assign the treatment to members of a subgroup provided there is any benefit for such individuals, even when this benefit is minute. In contrast, under treatment resource limits, it may be more advantageous to reserve treatment for subgroups with the greatest benefit from treatment. This issue has received attention in the recent work. Luedtke and van der Laan developed methods for estimation and evaluation of optimal ITRs with a constraint on the proportion receiving treatment [8]. Qiu et al. instead considered related problems in settings in which instrumental variables (IVs) are available [9]. In one of the settings they considered, the same resource constraint is imposed as in the study by Luedtke and van der Laan [8], but a binary IV is used to identify optimal ITRs even in settings in which there may be unmeasured confounders. In another setting considered in the study by Qiu et al. [9],

* **Corresponding author: Hongxiang Qiu**, Department of Statistics, the Wharton School, University of Pennsylvania, Philadelphia, PA 19104, United States, e-mail: qiuhx@upenn.edu

Marco Carone: Department of Biostatistics, University of Washington, Seattle, Washington, United States, e-mail: mcarone@uw.edu

Alex Luedtke: Department of Statistics, University of Washington, Seattle, Washington, United States, e-mail: aluedtke@uw.edu

the authors considered interventions on a causal IV or encouragement status and developed methods to estimate individualized encouragement (rather than treatment) rules with a constraint on the proportion receiving both encouragement and treatment [10]. They also developed nonparametrically efficient estimators of the average causal effect of optimal rules relative to a prespecified reference rule. Sun et al. [11] considered a setting in which the cost of treatment is random and dependent on baseline covariates. They developed methods to estimate optimal ITRs under a constraint on the expected additional treatment cost compared to control, though inference on the impact of implementing the optimal ITR in the population was not studied [11]. Sun [12] considered a related problem involving the development of optimal ITRs under resource constraints and established the asymptotic properties of the estimated optimal ITR. Their method appears viable when the class of ITRs is restricted by the user *a priori*.

In this article, we study estimation and inference for an optimal rule under two different cost constraints. The first is the same as appearing in the study by Sun et al. [11]. In contrast to earlier work on this setting, we do not constrain the class of ITRs considered and provide a means to obtain inference about the optimal ITR. The second constraint we consider places a cap on the total cost under the rule rather than on the incremental cost relative to control. To our knowledge, the latter problem has not previously been considered in the literature. Both of these estimation problems mirror the intervention-on-encouragement setting considered in the study by Qiu et al. [9] but involve different constraints and a more general cost function.

Similarly as in the study by Qiu et al. [9], the estimators that we develop are asymptotically efficient within a nonparametric model and enable the construction of asymptotically valid confidence intervals (CIs) for the impact of implementing the optimal rule. We develop our estimators using similar tools – such as semiparametric efficiency theory [13,14] and targeted minimum loss-based estimation (TMLE) [15,16] – as were used to tackle the related problem studied in the study by Qiu et al. [9]. Consequently, our proposed estimators are similar to that presented in the study by Qiu et al. [9]. Therefore, we will streamline the presentation by highlighting the key similarities and focusing on the differences between these related problems and estimation schemes.

The rest of this article is organized as follows. In Section 2, we describe the problem setup, introduce notation, and present the causal estimands along with basic causal conditions. In Section 3, we present additional causal conditions and the corresponding nonparametric identification results. In Section 4, we present our proposed estimators and their theoretical properties. In Section 5, we present a simulation illustrating the performance of our proposed estimators. We make concluding remarks in Section 6. Proofs, technical conditions, and additional simulation results can be found in the Supplementary material.

2 Setup and objectives

To facilitate comparisons with the study by Qiu et al. [9], we adopt similar notation. Suppose that we observe independent and identically distributed data units $O_1, O_2, \dots, O_n \sim P_0$, where P_0 is an unknown sampling distribution. A prototypical data unit O consists of the quadruplet (W, T, C, Y) , where $W \in \mathcal{W} \subseteq \mathbb{R}^p$ is the vector of baseline covariates, $T \in \{0, 1\}$ is the treatment status, $C \in [0, \infty)$ is the random treatment cost, and $Y \in \mathbb{R}$ is the outcome of interest. As a convention, we assume that larger values of Y are preferable. We use $V = V(W) \in \mathcal{V}$ to denote a fixed transformation of W upon which we allow treatment decisions to depend. For example, V may be a subset of covariates in W or a summary of W (e.g., body mass index as a summary of height and weight). In practice, V may be chosen based on prior knowledge on potential modifiers of the treatment effect as well as the cost of measuring various covariates. We distinguish between $V(W)$ and W because of their different roles. On the one hand, we will assume that the full covariate W contains all confounders and thus is used to identify causal effects, while $V(W)$ might not be sufficient for this purpose. On the other hand, some covariates in W may be expensive or difficult to measure in future applications, and thus, implementing an optimal ITR based on a subset $V(W)$ of covariate W may be desirable. In the rest of this article, we will use the shorthand notation V , V_i , and v to refer to $V(W)$, $V(W_i)$, and $V(w)$, respectively. We define an individualized (stochastic) treatment

rule (ITR) to be a function $\rho : \mathcal{V} \rightarrow [0, 1]$ that prescribes treatment with probability $\rho(v)$ according to an exogenous source of randomness for an individual with covariate value v . Any stochastic ITR that only takes values in $\{0, 1\}$ is referred to as a deterministic ITR.

In this work, we adopt the potential outcomes framework [17,18]. For each individual, we use $C(t)$ and $Y(t)$ to denote the potential treatment cost and potential outcome, respectively, corresponding to scenarios in which the individual has treatment status t . We use E to denote an expectation over the counterfactual observations and the exogenous random mechanism defining a rule, and E_0 to denote an expectation over observables alone under sampling from P_0 . We make the usual stable unit treatment value assumption.

Condition A1. (Stable unit treatment value assumption) The counterfactual data unit of one individual is unaffected by the treatment assigned to other individuals, and there is only a single version of the treatment, so that $T = t$ implies that $C = C(t)$ and $Y = Y(t)$.

Remark 1. The ITRs we consider are not truly individualized, because they are based on the value of covariate V rather than each individual's unique potential treatment effects $Y(1) - Y(0)$ and $C(1) - C(0)$. Nevertheless, depending on the resolution of V , these ITRs can be considerably more individualized than assigning everyone to either treatment or control. In this article, we adopt the conventional nomenclature and refer to the treatment rules we study as ITRs [see, e.g., 7,19–29].

We define $C(\rho)$ and $Y(\rho)$ to be the counterfactual treatment cost and outcome, respectively, for an ITR ρ under an exogenous random mechanism. We note that if $\rho(v) \in (0, 1)$ for an individual with covariate v , an exogenous random mechanism is used to randomly assign treatment with probability $\rho(v)$, and thus, $C(\rho)$ and $Y(\rho)$ are random for this given individual. If ρ were implemented in the population, then the population mean outcome would be $E[Y(\rho)]$, where we use E to denote expectation under the true data-generating mechanism involving potential outcomes $(C(t), Y(t))$ and exogenous randomness in ρ . We consider a generic treatment resource constraint requiring that a convex combination of the population average treatment cost and the population average additional treatment cost compared to control be no greater than a specified constant $\kappa \in (0, \infty]$. Consequently, an optimal ITR ρ_0 under this constraint is a solution in $\rho : \mathcal{V} \rightarrow [0, 1]$ to

$$\text{maximize } E[Y(\rho)] \text{ subject to } \alpha E[C(\rho)] + (1 - \alpha)E[C(\rho) - C(0)] \leq \kappa, \quad (1)$$

where $\alpha \in [0, 1]$ is also a constant specified by the investigator. Natural choices of α are $\alpha = 0$, corresponding to a constraint on the population average additional treatment cost compared to control, and $\alpha = 1$, corresponding to a constraint on the population average treatment cost. The first choice may be preferred when the control treatment corresponds to the current standard of care and a limited budget is available to fund the novel treatment to some patients. The second choice may be more relevant when both treatment and control incur treatment costs.

Remark 2. Our setup is similar to that in the study by Qiu et al. [9] if we view T and C defined here as the IV/encouragement Z and treatment status A defined in those prior works, respectively. However, the constraint in our setup is different from the constraint $E[\rho(V)C(\rho)] \leq \kappa$ considered previously. In IV settings, the constraint in (1) with $\alpha = 1$ is useful when assigning treatment *always* incurs a cost, regardless of whether encouragement is applied, such as in distributing a limited supply of an expensive drug within a health system based on the results of a randomized clinical trial. It is instead useful with $\alpha = 0$ when no encouragement is present under the standard of care but intervention on the encouragement is of interest when additional treatment resources are available. The constraint considered in the study by Qiu et al. [9,10] was instead useful in cases in which treatment only incurs a cost when paired with encouragement, such as when housing vouchers are used to encourage individuals to live in a certain area. In the general setting in which T is viewed as treatment status and C as a random treatment cost, the constraint in (1) with $\alpha = 0$ is identical to that considered in the study by Sun et al. [11] – we refer the readers to these works for a more in-depth discussion of the relation between the current problem setup and IV settings.

To evaluate an optimal ITR ρ_0 , we follow Qiu et al. [9] in considering three types of reference ITRs and develop methods for statistical inference on the difference in the mean counterfactual outcome between ρ_0 and a reference ITR $\rho_0^R : \mathcal{V} \rightarrow [0, 1]$. The first type of reference ITR considered, denoted by ρ^{FR} (FR = fixed rule), is any fixed ITR that may be specified by the investigator before the study. When $\alpha = 0$, it is usually most reasonable to consider the rule that always assigns control, namely, $v \mapsto 0$, because the constraint in (1) may arise due to limited funding for implementing treatment whereas the standard of care rule is to always assign control. The second type, denoted by ρ_0^{RD} (RD = random), prescribes treatment completely at random to individuals regardless of their baseline covariates. The probability of prescribing treatment is chosen such that the treatment resource is saturated (i.e., all available resources are used) or all individuals receive treatment, if such a probability exists. Symbolically, this ITR is given by $\rho_0^{\text{RD}} : v \mapsto \min\{1, (\kappa - \alpha E[C(0)]) / E[C(1) - C(0)]\}$ under the condition that $E[C(0)] \leq \kappa$ and $E[C(1) - C(0)] > 0$. Although ρ_0^{RD} has the same interpretation as the corresponding encouragement rule in the study by Qiu et al. [9], its mathematical expression is different due to the different resource constraints. This rule may be of interest if it is known *a priori* that treatment is harmless. The third type, denoted by ρ_0^{TP} (TP = true propensity), prescribes treatment according to the true propensity of the treatment implied by the study sampling mechanism P_0 , so that ρ_0^{TP} equals $w \mapsto P_0(T = 1 | W = w)$. This ITR may be of interest in two settings. In one setting, ρ_0^{TP} satisfies the treatment resource constraint. The investigator may wish to determine the extent to which the implementation of an optimal ITR would improve upon the standard of care. In the other setting, the treatment resource constraint is newly introduced and the standard-of-care ITR may lead to overuse of treatment resources. The investigator may then be interested in whether the implementation of an optimal constrained ITR would result, despite the new resource constraint, in a non-inferior mean outcome.

3 Identification of causal estimands

In this section, we present nonparametric identification results. Though these results are similar to those for individualized encouragement rules in the study by Qiu et al. [9], there are two key differences. First, the form of some of the conditions in the study by Qiu et al. [9] need to be modified to account for the novel resource constraint considered here. Second, two additional conditions are needed to overcome challenges that arise due to this new constraint.

We first introduce notation that will be useful when presenting our identification results and our proposed estimators. For any observed-data distribution P , we define pointwise the conditional mean functions $\mu_p^C(t, w) := E_P(C | T = t, W = w)$ and $\mu_p^Y(t, w) := E_P(Y | T = t, W = w)$, where we use E_P to denote an expectation over observables alone under sampling from P , and their corresponding contrasts due to different treatment status, $\Delta_p^C(w) := \mu_p^C(1, w) - \mu_p^C(0, w)$ and $\Delta_p^Y(w) := \mu_p^Y(1, w) - \mu_p^Y(0, w)$. We also define the average of these contrasts conditional on V as $\delta_p^C(v) := E_P[\Delta_p^C(W) | V = v]$ and $\delta_p^Y(v) := E_P[\Delta_p^Y(W) | V = v]$, and the propensity to receive treatment $\mu_p^T(w) := P(T = 1 | W = w)$. In addition, we define $v_p(t, v) := E_P[E_P(C | T = t, W) | V = v]$, $\phi_p := E_P[\mu_p^C(0, W)]$. These quantities play an important role in tackling the problem at hand. Throughout the article, for ease of notation, if f_P is a quantity or operation indexed by distribution P , we may denote f_{P_0} by f_0 . As an example, we may use Δ_0^Y to denote $\Delta_{P_0}^Y$.

We introduce additional causal conditions we will require, positivity and unconfoundedness. In one form or another, these conditions commonly appear in the causal inference literature [16], including in the IV literature [30–33].

Condition A2. (Strong positivity). There exists a constant $\varepsilon_T > 0$ such that $\varepsilon_T < \mu_0^T(w) < 1 - \varepsilon_T$ holds for P_0 – almost every w .

Condition A3. (Unconfoundedness of treatment). For each $t \in \{0, 1\}$, T and $(C(t), Y(t))$ are conditionally independent given $W = w$ for P_0 – almost every w .

Equipped with these conditions, we are able to state a theorem on the nonparametric identification of the mean counterfactual outcomes and average treatment effect (ATE) – these results can be viewed as a corollary of the well-known G-formula [34].

Theorem 1. (Identification of ATE and expected treatment resource expenditure). *Provided Conditions A1–A3 are satisfied, it holds that $\mathbb{E}[Y(t)|W = w] = \mu_0^Y(t, w)$, $\mathbb{E}[Y(1) - Y(0)|W = w] = \Delta_0^Y(w)$, and $\mathbb{E}[Y(1) - Y(0)|V = v] = \delta_0^Y(v)$ for P_0 – almost every w and v , and so, $\mathbb{E}[Y(\rho) - Y(\rho_0^R)] = E_0[\{\rho(V) - \rho_0^R(W)\}\Delta_0^Y(W)]$. In addition, it holds that $\mathbb{E}[C(t)|W = w] = \mu_0^C(t, w)$ for P_0 – almost every w , and so, $\mathbb{E}[C(\rho)] = E_0[\rho(V)\mu_0^C(1, W) + (1 - \rho(V))\mu_0^C(0, W)]$.*

In view of Theorem 1, the objective function in (1) can be identified as follows:

$$\mathbb{E}[Y(\rho)] = E_0[\rho(V)\mu_0^Y(1, W) + (1 - \rho(V))\mu_0^Y(0, W)] = E_0[\rho(V)\Delta_0^Y(W)] + E_0[\mu_0^Y(0, W)],$$

and, similarly, the expected cost is identified as $\mathbb{E}[C(\rho)] = E_0[\rho(V)\Delta_0^C(W)] + E_0[\mu_0^C(0, W)]$. It follows that the optimization problem (1) is equivalent to

$$\text{maximize } E_0[\rho(V)\delta_0^Y(V)] \text{ subject to } E_0[\rho(V)\delta_0^C(V)] + \alpha\phi_0 \leq \kappa. \quad (2)$$

This differs from equation (3) defining optimal individualized encouragement rules in the study by Qiu et al. [9]. We now present two additional conditions so that (2) is a fractional knapsack problem [35], thereby allowing us to use existing results from the optimization literature. These conditions are similar to those in the study by Sun et al. [11].

Condition A4. (Strictly costlier treatment). There exists a constant $\varepsilon_C > 0$ such that $\Delta_0^C(w) > \varepsilon_C$ holds for P_0 – almost every w .

Condition A5. (Financial feasibility of assigning treatment). The inequality $\alpha\phi_0 < \kappa$ holds.

Condition A4 is reasonable if the treatment is more expensive than control. When applied to an IV setting as outlined in Remark 2, this condition corresponds to the assumption that the IV is indeed an encouragement to take treatment. This condition is slightly stronger than its counterpart in the study by Sun et al. [11], which only requires that $\Delta_0^C \geq 0$. This stronger condition is needed to ensure the asymptotic linearity of our proposed estimator in Section 4. Under Condition A4, it is evident that Condition A5 is reasonable because if $\alpha\phi_0 > \kappa$, then no ITR satisfies the treatment resource constraint in view of the fact that $E_0[\rho(V)\delta_0^C(V)] \geq 0$, whereas if $\alpha\phi_0 = \kappa$, then only the trivial ITR $v \mapsto 0$ satisfies the constraint, and there is no need to estimate an optimal ITR.

Under these two additional conditions, (2) is a fractional knapsack problem [35] in which every subgroup defined by a different value of V corresponds to a different “item.” A solution in the special case in which $V(W) = W$ and $\alpha = 0$ was given in Theorem 1 in the study by Sun et al. [11]. We now state a more general result with the following differences: (i) the treatment decision may be based on a summary V rather than the entire covariate vector W , and (ii) α may take any value in $[0, 1]$ rather than only zero. We also explicitly state the randomization probability at the boundary for completeness and clarity. Despite these differences, the result we obtain is similar to Theorem 1 in the study by Sun et al. [11]. Define pointwise $\xi_0(v) := \delta_0^Y(v)/\delta_0^C(v)$, and write $\eta_0 := \inf\{\eta : E_0[I(\xi_0(V) > \eta)\delta_0^C(V)] \leq \kappa - \alpha\phi_0\}$ and $\tau_0 := \max\{\eta_0, 0\}$.

Theorem 2. (Optimal ITR). *Under Conditions A1–A5, a solution to (2) is explicitly given by*

$$\rho_0(v) := \begin{cases} \frac{\kappa - \alpha\phi_0 - E_0[I(\xi_0(V) > \tau_0)\delta_0^C(V)]}{E_0[I(\xi_0(V) = \tau_0)\delta_0^C(V)]} & : \text{if } \tau_0 > 0, \quad \xi_0(v) = \tau_0 \quad \text{and} \quad E_0[I(\xi_0(V) = \tau_0)\delta_0^C(V)] > 0 \\ I(\xi_0(v) > \tau_0) & : \text{otherwise.} \end{cases}$$

Here, the first case is the boundary case with the randomization probability that saturates the treatment resource.

We also note that the reference ITRs introduced in Section 2 are also identified under the aforementioned conditions. In particular, it can be shown that $\rho_0^{\text{RD}}(v) := \min\{1, (\kappa - \alpha\phi_0)/E_0[\Delta_0^C(W)]\}$ and $\rho_0^{\text{TP}} = \mu_0^T$.

4 Estimating and evaluating optimal ITRs

In this section, we present an estimator of an optimal ITR ρ_0 and an inferential procedure for its ATE relative to a reference ITR $\rho_0^{\mathcal{R}}$, where $\mathcal{R}$ is any of FR, RD, or TP. The proposed procedure is an adaptation of the method first proposed by Qiu et al. [9,10].

We begin by introducing some notations that are useful for defining the estimands. We define the parameter $\Psi_\rho(P) := E_P[\rho(V)\Delta_P^Y(W)]$ or $\Psi_\rho(P) := E_P[\rho(W)\Delta_P^Y(W)]$ for each ITR ρ and distribution $P \in \mathcal{M}$, depending on whether the domain of ρ is $\mathcal{V}$ or $\mathcal{W}$. Here, we consider the model $\mathcal{M}$ to be locally nonparametric at P_0 [13]. For $P \in \mathcal{M}$, the ATE of an optimal ITR ρ_P relative to a reference ITR $\rho_P^{\mathcal{R}}$ equals $\Psi_{\mathcal{R}}(P) := \Psi_{\rho_P}(P) - \Psi_{\rho_P^{\mathcal{R}}}(P)$. We are interested in making inference about $\psi_0 := \Psi_{\mathcal{R}}(P_0)$, where we have suppressed dependence on $\mathcal{R}$ from our shorthand notation.

4.1 Pathwise differentiability of the ATE

We first present a result regarding the pathwise differentiability of the ATE. Pathwise differentiability of the parameter of interest serves as the foundation for constructing asymptotically efficient estimators of this parameter, based on which an inferential procedure may be developed. Additional technical conditions are required and are provided in Section S1 in the Supplementary Material. For a distribution $P \in \mathcal{M}$, a function $\mu^C : \{0, 1\} \times \mathcal{W} \rightarrow \mathbb{R}$, an ITR ρ , and a decision threshold $\tau \in \mathbb{R}$, we define pointwise the following functions:

$$\begin{aligned} D(P, \rho, \tau, \mu^C)(o) &:= \rho(v) \left[\frac{y - \mu_P^Y(t, w)}{t + \mu_P^T(w) - 1} + \Delta_P^Y(w) \right] - \Psi_e(P) \\ &\quad - \tau \left\{ \rho(v) \left[\frac{c - \mu^C(t, w)}{t + \mu_P^T(w) - 1} + \Delta^C(w) \right] + \alpha \left[\frac{(1-t)(c - \mu^C(0, w))}{1 - \mu_P^T(w)} + \mu^C(0, w) \right] - \kappa \right\}; \\ G(P)(o) &:= D(P, \rho_P, \tau_P, \mu_P^C)(o); \\ D_1(P, \mu^C)(o) &:= \frac{(1-t)(c - \mu^C(0, w))}{1 - \mu_P^T(w)} + \mu^C(0, w) - E_P[\mu^C(0, W)]; \\ D_2(P, \mu^C)(o) &:= \frac{c - \mu^C(t, w)}{t + \mu_P^T(w) - 1} + \Delta^C(w) - E_P[\Delta^C(W)]; \\ G_{\text{RD}}(P)(o) &:= D(P, \rho_P^{\text{RD}}, 0, \mu_P^C)(o) - \frac{\alpha \Psi_{\rho_P^{\text{RD}}}(P) D_1(P, \mu_P^C)}{\kappa - \phi_P} - \frac{\Psi_{\rho_P^{\text{RD}}}(P) D_2(P, \mu_P^C)}{E_P[\Delta_P^C(W)]}; \\ G_{\text{TP}}(P)(o) &:= \frac{\mu_P^T(w)}{t + \mu_P^T(w) - 1} [y - \mu_P^Y(t, w)] + t \Delta_P^Y(w) - \Psi_{\rho_P^{\text{TP}}}(P); \\ G_{\text{FR}}(P)(o) &:= D(P, \rho^{\text{FR}}, 0, \mu_P^C)(o). \end{aligned} \tag{3}$$

One key condition we rely on is the following nonexceptional law assumption.

Condition B1. (Nonexceptional law). $P_0(\xi_0(V) = \tau_0) = 0$.

Under this condition, the true optimal ITR ρ_0 is identical to an indicator function. If all covariates are discrete, then we can plug in the empirical estimates into the identification formulae in Theorems 1 and 2 and show that the resulting estimators of the ATE are asymptotically normal by the delta method even when Condition B1 does not hold. We do not further pursue this simple case in this article, and thus need to rely on the nonexceptional law assumption, namely Condition B1, to account for continuous covariates. We list additional technical conditions in Supplement S1.

We can now provide a formal result describing the pathwise differentiability of the ATE parameter.

Theorem 3. (Pathwise differentiability of the ATE). *Let $\mathcal{R} \in \{\text{FR}, \text{RD}, \text{TP}\}$. Provided Conditions A1–A5 and B1–B5 are satisfied, the parameters $P \mapsto \Psi_{\rho_P}(P)$ and $P \mapsto \Psi_{\rho_P^{\mathcal{R}}}(P)$ are pathwise differentiable at P_0 relative to $\mathcal{M}$ with canonical gradients $G(P_0)$ and $G_{\mathcal{R}}(P_0)$, respectively.*

We note that the pathwise differentiability of $P \mapsto \Psi_{\rho_P^{\mathcal{R}}}(P)$ was established in Theorem 3 in the study by Qiu et al. [9] for $\mathcal{R} \in \{\text{FR}, \text{TP}\}$. The other results can be proven using similar techniques. We presented the proof of these results in Supplement S4.2. In view of Theorem 3, it follows that the ATE parameter $\Psi_{\mathcal{R}}$ is pathwise differentiable at P_0 with nonparametric canonical gradient:

$$D_{\mathcal{R}}(P_0) := G(P_0) - G_{\mathcal{R}}(P_0) \quad (4)$$

for $\mathcal{R} \in \{\text{FR}, \text{RD}, \text{TP}\}$.

Remark 3. We have noted similar additional terms related to the resource being used in the canonical gradient of the mean counterfactual outcome or ATE of optimal ITRs under resource constraints, for example, in the studies by Luedtke and van der Laan [8] and Qiu et al. [9]. In our problem, this additional term is

$$-\tau_0 \left\{ \rho_0(v) \left[\frac{c - \mu_0^C(t, w)}{t + \mu_0^T(w) - 1} + \Delta_0^C(w) \right] + \alpha \left[\frac{(1-t)(c - \mu_0^C(0, w))}{1 - \mu_0^T(w)} + \mu_0^C(0, w) \right] - \kappa \right\}.$$

Such terms appear to come from solving a fractional knapsack problem with truncation at zero and take the form of a product of (i) the threshold in the solution, and (ii) a term that equals the influence function of the resource being used under the solution when the resource is saturated. We conjecture that such structures generally exist for fractional knapsack problems.

4.2 Proposed estimator and asymptotic linearity

We next present our proposed nonparametric procedure for estimating an optimal ITR ρ_0 and the corresponding ATE ψ_0 . We will generally use subscript n to denote an estimator with sample size n and add a hat to a nuisance function estimator that is targeted toward estimating ϕ_0 .

- (1) Use the empirical distribution $\hat{P}_{W,n}$ of W as an estimate of the true marginal distribution of W . Compute estimates $\mu_n^Y, \mu_n^C, \mu_n^T, \delta_n^Y$ and δ_n^C of $\mu_0^Y, \mu_0^C, \mu_0^T, \delta_0^Y$ and δ_0^C , using flexible regression methods. Recall that $\mu_0^Y(t, w) = E_0[Y|T = t, W = w]$, $\mu_0^C(t, w) = E_0[C|T = t, W = w]$, $\delta_0^Y(v) = E_0[\mu_0^Y(1, W) - \mu_0^Y(0, W)|V = v]$, and $\delta_0^C(v) = E_0[\mu_0^C(1, W) - \mu_0^C(0, W)|V = v]$. Define pointwise $\Delta_n^C(w) := \mu_n^C(1, w) - \mu_n^C(0, w)$.
- (2) Estimate an optimal ITR:

(a) Estimate $\phi_0 = E_0[\mu_0^C(0, W)]$ with a one-step correction estimator:

$$\phi_n := \frac{1}{n} \sum_{i=1}^n \left[\mu_n^C(0, W_i) + \frac{(1 - T_i)(C_i - \mu_n^C(0, W_i))}{1 - \mu_n^T(W_i)} \right].$$

(b) Let $\xi_n := \delta_n^Y / \delta_n^C$, $\Gamma_n : \tau \mapsto \frac{1}{n} \sum_{i: \xi_n(V_i) > \tau} \Delta_n^C(W_i)$ and $\gamma_n : \tau \mapsto \frac{1}{n} \sum_{i: \xi_n(V_i) = \tau} \Delta_n^C(W_i)$. For any $k \in [0, \infty]$, define $\eta_n(k) := \inf\{\tau : \Gamma_n(\tau) \leq k - \alpha\phi_n\}$, $\tau_n(k) := \max\{\eta_n(k), 0\}$ and

$$d_{n,k} : v \mapsto \begin{cases} \frac{k - \alpha\phi_n - \Gamma_n(\eta_n(k))}{\gamma_n(\eta_n(k))} & : \text{ if } \xi_n(v) = \eta_n(k) \text{ and } \gamma_n(\eta_n(k)) > 0, \\ I\{\xi_n(v) > \eta_n(k)\} & : \text{ otherwise.} \end{cases}$$

The rule $d_{n,k}$ is the sample analog of an ITR that prescribes treatment to those with the highest values of $\xi_0(V)$, regardless of whether the treatment is harmful, until treatment resources run out.

(c) Compute k_n , which is used to define an estimate of ρ_0 for which the plug-in estimator is asymptotically linear under conditions, as follows:

– if $\tau_n(\kappa) > 0$ and there is a solution in $k \in [0, \infty)$ to

$$\frac{1}{n} \sum_{i=1}^n d_{n,k}(V_i) \left[\Delta_n^C(W_i) + \frac{C_i - \mu_n^C(T_i, W_i)}{t_i + \mu_n^T(W_i) - 1} \right] + \alpha\phi_n = \kappa, \quad (5)$$

then take k_n to be this solution;

– otherwise, set $k_n = \kappa$.

(d) Estimate ρ_0 using the sample analog of ρ_0 with treatment resource constraint k_n , namely,

$$\rho_n : v \mapsto \begin{cases} \frac{k_n - \alpha\phi_n - \Gamma_n(\tau_n(k_n))}{\gamma_n(\tau_n(k_n))} & : \text{ if } \xi_n(v) = \tau_n(k_n) \text{ and } \gamma_n(\tau_n(k_n)) > 0 \\ I\{\xi_n(v) > \tau_n(k_n)\} & : \text{ otherwise.} \end{cases}$$

(3) Obtain an estimate $\rho_n^{\mathcal{R}}$ of the reference ITR $\rho_0^{\mathcal{R}}$ as follows:

– For $\mathcal{R} = \text{FR}$, take $\rho_n^{\mathcal{R}}$ to be ρ^{FR} .

– For $\mathcal{R} = \text{RD}$,

(a) obtain a targeted estimate $\hat{\mu}_n^C(1, \cdot)$ of $\mu_0^C(1, \cdot)$: run an ordinary least-square linear regression with outcome C , covariate $1/(T + \mu_n^T(W) - 1)$, offset $\mu_n^C(T, W)$, and no intercept. Take $\hat{\mu}_n^C$ to be the fitted mean model;

(b) take $\rho_n^{\mathcal{R}}$ to be the constant function $w \mapsto \min\left\{1, (\kappa - \phi_n) / \frac{1}{n} \sum_{i=1}^n \hat{\Delta}_n^C(W_i)\right\}$, where we define point-wise $\hat{\Delta}_n^C(w) := \hat{\mu}_n^C(1, w) - \hat{\mu}_n^C(0, w)$.

– For $\mathcal{R} = \text{TP}$, take $\rho_n^{\mathcal{R}}$ to be μ_n^T .

(4) Estimate ATE of ρ_0 relative to the reference ITR $\rho_0^{\mathcal{R}}$ with a TMLE ψ_n :

(a) obtain a targeted estimate $\hat{\mu}_n^Y$ of μ_0^Y : run an ordinary least-square linear regression with outcome Y , covariate $[\rho_n(V) - \rho_n^{\mathcal{R}}(W)]/[T + \mu_n^T(W) - 1]$, offset $\mu_n^Y(T, W)$, and no intercept. Take $\hat{\mu}_n^Y$ to be the fitted mean function.

(b) with $\hat{P}_n$ being any distribution with components $\hat{\mu}_n^Y$ and $\hat{P}_{W,n}$, take

$$\psi_n := \Psi_{\rho_n}(\hat{P}_n) - \Psi_{\rho_n^{\mathcal{R}}}(\hat{P}_n) = \frac{1}{n} \sum_{i=1}^n [\rho_n(V_i) - \rho_{n,i}^{\mathcal{R}}][\hat{\mu}_n^Y(1, W_i) - \hat{\mu}_n^Y(0, W_i)],$$

where $\rho_{n,i}^{\mathcal{R}}$ is defined as $\rho_n^{\mathcal{R}}(W_i)$ or $\rho_n^{\mathcal{R}}(V_i)$ depending on the covariate used by the reference ITR.

The aforementioned procedure is similar to that proposed in the study by Qiu et al. [9]. One key difference is the use of the refined estimator k_n of κ obtained via the estimating equation (5), which is a key to ensuring the asymptotic linearity of ψ_n . Another difference is that the denominator of ξ_n is now δ_n^C , which is consistent with our different definition of the unit value for solving the fractional knapsack problem (2). Similar to TMLE for other problems, when C or Y has known bounds (e.g., the closed interval $[0, 1]$), to obtain a corresponding targeted estimate that respect the known bounds, we may use logistic regression rather than ordinary least squares [36].

The aforementioned procedure has both similarities and substantial differences compared to the estimation procedure proposed by Sun et al. [11]. The main difference is that our procedure is targeted towards efficient estimation of and inference about the ATE of ψ_0 of the optimal ITR under a nonparametric model, while Sun et al. [11] focused on estimating the optimal ITR ρ_0 and does not evaluate this optimal ITR. This leads to a key difference between the two procedures when estimating the optimal ITR: we need to solve an estimating equation (5), which is crucial to ensuring that the estimator ψ_n is asymptotically linear, while Sun et al. [11] do not. The requirement of solving (5) is related to the nature of the fractional knapsack problem discussed in Remark 3, and we conjecture that such a calibration on the resource used is necessary for general problems of the same nature. Our procedure is also related to the method in Sun [12]. Sun [12] relied on the availability of asymptotically normal estimators of both the average benefit and average resource used (Assumption 2.4), a nontrivial requirement when the propensity score μ_0^T is unknown in observational studies. Our procedure essentially produces such estimators: in Step 4, an asymptotically normal estimator of the ATE is constructed, whereas an asymptotically normal estimator of the expected resource is produced in Step 2 and used to calibrate the resource expenditure of the estimated optimal ITR ρ_n in Step 2(c).

Remark 4. In Step 1 of the aforementioned procedure, we estimate the functions δ_0^Y and δ_0^C using a naïve approach based on outcome regression. It is viable to use more advanced techniques such as the doubly robust methods in van der Laan and Rubin [15], van der Laan and Luedtke [37], Luedtke and van der Laan [38], and Kennedy [39] or R-learning as in Nie and Wager [40]. These methods were developed for conditional ATE estimation and might lead to better estimators of δ_0^Y and δ_0^C . It is also possible to develop multiply robust methods to estimate ξ_0 using influence function techniques. Such methods to estimate ξ_0 are beyond the scope of our article, whose main focus is on the inference for the ATE. Our theoretical analysis of the estimator only applies to naïve estimators based on outcome regression, but we expect only minor modifications to be required to study these more advanced estimators once their asymptotic behavior is characterized.

Remark 5. In Step 2(a), it is also viable to use other efficient estimators of ϕ_0 , for example, a TMLE. We note that estimating ϕ_0 is only one component of estimating the optimal ITR ρ_0 . Methods such as TMLE can be preferable to ensure that the estimator respects known bounds on the estimand. However, in our case, such an improvement in estimating ϕ_0 does not necessarily lead to an improvement in the estimation of ρ_0 .

We now present results on the asymptotic linearity and efficiency of our proposed estimator. We state and discuss the technical conditions required by the theorem below in Supplement S1.

Theorem 4. (Asymptotic linearity of ATE estimator) *Let $\mathcal{R} \in \{\text{FR}, \text{RD}, \text{TP}\}$. Under Conditions B1–B12, with the canonical gradient $D_{\mathcal{R}}(P_0)$ defined in (3) and (4), it holds that*

$$\psi_n - \psi_0 = \frac{1}{n} \sum_{i=1}^n D_{\mathcal{R}}(P_0)(O_i) + o_p(n^{-1/2}).$$

Therefore, $\sqrt{n}(\psi_n - \psi_0) \xrightarrow{d} N(0, \sigma_0^2)$, where $\sigma_0^2 := E_0[D_{\mathcal{R}}(P_0)(O)^2]$. Since ψ_n is asymptotically linear with influence function equal to the canonical gradient, ψ_n is also asymptotically efficient.

To conduct inference about ψ_0 , we can directly plug the estimators of nuisance functions into $D_{\mathcal{R}}(P_0)$ to obtain a consistent estimator of $D_{\mathcal{R}}(P_0)$, and then take the sample variance to obtain a consistent estimator of the asymptotic variance σ_0^2 . The proof of Theorem 4 can be found in Supplements S4.3 and S4.4.

Remark 6. It may be desirable to use cross-fitting [41,42] to estimate an optimal ITR for better finite-sample performance. The asymptotic linearity is maintained by a similar argument that is used to prove Theorem 4. We describe this algorithm in Section S3 in the Supplementary Material.

Remark 7. We note that, unlike the study by Qiu et al. [9] where the bound κ lies in $(0, 1]$ due to the binary nature of treatment status, the methods we propose here do not require knowledge of an upper bound on treatment costs. When such a bound is indeed known (e.g., one), our methods may still be applied as long as all special cases corresponding to $\kappa = \infty$ or $\kappa < \infty$ in Section 4 are replaced by κ being equal to or less than the known bound, respectively.

5 Simulation

5.1 Simulation setting

In this simulation study, we investigate the performance of our proposed estimator of the ATE of an optimal ITR relative to specified reference ITRs. We focus here on the setting $\alpha = 1$. This scenario is more difficult than the case $\alpha = 0$ because it requires the estimation of ϕ_0 .

We generate data from a model in which the treatment T is an IV and both the treatment cost C and outcome Y are binary. This data-generating mechanism satisfies all causal conditions and has an unobserved confounder between treatment cost and outcome. We first generate a trivariate covariate $W = (W_1, W_2, W_3)$, where $W_1 \sim \text{Unif}(-1, 1)$, $W_2 \sim \text{Bernoulli}(0.8)$, and $W_3 \sim N(0, 1)$ are mutually independent. We also simulate an unobserved treatment-outcome confounder $U \sim \text{Bernoulli}(0.5)$ independently of W , and then simulate T , C , and Y as follows:

$$\begin{aligned} T|W, U &\sim \text{Bernoulli}(\text{expit}(2.5W_1 + 0.5W_2W_3)), \\ C|T, W, U &\sim \text{Bernoulli}(\text{expit}(2T - 1 - W_1 + 0.2W_2 + 0.7W_3 + 2W_1W_2 + 0.5U)), \\ Y|T, C, W, U &\sim \text{Bernoulli}(\text{expit}(-0.3C + CW_2 - W_1 + 0.2W_2 - 0.9W_3 + 0.3CU)). \end{aligned}$$

We introduce U in the data-generating mechanism to emphasize that we do not require assumptions on the joint distribution of treatment cost and outcome conditional on covariates. We consider all three reference ITRs $\mathcal{R} \in \{\text{FR}, \text{RD}, \text{TP}\}$, where we set $\rho^{\text{FR}} : v \mapsto 0$. We set $\kappa = 0.68$, which is an active constraint with $\tau_0 > 0$ and $\rho_0^{\text{RD}} < 1$.

The ITRs we consider are based on all covariates – that is, we take $V(W) = W$. We estimate the nuisance functions using the Super Learner [43] with library including a logistic regression, generalized additive model with logit link [44], gradient boosting machine [45–47], support vector machine [48,49], and neural network [50,51]. Because none of the nuisance functions follows a logistic regression model, the resulting ensemble learner is not expected to achieve the parametric convergence rate. Since both C and Y are binary, we use logistic regression rather than ordinary least squares to obtain their corresponding targeted estimates in Section 4.2. We consider sample size $n \in \{500, 1,000, 4,000, 16,000\}$ and run 1000 Monte Carlo repetitions for each sample size. We implement the algorithm that incorporates cross-fitting discussed in Remark 6 and described in Section S3 in the Supplementary Material.

To evaluate the performance of our proposed estimator, we investigate the bias and root-mean-squared error (RMSE) of the estimator. We also investigate the coverage probability and the width of nominal 95% Wald CIs constructed using influence function-based standard error estimates. We further investigate the

probability that our confidence lower limit falls below the true ATE, that is, the coverage probability of the 97.5% Wald confidence lower bound.

5.2 Simulation results

Table 1 presents the performance of our proposed estimator in this simulation. For sample sizes 500, 1,000 and 4,000, the CI coverage of our proposed method is lower than the nominal coverage 95%. When sample size is larger (16,000), the CI coverage of our proposed method increases to 90–93%. The coverage of the confidence lower bounds is much closer to nominal (97.5%) for all sample sizes considered, though, and is always approximately nominal when the sample size is large. For all reference ITRs, the bias and RMSE of our proposed estimator appear to converge to zero faster than and at the same rate as the square root of sample size, respectively. All biases are negative, which is expected in view of Remark 6. All standard errors underestimate the variation of the estimator with the extent decreasing as sample size increases.

Figure 1 presents the width of the Wald CIs scaled by the square root of sample size n . Our theory indicates that the CI width should shrink at a root- n rate, and our simulation results are consistent with this. There are some outlying cases of extremely wide or narrow CIs. This is expected for small sample sizes because the estimator of σ_0^2 in Theorem 4 resembles a sample mean and might not be close to σ_0^2 with high probability when the sample size is small. In practice, this issue might be slightly mitigated by fine-tuning the involved machine learning algorithms.

As indicated in Theorem 4, theoretical guarantees for the validity of the Wald CIs rely on the nuisance function estimators converging to the truth sufficiently quickly. It appears that the undercoverage of our Wald CI in small samples may owe, in part, to poor estimation of these nuisance functions in small sample sizes. To illustrate how our procedure may perform with improved small-sample nuisance function estimators, we conducted another two simulations: one is identical to those reported earlier in all ways except that the nuisance function estimators μ_n^Y , μ_n^C , and μ_n^T are taken to be equal to the truth; the other is a simpler scenario under a lower dimension and a parametric model. The results are presented in Section S5

Table 1: Performance of estimators of ATEs in the simulation with nuisance functions estimated via machine learning

Performance measure	Sample size	FR	RD	TP
95% Wald CI coverage	500	74%	71%	70%
	1,000	78%	74%	73%
	4,000	90%	84%	88%
	16,000	93%	90%	93%
97.5% confidence lower bound coverage	500	94%	96%	96%
	1,000	97%	98%	96%
	4,000	98%	98%	98%
	16,000	97%	98%	97%
Bias	500	−0.018	−0.018	−0.020
	1,000	−0.014	−0.013	−0.013
	4,000	−0.003	−0.004	−0.003
	16,000	−0.000	−0.001	−0.000
RMSE	500	0.056	0.039	0.046
	1,000	0.039	0.025	0.031
	4,000	0.017	0.009	0.012
	16,000	0.009	0.004	0.005
Ratio of mean standard error to standard deviation	500	0.620	0.620	0.571
	1,000	0.683	0.673	0.637
	4,000	0.868	0.765	0.809
	16,000	0.913	0.870	0.906

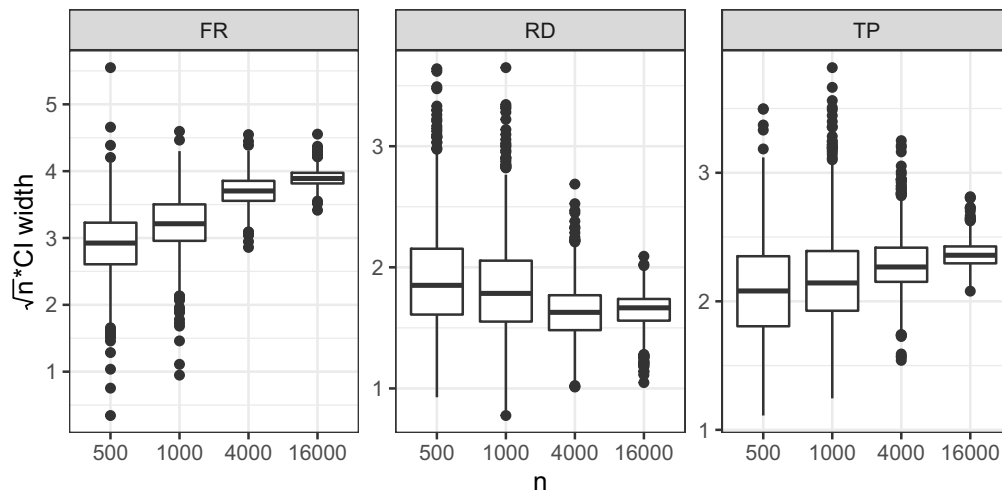


Figure 1: Boxplot of $\sqrt{n} \times \text{CI width}$ for ATE relative to each reference IER.

in the Supplementary material and suggest that our proposed estimator may achieve significantly better performance with improved machine learning estimators of the nuisance functions. This motivates seeking ways to optimize the finite-sample performance of the nuisance function estimators employed in future applications of the proposed method, possibly based on prior subject-matter expertise. The underestimation of standard errors in this simulation also motivates future work exploring whether there are standard error estimators with better finite-sample performance, for example, estimators based on the bootstrap.

6 Conclusion

There is extensive literature on estimating optimal ITRs and evaluating their performance. Among these works, only a few incorporated treatment resource constraints. In this article, we build upon the study by Sun et al. [11] and study the problem of estimating optimal ITRs under treatment cost constraints when the treatment cost is random. By using similar techniques as used in the study by Qiu et al. [9], we have proposed novel methods to estimate an optimal ITR and infer about the corresponding ATE relative to a prespecified reference ITR, under a locally nonparametric model. Our methods may also be applied to IV settings in the study by Qiu et al. [9] when the IV is intervened on.

Acknowledgments: This work was partially supported by the National Institutes of Health under award numbers DP2-LM013340 and R01HL137808. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding information: This work was partially supported by the National Institutes of Health under award numbers DP2-LM013340 and R01HL137808.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: Prof. Marco Carone is a member of the Editorial Board in the Journal of Causal Inference but was not involved in the review process of this article.

Data availability statement: Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

- [1] Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365(9454):176–86.
- [2] Varadhan R, Segal JB, Boyd CM, Wu AW, Weiss CO. A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *J Clin Epidemiol*. 2013;66(8):818–25.
- [3] Chakraborty B, Moodie EEM. Statistical methods for dynamic treatment regimes. *Statistics for biology and health*. New York, NY: Springer; 2013.
- [4] Luedtke AR, van der Laan MJ. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals Statistics*. 2016;44(2):713–42.
- [5] Murphy SA. Optimal dynamic treatment regimes. *J R Stat Soc B (Stat Methodol)*. 2003;65(2):331–55.
- [6] Robins JM. Optimal structural nested models for optimal sequential decisions. New York, NY: Springer; 2004. p. 189–326.
- [7] Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. *J Am Stat Assoc*. 2012;107(499):1106–18.
- [8] Luedtke AR, van der Laan MJ. Optimal individualized treatments in resource-limited settings. *Int J Biostat*. 2016;12(1):283–303.
- [9] Qiu H, Carone M, Sadikova E, Petukhova M, Kessler RC, Luedtke A. Optimal individualized decision rules using instrumental variable methods. *J Am Stat Assoc*. 2021;116(533):174–91.
- [10] Qiu H, Carone M, Sadikova E, Petukhova M, Kessler RC, Luedtke A. Correction to: optimal individualized decision rules using instrumental variable methods. *J Am Stat Assoc*. 2021;(just-accepted):1–2.
- [11] Sun H, Du S, Wager S. Treatment allocation under uncertain costs. 2021. arXiv: <http://arXiv.org/abs/arXiv:210311066v1>.
- [12] Sun L. Empirical welfare maximization with constraints. 2021. arXiv: <http://arXiv.org/abs/arXiv:210315298v1>.
- [13] Pfanzagl J. Estimation in semiparametric models. In: *Estimation in semiparametric models*. New York, NY, USA: Springer; 1990. p. 17–22.
- [14] van der Vaart AW. Asymptotic statistics. Cambridge, England: Cambridge University Press; 1998.
- [15] van der Laan M, Rubin D. Targeted maximum likelihood learning. *Int J Biostat*. 2006;2(1):Article 11. doi: 10.2202/1557-4679.1043.
- [16] van der Laan MJ, Rose S. Targeted learning in data science. New York, NY, USA: Springer; 2018.
- [17] Neyman J. Sur les applications de la théorie des probabilités aux expériences agricoles: Essay des principes. (Excerpts reprinted and translated to English, 1990). *Stat Sci*. 1923;5:463–72.
- [18] Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688–701.
- [19] Butler EL, Laber EB, Davis SM, Kosorok MR. Incorporating patient preferences into estimation of optimal individualized treatment rules. *Biometrics* 2018;74(1):18–26.
- [20] Chen J, Fu H, He X, Kosorok MR, Liu Y. Estimating individualized treatment rules for ordinal treatments. *Biometrics* 2018;74(3):924–33.
- [21] Imai K, Li ML. Experimental evaluation of individualized treatment rules. *J Am Stat Assoc*. 2021;1–15. doi: 10.1080/01621459.2021.1923511.
- [22] Laber E, Zhao Y. Tree-based methods for individualized treatment regimes. *Biometrika*. 2015;102(3):501–14.
- [23] Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A “SMART” design for building individualized treatment sequences. *Annual Rev Clin Psychol*. 2012;8:21–48.
- [24] Petersen ML, Deeks SG, van der Laan MJ. Individualized treatment rules: Generating candidate clinical trials. *Stat Med* 2007;26(25):4578–601.
- [25] Qian M, Murphy SA. Performance guarantees for individualized treatment rules. *Annal Stat*. 2011;39(2):1180.
- [26] Song R, Kosorok M, Zeng D, Zhao Y, Laber E, Yuan M. On sparse representation for optimal individualized treatment selection with penalized outcome weighted learning. *Stat*. 2015;4(1):59–68.
- [27] van der Laan MJ, Petersen ML. Causal effect models for realistic individualized treatment and intention to treat rules. *Int J Biostat*. 2007;3(1):Article 3. doi: 10.2202/1557-4679.1022.
- [28] Zhao YQ, Zeng D, Laber EB, Song R, Yuan M, Kosorok MR. Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*. 2015;102(1):151–68.
- [29] Zhou X, Mayer-Hamblett N, Khan U, Kosorok MR. Residual weighted learning for estimating individualized treatment rules. *J Am Stat Assoc*. 2017;112(517):169–87.
- [30] Abadie A. Semiparametric instrumental variable estimation of treatment response models. *J Econom*. 2003;113(2):231–63.
- [31] Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica*. 1994;62(2):467–75.
- [32] Tchetgen Tchetgen EJ, Vansteelandt S. Alternative identification and inference for the effect of treatment on the treated with an instrumental variable. *Harvard University Biostatistics Working Paper Series*. 2013.
- [33] Wang L, Tchetgen Tchetgen E. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J R Stat Soc B (Stat Methodol)*. 2018;80(3):531–50.

- [34] Robins J. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Math Modell.* 1986;7(9–12):1393–512.
- [35] Dantzig GB. Discrete-variable extremum problems. *Operat Res.* 1957;5(2):266–88.
- [36] Gruber S, Van Der Laan MJ. A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. *Int J Biostat.* 2010;6(1):Article 26. doi: 10.2202/1557-4679.1260.
- [37] van der Laan MJ, Luedtke AR. Targeted learning of the mean outcome under an optimal dynamic treatment rule. *J Causal Inference.* 2014;3(1):61–95.
- [38] Luedtke AR, van der Laan MJ. Super-learning of an optimal dynamic treatment rule. *Int J Biostat.* 2016;12(1):305–32.
- [39] Kennedy EH. Towards optimal doubly robust estimation of heterogeneous causal effects. 2020. arXiv: <http://arXiv.org/abs/arXiv:200414497v3>. Available from: <http://arxiv.org/abs/2004.14497>.
- [40] Nie X, Wager S. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika.* 2021;108(2):299–319.
- [41] Newey WK, Robins JR. Cross-fitting and fast remainder rates for semiparametric estimation. 2018. arXiv: <http://arXiv.org/abs/arXiv:180109138v1>.
- [42] Zheng W, van der Laan MJ. Cross-validated targeted minimum-loss-based estimation. New York, NY: Springer; 2011. p. 459–74.
- [43] van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genetics Mol Biol.* 2007;6(1):Article 25. doi: 10.2202/1544-6115.1309.
- [44] Hastie T, Tibshirani R. Generalized additive models. London: Chapman and Hall; 1990.
- [45] Friedman JH. Greedy function approximation: a gradient boosting machine. *Annal Stat.* 2001;29(5):1189–232.
- [46] Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal.* 2002;38(4):367–78.
- [47] Mason L, Baxter J, Bartlett PL, Frean M. Boosting algorithms as gradient descent; 2000. p. 512–8.
- [48] Bennett KP, Campbell C. Support vector machines: hype or hallelujah? *SIGKDD Explor Newsl.* 2000;2(2):1–13.
- [49] Cortes C, Vapnik V. Support-vector networks. *Machine Learning.* 1995;20(3):273–97.
- [50] Bishop CM. Neural networks for pattern recognition. Oxford, England: Oxford University Press; 1995.
- [51] Ripley BD. Pattern recognition and neural networks. Cambridge, England: Cambridge University Press; 2014.