Statistica Sinica Preprint No: SS-2024-0420	
Title	Learning Optimal Treatment Regimes with Survival Data under Imperfect Compliance: An Instrumental Variable Approach
Manuscript ID	SS-2024-0420
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0420
Complete List of Authors	Yifan Cui, Jianhua Guo, Wendong Li, Frank Tanser and Dongdong Xiang
Corresponding Authors	Dongdong Xiang
E-mails	ddxiang@sfs.ecnu.edu.cn
Notice: Accepted author version.	

Statistica Sinica

Learning Optimal Treatment Regimes with Survival Data under Imperfect Compliance: An Instrumental Variable Approach

Yifan Cui[1], Jianhua Guo[2], Wendong Li[3], Frank Tanser[4], Dongdong Xiang[3]

[1]*Zhejiang University,* [2]*Beijing Technology and Business University,*

[3]*East China Normal University and* [4]*Africa Health Research Institute*

Abstract: Estimating individualized optimal treatment regimes (OTR) is a central task for precision medicine. The clinical outcome of interest is often censored survival time due to reasons such as early dropout. Additionally, it is hard to completely rule out confounding by unmeasured factors in observational studies and randomized trails subject to imperfect compliance. These issues make estimating OTR extremely challenging. In this paper, we propose an instrumental variable (IV) approach to estimate OTR in the presence of data censoring and unmeasured confounding subject to imperfect compliance. By introducing a binary IV into the outcome-weighted learning framework, we establish the identification of OTR based on a no unmeasured common effect modifier assumption. We also derive a doubly robust estimator with cross-fitting to protect against model misspecification. A comparison between our proposed treatment regimes and intention-to-treat analysis further shows the superiority of our methods in

The authors are listed in alphabetical order. Corresponding author: Dongdong Xiang. Email: ddxiang@sfs.ecnu.edu.cn.

practice. We illustrate the proposed methods using simulation study and a real application to an HIV dataset, providing further empirical evidence that living in a community with high coverage of antiretroviral therapy reduces the risk of acquiring HIV.

Key words and phrases: Causal inference, unmeasured confounding, survival data, imperfect compliance, instrumental variable, optimal treatment regimes.

1. Introduction

The problem of estimating optimal treatment regimes (OTR) plays a central role in data-driven personalization and precision operation. A significant amount of work has been devoted to estimating OTR based on data from clinical trials or observational studies (Athey and Wager, 2021; Chakraborty and Moodie, 2013; Kitagawa and Tetenov, 2018; Murphy, 2003; Qian and Murphy, 2011; Zhang et al., 2012; Zhao et al., 2012); see Kosorok and Laber (2019) for a recent review. However, one major challenge in applying this line of work to medical applications is the issue of non-compliance. For instance, noncompliant behavior of patients due to the existence of unmeasured confounding frequently interferes with the effectiveness of treatments for various medical conditions and can have serious consequences. An additional difficulty arises when the clinical outcome of interest is censored survival time due to reasons such as early dropout. These two challenges

combine to make the estimation of OTR particularly difficult. Therefore, it is important to develop new methods that are appropriate for estimating OTR in the presence of both data censoring and imperfect compliance, which is the focus of this article.

In the literature, learning OTR with censored survival data has mostly been considered from a non-causal perspective. Adapting the outcome-weighted learning framework (Zhao et al., 2012), Zhao et al. (2015) proposed two classification approaches, inverse censoring outcome-weighted learning, and doubly robust outcome-weighted learning, both of which require semi-parametric estimation of the conditional censoring probability given the patient characteristics and intervention. Zhu et al. (2017) adopted the accelerated failure time model to estimate an interpretable single-tree treatment decision rule. In addition, Cui et al. (2017) proposed a random forest approach for right-censored outcome-weighted learning. Furthermore, Jiang et al. (2017) proposed a doubly robust approach to estimate OTR that optimize a user-specified function of the survival curve. More recently, Qi et al. (2020) proposed multi-armed angle-based direct learning for estimating OTR for various types of outcomes including survival data; Xue et al. (2022) proposed an angle-based approach to search the optimal dynamic treatment regimes under a multicategory treatment framework for

survival data; Cho et al. (2023) proposed using generalized random survival forests to estimate optimal dynamic treatment regimes for survival outcomes with dependent censoring. However, none of these methods considered non-compliance, which refers to situations where participants do not adhere fully to their assigned treatment or intervention protocols as specified by the study design. Patient non-compliance is often closely related to unmeasured confounding. In the presence of unmeasured confounding, non-compliance can amplify its effects by further obscuring the true relationship between the exposure and outcome. On the other hand, although there is extensive recent literature on learning OTR with an instrumental variable (IV) to deal with unmeasured confounding in both point exposure settings (Athey and Wager, 2021; Cui, 2021; Cui and Tchetgen Tchetgen, 2021a,b; Han, 2021; Qiu et al., 2021; Pu and Zhang, 2021) and longitudinal settings (Liao et al., 2021; Fu et al., 2022), censored survival data are ubiquitous in clinical trials and other biomedical research studies. Unfortunately, these IV-based methods fail to consider data censoring.

In summary, few methods for estimating OTR can comprehensively address data censoring and imperfect compliance due to unmeasured confounding, leading to limited power once all these features emerge. In this paper, we propose an instrumental variable (IV)-based classification ap-

proach for estimating OTR for randomized trials with imperfect compliance as well as observational studies with censored survival outcomes. An IV is defined as a pre-treatment variable that is independent of all unmeasured confounders and does not have a direct causal effect on the outcome other than through the treatment (Angrist et al., 1996; Imbens and Angrist, 1994). For instance, in a double-blind placebo-controlled randomized trial, random assignment is a common example of an ideal IV for the causal effect of treatment when some patients fail to comply with assigned treatments. We focus on a statistical setting in which we observe independent and identically distributed (i.i.d.) tuples $\{X_i, Z_i, A_i, Y_i, \Delta_i\}$ for $i = 1, \ldots, n$ where X_i denotes subject covariates, $Y_i = \min(T_i, C_i)$, T_i denotes the survival time, C_i denotes the censoring time, $\Delta_i = \mathbb{1}(T_i \leq C_i)$, Z_i denotes treatment assignment, and A_i denotes treatment taken. The goal of a decision-maker is to estimate the OTR given the observed dataset. Our identification condition of OTR relies on a no unmeasured common effect modifier assumption, which essentially rules out an unmeasured common effect modifier of the additive effect of treatment on the outcome, and the additive effect of the IV on treatment. Based on this assumption, we integrate a binary IV into the outcome-weighted learning framework, adjusting for data censoring using inverse-probability weighting and augmented inverse-propensity weighting.

The maximizer of the augmented inverse-propensity weighting value function is doubly robust and can be implemented efficiently with cross-fitting.

We also provide a sharp comparison between our proposed treatment regimes and estimating OTR based on an intention-to-treat (ITT) analysis. It is relatively straightforward to see that the IIT regime can deteriorate if the IV has a discouraging effect on average for certain strata to uptake the intervention. This is because the sign of the conditional average treatment effect is misidentified. In contrast, the proposed regimes remain optimal when the no unmeasured common effect modifier assumption holds. Moreover, we emphasize the importance of the policy class when selecting among different IV methods. Specifically, if the IV encourages patients within strata to uptake the intervention on average and the no unmeasured common effect modifier assumption holds, meaning the OTRs are correctly identified for both approaches, our proposed methods outperform the IIT analysis when the OTR does not belong to a pre-specified decision function class. For example, if one uses a linear interpretable decision rule while the OTR is nonlinear, our methods prove superior. The following figure illustrates this scenario. In this plot, the doubly robust approach of Zhao et al. (2015), the IIT analysis, and our proposed doubly robust approach are considered and denoted as DRO, IV0-DR, and IV-DR, respectively. The

detailed simulation setting can be found in Case 2 of Section 4, where the OTR is nonlinear. As shown in Figure 1, our approach is significantly more accurate than the IIT analysis in this context.

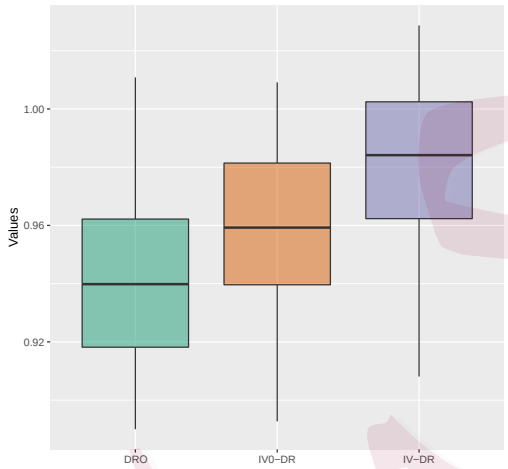

Figure 1: Boxplots of values of estimated rules using different methods, representing the survival time with higher values being preferable.

The rest of this paper is organized as follows. In Section 2, we formally introduce the model and identification of OTR under imperfect compliance and right censoring. In Section 3, we compare the proposed methods with the ITT analysis. Extensive simulation studies in Section 4 and an application on recommending living communities based on community antiretroviral therapy coverage and its impact on HIV incidence in Section 5 confirm the effectiveness of our approaches. Section 6 concludes the paper

with several remarks. Technique details are presented in the Supplementary Material.

2. Methodology

2.1 OTR and value function framework

First, we briefly introduce the basic notations and the value function framework for OTR estimation. Let $\tilde{T}$ denote the true survival time and h denote the end of the study. Following restricted mean survival time literature, we consider a deterministic transformation of $\tilde{T}$, $T = \min\{\tilde{T}, h\}$, as the outcome of interest, since there is no information on survival beyond h. Moreover, let $A \in \{+1, -1\}$ denote the binary treatment indicator and $X = (X_1, ..., X_p)^T \in \mathbb{R}^p$ denote the observed covariate vector. Our goal is to identify a treatment regime d, which is a mapping from the patient-level covariate space $\mathbb{R}^p$ to the treatment space $\{+1, -1\}$ that maximizes the expected potential outcome for the population. The goal can be formulated as estimating an OTR:

$$d^* = \arg\max_d V(d) = \arg\max_d \mathbb{E}[T_{d(X)}],$$

where $T_{d(X)}$ is the potential outcome under a hypothetical intervention that assigns treatment according to regime d.

Let T_a denote a person's potential outcome under an intervention that sets treatment to value a. In the context of randomized trails under perfect compliance, where patients will follow the treatment regimes assigned to them, a significant amount of work has been devoted to estimating OTR based on the following unconfoundedness assumption:

Assumption 1. *(Unconfoundedness)* $T_a \perp\!\!\!\perp A|X$ for $a = \pm 1$.

This assumption essentially rules out the existence of an unmeasured factor U that confounds the effect of A on T upon conditioning on X. It is straightforward to verify that under Assumption 1, one can identify the value function $V(d)$ for a given decision rule d. Furthermore, the OTR is identified from the observed data

$$d^*(X) = \text{sign}\{\tau(X)\},$$

where $\tau(X) = \mathbb{E}(T|X, A = 1) - \mathbb{E}(T|X, A = -1) = \mathbb{E}(T_1 - T_{-1}|X)$ denotes the conditional average treatment effect, $\text{sign}(x) = 1$ if $x > 0$ and $\text{sign}(x) = -1$ if $x < 0$. It has been established by Qian and Murphy (2011) that learning OTR under Assumption 1 can be equivalently formulated as

$$d^* = \arg\max_d \mathbb{E}\left[\frac{\mathbb{1}\{d(X) = A\}T}{\Pr(A|X)}\right],$$

where $\Pr(A|X)$ is the probability of taking treatment A given X. Rather than directly maximizing the above value function, Rubin and van der Laan

(2012) and Zhao et al. (2012) transformed this problem into a weighted classification problem,

$$d^* = \arg\min_{d\in\mathcal{D}} \mathbb{E}\left[\frac{T}{\Pr(A|X)}\mathbb{1}\{A \neq d(X)\}\right],$$

with 0-1 loss and weight $T/\Pr(A|X)$, where $\mathcal{D}$ is a certain policy class. The ensuing classification approach was shown to have appealing robustness properties, particularly in a randomized study where no model assumption on T is needed. Subsequent work has provided further extension and refinements of the classification perspective; see Zhao et al. (2015); Cui et al. (2017); Zhou and Kosorok (2017); Cui and Tchetgen Tchetgen (2021b) and the reference therein.

2.2 Learning OTR with imperfect compliance and right censoring

In randomized trials, we often encounter the issue that imperfect patient compliance and data censoring exist simultaneously. Imperfect compliance refers to situations where patients do not strictly adhere to assigned treatment regimes, while data censoring occurs when the survival times of some individuals are unknown or incomplete due to factors such as loss to follow-up or study termination. The first problem invalidates Assumption 1, while the second problem makes T unobservable. These two challenges collec-

tively pose significant challenges to the accurate estimation of OTR.

In this part, we address both problems simultaneously to effectively estimate OTR. Specifically, we no longer rely on Assumption 1 and thus allow for imperfect compliance due to unmeasured confounding. Let $Z \in \{+1, -1\}$ denote binary treatment assignment and $A \in \{+1, -1\}$ denote treatment taken, where Z and A may not be equal. In such cases, Z can be used as a natural IV. Let U, possibly vector-valued, be an unmeasured confounder of the effect of A on T. Meanwhile, due to data censoring, we can only observe $Y = \min\{T, C\}$ and the censoring indicator $\Delta = \mathbb{1}(T \leq C)$, where C denotes the potential censoring time. As a results, we observe data compromising n i.i.d. subjects, $\{X_i, Z_i, A_i, Y_i = \min\{T_i, C_i\}, \Delta_i = \mathbb{1}(T_i \leq C_i)\}$ for $i = 1, \ldots, n$.

Let $T_{z,a}$ represent the potential outcome if a person's IV and treatment value were set to z and a, respectively. We assume that the following latent unconfoundedness assumption holds, which essentially states that together X and U suffice to account for confounding of the joint effect of (Z, A) on T.

Assumption 2. *(Latent unconfoundedness)* $T_{z,a} \perp\!\!\!\perp (Z, A)|X, U$ for $z, a = \pm 1$.

We then make the following core IV assumptions.

Assumption 3. *(IV relevance)* $Z \not\perp\!\!\!\perp A|X$.

Assumption 4. *(Exclusion restriction)* $T_{z,a} = T_a$ for $z, a = \pm 1$ almost surely.

Assumption 5. *(IV independence)* $Z \perp\!\!\!\perp U|X$.

Assumption 6. *(IV positivity)* $0 < \Pr(Z = 1|X) < 1$ almost surely.

Assumption 7. *(No unmeasured common effect modifier)*

$$Cov\left\{\widetilde{\delta}(X,U), \widetilde{\tau}(X,U)|X\right\} = 0 \text{ almost surely,}$$

where $\widetilde{\delta}(X,U) \triangleq \Pr(A = 1|Z = 1, X, U) - \Pr(A = 1|Z = -1, X, U)$ and $\widetilde{\tau}(X,U) \triangleq \mathbb{E}(T_1 - T_{-1}|X,U)$.

Remark 1. Assumption 3 requires that the IV is associated with the treatment conditional on baseline covariates. Note that Assumption 3 does not rule out confounding of the Z-A association by an unmeasured factor, however, if present, such factor must be independent of U. Assumption 4 states that there can be no direct causal effect of treatment assignment Z on survival time T not mediated by treatment taken A. In a double-blinded randomized controlled trial with non-compliance, this assumption holds naturally. Assumption 5 assumes a conditional independence of Z and U given the baseline covariates. The positivity Assumption 6 is standard and needed for nonparametric identification. Assumption 7 essentially

rules out the presence of unmeasured common effect modifiers that simultaneously influence the additive effect of the treatment assignment Z on the received treatment A, and the additive effect of the actual treatment A on the outcome T (Cui and Tchetgen Tchetgen, 2021b). While not always intuitive, this assumption has been adopted in recent literature on instrumental variable methods for OTR learning (Michael et al., 2024; Ye et al., 2023). Intuitively, Assumption 7 rules out the possibility that individuals who are more likely to comply (given X and U) systematically experience larger or smaller treatment effects. This helps to separate the instrumental mechanism from effect modification due to unobserved factors.

Based on the above assumptions, it has been well demonstrated in the literature that when T is observable, the OTR can be nonparametrically identified based on the weighted classification framework mentioned in Section 2.1 (Cui and Tchetgen Tchetgen, 2021b). However, since we can only observe the censored version of T, i.e., Y, the OTR becomes unidentifiable in the presence of both imperfect compliance and data censoring. This dual challenge requires a novel approach to accurately identify the value function $V(d)$. To address this issue, we introduce Assumptions 8–9. In particular, Assumption 8 imposes a conditional independent censoring condition, which assumes that the censoring time C is independent of the survival time T

and the unmeasured confounder U , conditional on X, A, Z. This may be plausible in applications where censoring is mainly due to administrative reasons or loss to follow-up unrelated to unmeasured health status. While this may be restrictive in certain settings, it serves as a working assumption in practice to enable identification under both unmeasured confounding and censoring. See Section 6 for more discussions.

Assumption 8. (Independent censoring) $C \perp\!\!\!\perp (T, U)|X, A, Z$.

Assumption 9. (Censoring positivity) $\Pr(C \leq h|X, A, Z) \leq 1 - M_c$ for some $0 < M_c \leq 1$.

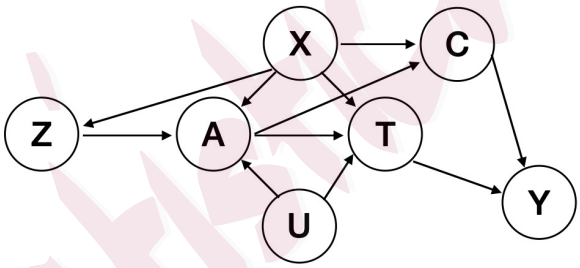

Figure 2: Causal directed acyclic graph with unmeansured confounding and censoring.

Figure 2 provides a graphical representation of the above Assumptions

2-9. It can be shown that

$$d^* = \arg\max_{d\in\mathcal{D}} \mathbb{E}[T_{d(X)}] = \arg\max_{d\in\mathcal{D}} \mathbb{E}\left[\frac{ZAT\mathbb{1}\{A=d(X)\}}{\delta(X)\Pr(Z|X)}\right],$$

where $\delta(X) = \Pr(A=1|Z=1,X) - \Pr(A=1|Z=-1,X)$. Let $S_C(t|X,A,Z) = \Pr(C>t|X,A,Z)$ be the conditional treatment specific survival function for the censoring time given covariates. Recall that $T = \min\{\tilde{T}, h\}$. Then we have

$$\begin{aligned}\mathbb{E}\left[\frac{\Delta Y}{S_C(Y|X,A,Z)}|X,A,Z\right] &= \mathbb{E}\left[\frac{h\mathbb{1}\{\tilde{T}\geq h\}\mathbb{1}\{C>h\}}{S_C(h|X,A,Z)} + \frac{\tilde{T}\mathbb{1}\{\tilde{T}<h\}\mathbb{1}\{C>\tilde{T}\}}{S_C(\tilde{T}|X,A,Z)}|X,A,Z\right] \\ &= \mathbb{E}\left[h\mathbb{1}\{\tilde{T}\geq h\} + \tilde{T}\mathbb{1}\{\tilde{T}<h\}|X,A,Z\right] = E(T|X,A,Z).\end{aligned}$$

Considering the conditional expectation of T, the following Theorem 1 gives one of our main identification results, and states that the OTR can be nonparametrically identified under Assumptions 2-9, therefore generalizing the identification of OTR under imperfect compliance with a valid IV to account for data censoring.

Theorem 1. Under Assumptions 2-9, for any $\mathcal{D}$, we have

$$\arg\max_{d\in\mathcal{D}} \mathbb{E}[T_{d(X)}] = \arg\max_{d\in\mathcal{D}} \mathbb{E}\left[\frac{ZA\Delta Y\mathbb{1}\{A=d(X)\}}{S_C(Y|X,A,Z)\delta(X)\Pr(Z|X)}\right]. \tag{2.1}$$

It is clear that maximization task (2.1) is equivalent to

$$\arg\min_{d\in\mathcal{D}} \mathbb{E}\left[W\mathbb{1}\{A\neq d(X)\}\right], \tag{2.2}$$

where

$$W = \frac{ZA\Delta Y}{S_C(Y|X,A,Z)\delta(X)\Pr(Z|X)}.$$

Minimization task (2.2) can be interpreted as a weighted classification problem in which one aim to classify A using X by minimizing the weighted misclassification error given by the weighted outcome W. It should be noted that when, for example, $Z \neq A$, the weight W may not be positive. To solve this problem, we further modify the weights based on the following equality, as inspired by Liu et al. (2016):

$$\arg\min_{d\in\mathcal{D}} \mathbb{E}\left[W\mathbb{1}\{A \neq d(X)\}\right] = \arg\min_{d\in\mathcal{D}} \mathbb{E}\left[|W|\mathbb{1}\{\text{sign}(W)A \neq d(X)\}\right]. \tag{2.3}$$

Noticing that direct optimization is tricky because of the discontinuous indicator function $\mathbb{1}\{\cdot\}$, we consider a convex relaxation of (2.3) by using the hinge loss function $\phi(t) = \max(1-t, 0)$, one of the most popular loss functions in the context of classification. Furthermore, to avoid overfitting, a regularization term is added to penalize the complexity of the decision function. As a result, we propose to estimate the OTR by minimizing the following regularized objective function:

$$\hat{g} = \arg\min_{g\in\mathcal{G}} \frac{1}{n}\sum_{i=1}^{n} |W_i|\phi[\text{sign}(W_i)A_i g(X_i)] + \lambda_n ||g||^2, \tag{2.4}$$

where λ_n is a tuning parameter, $||g||$ is some norm, $\mathcal{G}$ is a function class,

and g encodes the decision function d. The estimated OTR is given by $\hat{d} = \text{sign}(\hat{g})$.

In (2.4), W_i is unknown and should be estimated from the observed data. To this end, we denote

$$\hat{W}^{IW} = \frac{ZA\Delta Y}{\hat{S}_C(Y|X,A,Z)\hat{\delta}(X)\hat{\Pr}(Z|X)},$$

and they can be substituted for W_i, where IW stands for "inverse weighted estimator", $\hat{\delta}(X)$ and $\hat{\Pr}(Z|X)$ can be obtained from logistics regression models, and $\hat{S}_C(t|X,A,Z)$ can be obtained by using the Cox proportional hazards model (Cox, 1972), survival forests (Ishwaran et al., 2008; Zhu and Kosorok, 2012), or neural nets (Zhong et al., 2021, 2022). Then, minimization task (2.4) can be solved efficiently using support vector machines (SVM) techniques. We refer to Zhao et al. (2012) for solving this problem with linear and nonlinear decision rules.

2.3 Doubly robust learning of OTR

In general, a misspecified model for C given (X, A, Z) in the formulation above may result in biased estimate, and inverse-probability of censoring weighting is generally not robust to estimation errors in S_C. Moreover, although $\hat{W}^{IW}$ is attractive in terms of its simplicity, this estimation throws away all observations with $\Delta_i = 0$, and this may hurt us in terms of ef-

ficiency. To mitigate model misspecification and improve estimation efficiency, we propose a doubly robust estimator. The estimator remains consistent if either the survival model or the censoring model is correctly specified, provided that the propensity score models are correctly specified. Specifically, let $\tilde{\mathbb{E}}(T|T > t, X, A, Z)$ denote a working model for the conditional mean residual lifetime given (X, A, Z) derived from the survival model for $S_T(t|X, A, Z)$, where $S_T(t|X, A, Z) = \Pr(T > t|X, A, Z)$ is the conditional treatment specific survival function for the survival time given covariates, and let $\tilde{S}_C(t|X, A, Z)$ denote a working model for $S_C(t|X, A, Z)$. Then we propose the following augmented value function,

$$\tilde{V}(d) = \mathbb{E}([\frac{\Delta Y}{\tilde{S}_C(Y|X, A, Z)} + \frac{(1-\Delta)\tilde{\mathbb{E}}(T|T > Y, X, A, Z)}{\tilde{S}_C(Y|X, A, Z)} - \int_0^Y \frac{\lambda_C(s|X, A, Z)}{\tilde{S}_C(s|X, A, Z)}\tilde{\mathbb{E}}(T|T > s, X, A, Z)\mathrm{d}s]\frac{ZA\mathbb{1}\{A = d(X)\}}{\delta(X)\Pr(Z|X)}).$$

where $\lambda_C(s|X, A, Z)$ is the hazard function of the censoring time. It can be easily shown that $\tilde{V}(d)$ is equivalent to $V(d)$ if either working model is correct. Define

$$Q(X, A, Z, \Delta, Y, S_C, S_T) = \frac{\Delta Y + (1-\Delta)\mathbb{E}(T|T > Y, X, A, Z)}{S_C(Y|X, A, Z)} - \int_0^Y \frac{\lambda_C(s|X, A, Z)}{S_C(s|X, A, Z)}\mathbb{E}(T|T > s, X, A, Z)\mathrm{d}s.$$

We then give the identification of OTR via doubly robust transformation of the response.

Theorem 2. Under Assumptions 2-9, for any $\mathcal{D}$, we have

$$\arg\max_{d\in\mathcal{D}} \mathbb{E}[T_{d(X)}] = \arg\max_{d\in\mathcal{D}} \mathbb{E}\left[Q(X, A, Z, \Delta, Y, S_C, S_T)\frac{ZA\mathbb{1}\{A = d(X)\}}{\delta(X)\Pr(Z|X)}\right].$$

The above Theorem 2 shows that progress can be made towards identifying OTR using doubly robust transformation of the survival response. The proofs of Theorems 1-2 are provided in Appendix A. To save space, the theoretical results of Fisher consistency, excess risk bound and universal consistency of the estimated treatment regimes are also provided in the supplementary material. In such cases, the weighted classification approach can be applied to estimate the OTR with weights

$$\hat{W}^{DR} = Q(X, A, Z, \Delta, Y, \hat{S}_C, \hat{S}_T)\frac{ZA}{\hat{\delta}(X)\hat{\Pr}(Z|X)}.$$

Meanwhile, the proposed doubly robust procedure can be well combined with the cross-fitting technique to yield better statistical properties (Zivich and Breskin, 2021). The cross-fitting procedure goes as follows. We randomly split data into K folds, and the cross-fitted estimator of the doubly robust value function is given by

$$\frac{1}{K}\sum_{k=1}^{K} P_{n,k}\left\{Q(X, A, Z, \Delta, Y, \hat{S}_{C,-k}, \hat{S}_{T,-k})\frac{ZA\mathbb{1}\{A = d(X)\}}{\hat{\delta}_{-k}(X)\hat{\Pr}_{-k}(Z|X)}\right\},$$

where $P_{n,k}$ denotes empirical averages only over the k-th fold, and the subscript $_{-k}$ denotes the nuisance estimators constructed excluding the k-th fold.

3. A Comparison with Intention-To-Treat Analysis and Choice of Decision Class

In this section, we make a comparison of the proposed OTR with an intention-to-treat (ITT) analysis. In medicine, an ITT analysis of a randomized trial is based on the initial treatment assignment Z and not on the treatment eventually received A. In other words, ITT analysis ignores non-compliance. As given in Cui and Tchetgen Tchetgen (2021b) and Qiu et al. (2021), the IIT analysis identifies the so-called complier OTR under a set of assumptions essentially excluding defiers in the population. Recall that the proposed methods essentially target at Equation (2.2), with the weights W^{IW} and W^{DR}. We denote the resulting regime from Equation (2.2) as $\tilde{d}$. In contrast, if one considers ITT analysis, on a population level, one aims at optimizing

$$d^{ITT} = \arg\min_{d\in\mathcal{D}} \mathbb{E}\left[W\mathbb{1}\{Z \neq d(X)\}\right], \tag{3.5}$$

with weights W equal to

$$\frac{\Delta Y}{S_C(Y|X,A,Z)\Pr(Z|X)},$$

or

$$\frac{Q(X,A,Z,\Delta,Y,S_C,S_T)}{\Pr(Z|X)}.$$

Interestingly, by Proposition 5 given in the supplementary file, the objective of the IIT analysis in (3.5), that is d^{ITT}, can be written as Equation (2.2) with weights

$$W_0^{IW} \triangleq \frac{ZA\Delta Y}{S_C(Y|X,A,Z)\Pr(Z|X)},$$

or

$$W_0^{DR} \triangleq Q(X,A,Z,\Delta,Y,S_C,S_T)\frac{ZA}{\Pr(Z|X)},$$

respectively. It is important to note that the only difference between W^{IW}, W^{DR} and W_0^{IW}, W_0^{DR} is a term of $\delta(X)$ on the denominator. If $\delta(X) > 0$ almost surely, we can see that the two optimizations of ours and ITT analysis lead to the same global optimal as long as $d^* \in \mathcal{D}$, where recall that d^* is the global optimal defined as the solution to optimization (2.2) with $\mathcal{D}$ being the unrestricted policy class, i.e., the decision function g belongs to the class of all measurable functions. However, in many real-world applications, $\delta(x) > 0$ for every x might not be realistic. The assumption of $\delta(x) > 0$ implies that the IV has an encouragement on average for patients within strata to uptake the intervention, which is hardly to be true as there might always be a subgroup of people who is inclined to comply, which is known as strategic self-anticonformity or reverse psychology. The idea behind reverse psychology is that by pushing for the opposite, the individual

would choose to engage in the behavior that is truly desired. We note that $\delta(x)$ is different from the commonly assessed compliance rate in IV analysis as the latter one is an average. For example, the average compliance rate can be large, say 80% or 90%, but there might be some $\delta(x) < 0$. This in fact matches our intuition about human behavior: there might be a subgroup of participants who are anti-conformists, though this proportion of people might not be that many.

Moreover, as illustrated in Figure 1 in the introduction, one might want to consider a parsimonious or interpretable decision rule in certain applications, for example, a linear decision rule (Mo et al., 2021) or an interpretable policy tree (Sverdrup et al., 2020). In such scenarios, even if $\delta(X) > 0$ almost surely, the proposed OTR are theoretically better than ITT analysis.

In conclusion, if one uses the proposed estimation pipeline, then the resulting estimated treatment regime outperforms ITT analysis, that is,

Proposition 1. Under Assumptions 2-9, for any given class $\mathcal{D}$, we have $\mathbb{E}[T_{\tilde{d}(X)}] \geq \mathbb{E}[T_{d^{ITT}(X)}]$.

4. Simulation Studies

In this section, we report simulation results of the proposed two estimators based on $\hat{W}^{IW}$ and $\hat{W}^{DR}$ (denoted as IV-IW and IV-DR), respectively. To

the best of our knowledge, there is currently no method that can effectively solve the problem of non-compliance and data censoring at the same time. For comparison, the following methods for learning OTR are considered: (i) Cox regression; (ii) the inverse censoring weighted and the doubly robust methods (Zhao et al., 2015, denoted as ICO and DRO); and (iii) ITT analysis (denoted as IV0-IW and IV0-DR) described in Section 3.

We generated $X = (X_1, ..., X_5)^T$ from a uniform distribution on $[-1, 1]^5$. The IV Z was a Bernoulli event with probability 0.5. Further details on the generations of A, U, T and C are provided later. The proposed methods were implemented according to Section 2. Specifically, we chose the proportional hazards model as the working model for T and C given (X, A), where $\hat{\delta}(X) = \hat{\Pr}(A|X, Z = 1) - \hat{\Pr}(A|X, Z = -1)$ and $\hat{\Pr}(A|X, Z)$, $\hat{\Pr}(Z|X)$ were estimated using logistic regression. For each case, the training sample size was 1000, and a test dataset with 10000 individuals was generated to evaluate the performance of different methods. We repeated the simulation 500 times. For demonstration, we presented the results based on linear kernels, and the performances of Gaussian kernels were comparable to linear kernels based on the empirical results. We used five-fold cross-validation for choosing the tuning parameter λ_n over a prespecified grid. In the following, we consider four generative models.

Case I. The treatment A was a Bernoulli event with probability $\Pr(A = 1|X, Z, U) = \Phi(0.5X_1 - 0.8U)(1 - \Delta) + \mathbb{1}(Z = 1)\Delta$, where $\Delta = \Phi(0.5X_1)$ and U was generated from a bridge distribution (Wang and Louis, 2003) with parameter $\phi = 0.5$. The survival time T is the minimum of $\tilde{T}$ and $h = 2$, where $\tilde{T}$ is generated with hazard rate function

$$\lambda_{\tilde{T}}(t, |X, A, U) = \lambda_{\tilde{T}0}(t)\exp\{0.6X_1 - 0.8X_2 + (-0.4X_1 - 0.2X_2 - 0.4X_3)A + 0.5U\},$$

and $\lambda_{\tilde{T}0}(t) = 2t$. The censoring time C is generated with hazard rate function

$$\lambda_C(t, |X, A) = \lambda_{C0}(t)\exp\{0.6X_1 + (0.4X_1 - 0.6X_2)A\}$$

and $\lambda_{C0}(t) = 2t$. The censoring rate is around 49%. The compliance rate is around 74%. The optimal decision rule is linear with $d^*(X) = -\text{sign}(-0.4X_1 - 0.2X_2 - 0.4X_3)$.

Case II. The treatment A was a Bernoulli event with probability $\Pr(A = 1|X, Z, U) = \Phi(0.5X_1 - 0.8U)(1 - \Delta) + \mathbb{1}(Z = 1)\Delta$, where $\Delta = \Phi(0.5X_1)$ and U was generated from a bridge distribution with parameter $\phi = 0.5$. The survival time T is the minimum of $\tilde{T}$ and $h = 2$, where $\tilde{T}$ is generated with hazard rate function

$$\lambda_{\tilde{T}}(t, |X, A, U) = \lambda_{\tilde{T}0}(t)\exp\{0.6X_1 - 0.8X_2 + \sin(2\pi X_1)A + 0.5U\}.$$

The censoring time C is generated with hazard rate function

$$\lambda_C(t, |X, A) = \lambda_{C0}(t)\exp\{0.6X_1 + (0.4X_1 - 0.6X_2)A\}.$$

The censoring rate is around 47%. The compliance rate is around 74%. The optimal decision rule is nonlinear with $d^*(X) = -\text{sign}(\sin(2\pi X_1))$.

Case III. The treatment A was a Bernoulli event with probability $\Pr(A = 1|X, Z, U) = \text{expit}\{2X_1 + 2.5Z - 2U\}$, where U was generated from a uniform distribution on $[-1, 1]$. The survival time T is the minimum of $\tilde{T}$ and $h = 3$, where $\tilde{T}$ is generated with linear model

$$\tilde{T} = 2.4 + 0.5X_1 + 0.4X_2 + (0.6X_4 - 0.4X_5)A + 1.5U + \epsilon,$$

where $\epsilon \sim N(0, 0.1^2)$. The censoring time C is generated with hazard rate function

$$\lambda_C(t, |X, A) = \lambda_{C0}(t)\exp\{-1.6 + 0.6X_1 + (0.4X_1 - 0.6X_2)A\}.$$

The censoring rate is around 67%. The compliance rate is around 86%. The optimal decision rule is linear with $d^*(X) = \text{sign}(0.6X_1 - 0.4X_2)$.

Case IV. The treatment A was a Bernoulli event with probability $\Pr(A = 1|X, Z, U) = \text{expit}\{2X_1 + 1.8X_3Z + 1.4Z - 3U\}$, where U was generated from a uniform distribution on $[-1, 1]$. The other setups are the

same as in Case 3. The censoring rate is around 67%. The compliance rate is around 64%. The assumption of $\delta(X) > 0$ holds for all individuals in the above three cases, while it is not always the case here because of the interaction between X_3 and Z.

Given that the data generating mechanism is known in each case, we compute the value function in the test set for each of the 500 replications, evaluated at the estimated optimal treatment regimes. The results for Cases 1–4 are presented in Figure 3, where a larger value indicates a longer survival time. We chose to use boxplots, a common practice in the literature, as they effectively illustrate the distributional variability across replications, including median performance and quantile-based comparisons. Overall, the proposed IV-IW and IV-DR methods perform well, yielding longer survival times under both censoring and imperfect compliance. From the boxplots, it can be seen that in Cases 1, 3, and 4, the value functions of the proposed regimes are close to the empirical optimal value, suggesting strong performance. While there is some overlap in the interquartile ranges, the distributional differences—particularly in the medians and the upper and under quartiles—consistently indicate that IV-DR outperforms IV0-DR and other baseline methods. In Case 2, a noticeable gap remains between all estimated methods and the optimal value. This is expected, as the optimal

regime in this case is nonlinear and lies outside the policy class.

When the optimal decision rule is linear and $\delta(X) > 0$ (i.e., Cases 1 and 3), the proposed IV-IW and IV-DR methods perform similarly to the IV0-IW and IV0-DR methods of ITT analysis. Otherwise, when the optimal decision rule is nonlinear (i.e., Case 2) or when there are some $\delta(x) < 0$ (i.e., Case 4), the value functions of the proposed IV-IW and IV-DR methods are significantly larger. These observations align with the discussion in Section 3, which further demonstrate the superiority of our methods. Throughout, Cox, ICO and DRO perform worse than other methods, which is expected because they do not account for unmeasured confounding.

It should be noted that our work is primarily motivated by the presence of imperfect compliance, where treatment assignment and treatment taken may differ, leading to unmeasured confounding. While our method remains applicable in settings without such confounding, its primary advantage lies in providing robust estimation under noncompliance. Therefore, the no-confounding case is not the main focus of our study.

5. Real Data Analysis

In this section, we apply the proposed methods to an HIV dataset studied in Tanser et al. (2013), which comes from one of Africa's largest population-

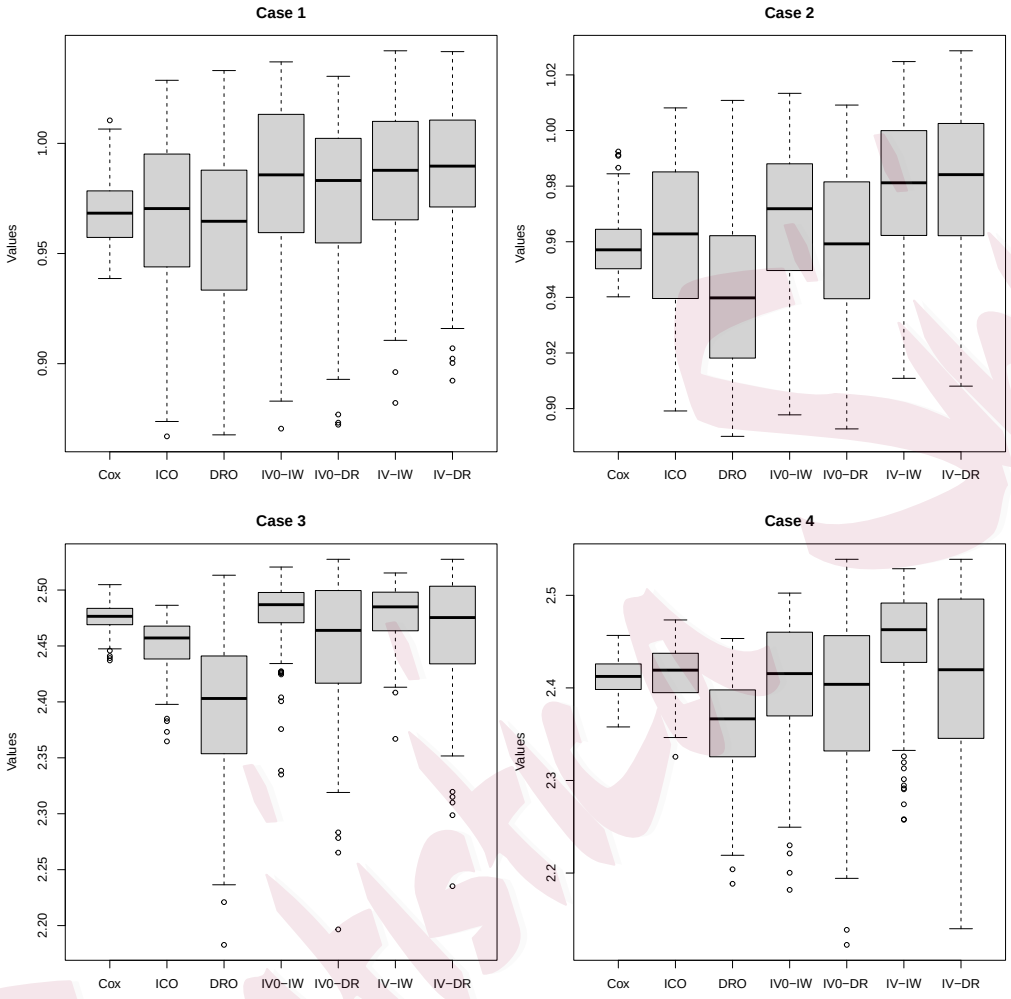

Figure 3: Boxplots of values of estimated rules using different methods, representing the survival time with higher values being preferable. The empirical optimal values are 1.025, 1.127, 2.512, 2.513 for four cases, respectively.

based prospective cohort studies to follow up individuals who were HIV-uninfected at baseline and can be downloaded from the AHRI data repository (`https://data.ahri.org/index.php/home`). In this study and several subsequent studies (Tanser et al., 2020), strong evidence was found that nurse-led, devolved, public-sector antiretroviral therapy (ART) programmes in rural sub-Saharan African settings can reduce HIV incidence. With accounting for unmeasured confounding and data censoring, we show further evidence that living in a community with high coverage of ART substantially reduces the risk of acquiring HIV.

The HIV dataset is reorganized as follows to fit in the current context. First, 6177 individuals who were HIV-negative on 5 June 2008 and had complete covariates and instrumental variable are involved in our analysis. We also set 1 January 2013 as the end of the study. Individuals who were still HIV-negative at the end or quit the study earlier are considered as censoring, and individuals who were HIV-positive before the end are considered as failures. The following seven covariates are considered: number of partners in the past 12 months; marital status; wealth; age; gender; mode of community; and community HIV prevalence rate. ART coverage is defined as the proportion of all HIV-infected individuals receiving ART at every location. Note that ART coverage and HIV prevalence of an individual's

community were measured by means of a moving two-dimensional Gaussian kernel with a search radius of 3 km. For demonstration, ART coverage is dichotomized at 30%. In other words, $A = -1$ if ART coverage is less than 30% and $A = 1$ otherwise. We choose the travel distance to the nearest ART facility as the instrumental variable, i.e., $Z = -1$ if the distance is 3.8 km or more and $Z = 1$ otherwise. This instrumental variable is found to be strongly associated with ART coverage, and 83% of the individuals have the same values of A and Z.

In addition to the proposed methods, the methods considered in the simulation studies are also implemented. All the methods are implemented with a linear basis, and we use Cox regression model for both survival time and censoring time. To evaluate the performance of the considered methods, a cross-validated analysis is employed. Specifically, at each run, the whole dataset is randomly partitioned into five parts, where one part is treated as training data to estimate the individualized treatment rules and the rest four parts are treated as test data. The cross-validated values are obtained by averaging the empirical values on all five test data, where empirical values of estimated treatment rule $\hat{d}$ were evaluated with

$$V_1 = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_i^{IW} \mathbb{1}\{A_i = \hat{d}(X_i)\}$$

and

$$V_2 = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_i^{DR} \mathbb{1}\{A_i = \hat{d}(X_i)\}$$

where $\hat{W}_i^{IW}$ and $\hat{W}_i^{DR}$ are estimated by respectively logistic regression and doubly robust method using test data and n is the test sample size. The above procedure is repeated 100 times. Tuning parameters are selected in the same way as Section 4. The results are shown in Table 1, where larger values indicate longer survival time. From the table, we observe that the proposed IV-IW and IV-DR methods generally outperform the rest with larger value functions. Moreover, by using doubly robust methods, we achieve uniformly better results in terms of larger mean values and smaller standard errors. Moreover, Table 2 presents the percentage of assigning $A = 1$ according to the estimated OTR for each method. A higher percentage provides further evidence that high coverage of ART reduces the risk of acquiring HIV.

6. Discussion

In this paper, we have proposed a general IV approach for learning optimal treatment regimes with survival data under imperfect compliance. We established identification of the optimal regimes $\arg\max_{d\in\mathcal{D}} \mathbb{E}[T_{d(X)}]$ under right-censoring with the aid of a binary IV. We also constructed doubly ro-

Table 1: Real data analysis: mean $\times 10^{-2}$ (standard error $\times 10^{-2}$) of V_1 and V_2.

Method	V_1	V_2
Cox	52.73 (0.88)	63.15 (1.43)
ICO	46.30 (0.65)	58.59 (1.30)
DRO	47.51 (0.71)	57.20 (1.24)
IV0-IW	56.64 (0.94)	68.34 (1.72)
IV0-DR	58.47 (0.98)	70.08 (1.81)
IV-IW	57.77 (0.95)	70.36 (1.79)
IV-DR	59.25 (0.99)	71.54 (1.86)

Table 2: Percentage of assigning $A = 1$ according to the estimated OTR.

Method	Cox	ICO	DRO	IV0-IW	IV0-DR	IV-IW	IV-DR
Percentage	0.62	0.38	0.44	0.67	0.68	0.70	0.73

bust classification-based estimators. We made a sharp comparison between our proposed treatment regimes and optimal treatment regimes based on an intention-to-treat analysis. Our approaches were illustrated via extensive simulation studies and a real data application.

We conclude by outlining several promising directions for future research. First, although our current framework focuses on a binary IV, it can be extended to accommodate more general instruments, such as multi-level or continuous IVs. This extension is particularly relevant in modern applications involving multiple treatment arms or continuous dosage levels (Sun et al., 2017; Chen et al., 2016; Qi et al., 2020). Second, weak IVs can be problematic. It may be possible to estimate OTRs by empirically strengthening instruments (Zubizarreta et al., 2013; Baiocchi et al., 2010; Ertefaie et al., 2018). Third, estimating OTRs for point process treatment and outcome with IVs (Jiang et al., 2023) should make a fruitful avenue of future research. Fourth, in certain trials such as psychological studies, treatment assignment might be an invalid IV. It would be interesting to consider other approaches such as invalid IV methods (Liu et al., 2020; Sun et al., 2022, 2023; Tchetgen Tchetgen et al., 2021), single proxy control framework (Park et al., 2024), and proximal causal inference framework (Cui et al., 2024; Miao et al., 2018; Tchetgen Tchetgen et al., 2020; Ying

et al., 2022). In addition, while our method addresses imperfect compliance in observational data via an IV strategy, the learned treatment regime is designed to be deployed assuming full compliance. In practice, however, adherence may be partial or heterogeneous, potentially attenuating real-world effectiveness. Incorporating compliance behavior into the regime learning framework would therefore be an important extension. Furthermore, although standard IV models imply testable inequality constraints on the observed data distribution, how such implications translate to the censoring-adjusted IV framework remains unclear. Establishing analogous testable conditions in the presence of censoring would allow formal assessment of IV validity and strengthen empirical applications. Finally, our framework relies on the conditional independence assumption between C and U. This assumption may not hold in practice, especially in scenarios where unmeasured risk factors influence both dropout and survival. Future methodological developments are needed to relax this assumption, potentially through models that allow for dependent censoring.

Supplementary Materials

Detailed proofs of Theorems 1-2 and Proposition 1 as well as the theoretical results of Fisher consistency, excess risk bound and universal consistency of

the estimated treatment regimes.

Acknowledgements

The authors thank to the editors and anonymous referees for their valuable comments and constructive suggestions that improve the quality of this work significantly. This work was supported by National Key R&D Program of China (2024YFA1015600, 2022YFA1003801, 2021YFA1000101, 2021YFA1000102, 2020YFA0714102), National Natural Science Foundation of China (12431009, 12471254, 12471266, 12201382, 12071144, U23A2064), Shanghai Pilot Program for Basic Research (TQ20240201), Basic Research Project of Shanghai Science and Technology Commission (22JC1400800).

References

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444–455.

Athey, S. and S. Wager (2021). Policy learning with observational data. *Econometrica 89*(1), 133–161.

Baiocchi, M., D. S. Small, S. Lorch, and P. R. Rosenbaum (2010). Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association 105*(492), 1285–1296.

REFERENCES

Chakraborty, B. and E. Moodie (2013). *Statistical methods for dynamic treatment regimes*. Springer.

Chen, G., D. Zeng, and M. R. Kosorok (2016). Personalized dose finding using outcome weighted learning. *Journal of the American Statistical Association 111*(516), 1509–1521.

Cho, H., S. T. Holloway, D. J. Couper, and M. R. Kosorok (2023). Multi-stage optimal dynamic treatment regimes for survival outcomes with dependent censoring. *Biometrika 110*(2), 395–410.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological) 34*(2), 187–202.

Cui, Y. (2021). Individualized decision-making under partial identification: Three perspectives, two optimality results, and one paradox. *Harvard Data Science Review 3(3)*, 1–19.

Cui, Y., H. Pu, X. Shi, W. Miao, and E. Tchetgen Tchetgen (2024). Semiparametric proximal causal inference. *Journal of the American Statistical Association 119*(546), 1348–1359.

Cui, Y. and E. Tchetgen Tchetgen (2021a). On a necessary and sufficient identification condition of optimal treatment regimes with an instrumental variable. *Statistics & Probability Letters 178*, 109180.

Cui, Y. and E. Tchetgen Tchetgen (2021b). A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association 116*(533), 162–173.

REFERENCES

Cui, Y., R. Zhu, and M. Kosorok (2017). Tree based weighted learning for estimating individualized treatment rules with censored data. *Electronic journal of statistics 11*(2), 3927.

Ertefaie, A., D. S. Small, and P. R. Rosenbaum (2018). Quantitative evaluation of the trade-off of strengthened instruments and sample size in observational studies. *Journal of the American Statistical Association 113*(523), 1122–1134.

Fu, Z., Z. Qi, Z. Wang, Z. Yang, Y. Xu, and M. R. Kosorok (2022). Offline reinforcement learning with instrumental variables in confounded markov decision processes. *arXiv preprint arXiv:2209.08666*.

Han, S. (2021). Identification in nonparametric models for dynamic treatment effects. *Journal of Econometrics 225*(2), 132–147.

Imbens, G. and J. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Ishwaran, H., U. B. Kogalur, E. H. Blackstone, and M. S. Lauer (2008). Random survival forest. *Annals of Applied Statistics 2*(3), 841–860.

Jiang, R., W. Lu, R. Song, M. G. Hudgens, and S. Naprvavnik (2017). Doubly robust estimation of optimal treatment regimes for survival data with application to an hiv/aids study. *Annals of Applied Statistics 11*(3), 1763.

Jiang, Z., S. Chen, and P. Ding (2023). An instrumental variable method for point processes: generalized wald estimation based on deconvolution. *Biometrika 110*(4), 989–1008.

Kitagawa, T. and A. Tetenov (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica 86*(2), 591–616.

Kosorok, M. R. and E. B. Laber (2019). Precision medicine. *Annual Review of Statistics and Its Application 6*(1), 263–286.

Liao, L., Z. Fu, Z. Yang, Y. Wang, M. Kolar, and Z. Wang (2021). Instrumental variable value iteration for causal offline reinforcement learning. *arXiv preprint arXiv:2102.09907*.

Liu, Y., Y. Wang, R. M. Kosorok, Y. Zhao, and D. Zeng (2016). Robust hybrid learning for estimating personalized dynamic treatment regimens. *arXiv no. 1611.02314*.

Liu, Z., T. Ye, B. Sun, M. Schooling, and E. Tchetgen Tchetgen (2020). On mendelian randomization mixed-scale treatment effect robust identification (mr misteri) and estimation for causal inference. *arXiv preprint arXiv:2009.14484*.

Miao, W., Z. Geng, and E. J. Tchetgen Tchetgen (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika 105*(4), 987–993.

Michael, H., Y. Cui, S. A. Lorch, and E. J. Tchetgen Tchetgen (2024). Instrumental variable estimation of marginal structural mean models for time-varying treatment. *Journal of the American Statistical Association 119*(546), 1240–1251.

Mo, W., Z. Qi, and Y. Liu (2021). Learning optimal distributionally robust individualized treatment rules. *Journal of the American Statistical Association 116*(534), 659–674.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical*

Society: Series B (Statistical Methodology) 65(2), 331–355.

Park, C., D. B. Richardson, and E. J. Tchetgen Tchetgen (2024). Single proxy control. *Biometrics 80*(2), ujae027.

Pu, H. and B. Zhang (2021). Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. *Journal of the Royal Statistical Society Series B 83*(2), 318–345.

Qi, Z., D. Liu, H. Fu, and Y. Liu (2020). Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *Journal of the American Statistical Association 115*(530), 678–691.

Qian, M. and S. A. Murphy (2011). Performance guarantees for individualized treatment rules. *Annals of statistics 39*(2), 1180–1210.

Qiu, H., M. Carone, E. Sadikova, M. Petukhova, R. C. Kessler, and A. Luedtke (2021). Optimal individualized decision rules using instrumental variable methods (with discussion). *Journal of the American Statistical Association 116*(533), 174–191.

Rubin, D. B. and M. J. van der Laan (2012). Statistical issues and limitations in personalized medicine research with clinical trials. *The International Journal of Biostatistics 8*(1), 18.

Sun, B., Y. Cui, and E. Tchetgen Tchetgen (2022). Selective machine learning of the average treatment effect with an invalid instrumental variable. *Journal of Machine Learning Research 23*(204), 1–40.

Sun, B., Z. Liu, and E. Tchetgen Tchetgen (2023). Semiparametric efficient g-estimation with invalid instrumental variables. *Biometrika 110*(4), 953–971.

Sun, H., B. A. Craig, and L. Zhang (2017). Angle-based multicategory distance-weighted svm. *The Journal of Machine Learning Research 18*(1), 2981–3001.

Sverdrup, E., A. Kanodia, Z. Zhou, S. Athey, and S. Wager (2020). policytree: Policy learning via doubly robust empirical welfare maximization over trees. *Journal of Open Source Software 5*(50), 2232.

Tanser, F., T. Bärnighausen, E. Grapsa, J. Zaidi, and M.-L. Newell (2013). High coverage of art associated with decline in risk of hiv acquisition in rural kwazulu-natal, south africa. *Science 339*(6122), 966–971.

Tanser, F., H.-Y. Kim, A. Vandormael, C. Iwuji, and T. Bärnighausen (2020). Opportunities and challenges in hiv treatment as prevention research: results from the anrs 12249 cluster-randomized trial and associated population cohort. *Current Hiv/Aids Reports 17*, 97–108.

Tchetgen Tchetgen, E., B. Sun, and S. Walter (2021). The genius approach to robust mendelian randomization inference. *Statistical Science 36*(3), 443–464.

Tchetgen Tchetgen, E. J., A. Ying, Y. Cui, X. Shi, and W. Miao (2020). An introduction to proximal causal learning. *arXiv preprint arXiv:2009.10982*.

Wang, Z. and T. A. Louis (2003). Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika 90*(4), 765–775.

Xue, F., Y. Zhang, W. Zhou, H. Fu, and A. Qu (2022). Multicategory angle-based learning for estimating optimal dynamic treatment regimes with censored data. *Journal of the American Statistical Association 117*(539), 1438–1451.

Ye, T., A. Ertefaie, J. Flory, S. Hennessy, and D. S. Small (2023). Instrumented difference-in-differences. *Biometrics 79*(2), 569–581.

Ying, A., Y. Cui, and E. J. T. Tchetgen (2022). Proximal causal inference for marginal counterfactual survival curves. *arXiv preprint arXiv:2204.13144*.

Zhang, B., A. A. Tsiatis, M. Davidian, M. Zhang, and E. Laber (2012). Estimating optimal treatment regimes from a classification perspective. *Stat 1*(1), 103–114.

Zhao, Y., D. Zeng, E. B. Laber, R. Song, M. Yuan, and M. R. Kosorok (2015). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika 102*(1), 151–168.

Zhao, Y., D. Zeng, A. J. Rush, and M. R. Kosorok (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association 107*(499), 1106–1118.

Zhong, Q., J. Mueller, and J.-L. Wang (2022). Deep learning for the partially linear cox model. *The Annals of Statistics 50*(3), 1348–1375.

Zhong, Q., J. W. Mueller, and J.-L. Wang (2021). Deep extended hazard models for survival analysis. *Advances in Neural Information Processing Systems 34*, 15111–15124.

Zhou, X. and M. R. Kosorok (2017). Augmented outcome-weighted learning for optimal treatment regimes. *arXiv preprint arXiv:1711.10654*.

Zhu, R. and M. R. Kosorok (2012). Recursively imputed survival trees. *Journal of the American Statistical Association 107*(497), 331–340.

Zhu, R., Y.-Q. Zhao, G. Chen, S. Ma, and H. Zhao (2017). Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics 73*(2), 391–400.

Zivich, P. N. and A. Breskin (2021). Machine learning for causal inference: on the use of cross-fit estimators. *Epidemiology 32*(3), 393–401.

Zubizarreta, J. R., D. S. Small, N. K. Goyal, S. Lorch, and P. R. Rosenbaum (2013). Stronger instruments via integer programming in an observational study of late preterm birth outcomes. *The Annals of Applied Statistics 7*(1), 25–50.

REFERENCES

Yifan Cui

Center for Data Science, Zhejiang University

E-mail: (cuiyf@zju.edu.cn)

Jianhua Guo

School of Mathematics and Statistics, Beijing Technology and Business University

E-mail: (jhguo@btbu.edu.cn)

Wendong Li

School of Statistics, East China Normal University

E-mail: (wdli@sfs.ecnu.edu.cn)

Frank Tanser

Africa Health Research Institute

E-mail: (ftanser@gmail.com)

Dongdong Xiang

School of Statistics, East China Normal University

E-mail: (ddxiang@sfs.ecnu.edu.cn)