Individualized treatment regimens under correlated data with multiple outcomes

Misha Dolmatov*, Erica Moodie, David Stephens
Department of Epidemiology, Biostatistics and Occupational Health,
McGill University, Montreal, QC, Canada

and

Dipankar Bandyopadhyay
Department of Biostatistics, School of Public Health,
Virginia Commonwealth University, Richmond, VA, USA

Abstract

Precision medicine involves developing individualized treatment regimes (ITRs) which allow for treatment decisions to be tailored to patient characteristics. Naturally, the identification of the optimal regime, that is, the rule which maximizes patient outcomes, is of interest. Several procedures for estimating optimal ITRs from observational data have been proposed; however, relatively few methods exist for estimating optimal ITRs in the presence of competing risks. Previous approaches either target one particular cause of failure, or rely on singly-robust estimators. We propose a novel doubly-robust regression-based method for estimating optimal ITRs which accounts for the uncertainty related to the unobserved cause of failure by averaging over all possible causes, or targeting the most likely cause. Our approach is straightforward to implement, and we demonstrate an extension to incorporate clustering, motivated by the question of for whom kidney transplantation with hepatitis C virus (HCV)-positive donors is safe, using data from the Organ Procurement and Transplantation Network. Our analysis suggests that a large portion of HCV-negative kidney recipients would see their overall survival unchanged if they were instead provided a kidney from an HCV-positive donor. The estimated treatment rules could be used to provide more efficient allocation of HCV-positive kidneys, increasing the donor pool.

Keywords: Accelerated failure time model; Clustered data; Causal inference; Survival analysis; Precision medicine; Kidney transplantation

*The authors gratefully acknowledge the United Network for Organ Sharing for providing the motivating OPTN data, and the context for the work. The work is partially funded by grants R01DE031134 and R21DE031879 from the United States National Institutes of Health.

1 Introduction

The field of personalized medicine involves developing adaptive treatment rules which prescribe the most beneficial treatment based on available patient information. This framework is especially relevant when a given exposure has associated risks, which must be taken into account when assessing the overall benefit of treatment assignment. In such scenarios, patient covariate information can be used to guide decision-making and tailor treatment to an individual's personal level of risk. Covariate-tailored decision rules are referred to as individualized treatment regimes (ITRs). Naturally, the identification of the optimal regime, that is, the rule which maximizes patient outcomes, is of immense interest.

Our motivating question is focused on the problem of kidney transplantation with hepatitis C virus (HCV)-positive donors. Historically, kidney transplants from HCV-positive donors to HCV-negative recipients have been controversial due to concerns of viral transmission (Pereira et al., 1991). However, a recent study using data from the Organ Procurement and Transplant Network (OPTN) registry investigated the feasibility of using kidneys from HCV-positive donors in order to increase the donor pool (Gupta et al., 2017), and showed that HCV-negative recipients with HCV-positive donors had lower graft and patient survival than their counterparts with HCV-negative donors. However, patients with a recipient-donor mismatch in HCV status still had superior survival outcomes when compared to those who remained on the waitlist. These results point towards mismatched transplants being a viable alternative for certain patient subgroups. In other words, if one could identify the patient subgroups that are less at risk of poor outcomes due to the HCV status of their transplant donor, this would yield an easy way to increase the number of available donors while limiting adverse events. We revisit the OPTN dataset, approaching this problem from the ITR perspective in order to identify additional tailoring variables which may permit the use of HCV+ kidneys for safe transplantations.

Estimation of the optimal regime from observational data requires careful consideration of the relationship between the treatment assignment mechanism, patient covariate information, and the study design. Several regression-based approaches for estimating ITRs from observational data have been proposed, including Q-learning (Murphy, 2003), G-estimation (Robins, 2004), and dynamic weighted ordinary least squares (dWOLS; Wallace and Moodie, 2015). Variants of these methods that allow for time-to-event ("survival") outcomes subject to right-censoring have also been suggested. For example, dynamic weighted survival modelling (DWSurv; Simoneau et al., 2020) was developed as an extension to the doubly-robust weighted linear regression approach of dWOLS. The method aims to estimate the optimal regime via weighted accelerated failure time (AFT; Orbe and Núñez-Antón, 2006) models at each stage of treatment with inverse probability of being censored (IPC) weights, combined with balancing weights (such as overlap or inverse probability of treatment weights) to account for non-random censoring and non-randomized treatment assignment.

An important consideration when dealing with survival data is that of competing risks, i.e., events that prevent the occurrence of other types of events (Prentice et al., 1978). For example, in kidney transplantation, if treatment recommendations were made to maximize the time to graft rejection (post transplantation), deaths due to a cause unrelated to graft failure would constitute a competing risk. Despite being of vast clinical importance, research on the topic has been relatively limited in the dynamic treatment regime (DTR) literature. Most of the methods developed to identify competing risk DTRs have revolved around providing novel ways of estimating the Fine and Gray (Fine and Gray, 1999) cumulative incidence function (CIF) for the event type of interest in multi-stage designs (Yavuz et al., 2016; Chen et al., 2018). For instance, Morzywołek et al. (2022) estimates the CIFs associated with a fixed regime via dynamic-regime marginal structural models. In contrast, a recent work (Choi and Cho, 2019)

has investigated the possibility of using AFT models for competing risk problems, allowing for treatment effects to be directly interpreted in terms of the survival time.

The contributions discussed above typically consider each event type separately for DTR estimation. However, since the cause of failure of each subject is unknown at treatment assignment, optimal decision-making requires consideration of all competing risks in order to maximize overall event-free survival. One possible solution, based on Q-learning (Clifton and Laber, 2020), estimates a set-valued DTR that results in a set of recommended treatments that are not unfavourable according to any of the competing outcomes (Laber et al., 2014). As Q-learning is a singly-robust procedure, this approach is inherently more sensitive to model misspecification. Finally, none of the methods proposed thus far allow for clustered data (subjects within centers), which are of particular interest in precision medicine, with large multi-center studies often being required for sufficient power to investigate treatment effect heterogeneity.

This article extends the weighted AFT methodology of DWSurv to account for competing risks and clustering. Our method, based on a doubly-robust implementation of generalized estimating equations (GEE), permits consistent estimation of parameters of interest while allowing misspecification of a subset of nuisance models employed to address confounding, loss to follow up, or the median survival time itself. The estimated optimal treatment rule is expressed as a function of the optimal ITRs for each competing cause of failure, in order to account for the uncertainty regarding the true cause of failure at treatment allocation. Our flexible approach allows for different functions to be considered and compared based on the resulting treatment rule and its ability to maximize the overall survival time. The proposed extensions provide straightforward, regression-based methods for determining optimal ITRs in the presence of competing risks.

The rest of the paper is organized as follows. Starting with definitions and assumptions,

Section 2 presents the development of the optimal ITR, related estimation and inference, and regime evaluation. Section 3 explores numerical characteristics and behaviour of the proposed ITRs in a number of finite sample simulation settings. The proposed methodology is illustrated via application to the motivating OPTN kidney transplant allocation data in Section 4. Finally, Section 5 concludes, with a discussion. Additional materials, consisting of plots and tables from simulation studies and real data illustration, and some theoretical details are relegated to Appendices A-C.

2 Methodology

2.1 Definitions and assumptions

To each subject, we associate the vector of observable quantities $(\boldsymbol{X}, A, \Delta, C, K, R, T)$. Within the competing risk framework, let $K \in \{1, \ldots, \kappa\}$ be the cause of failure indicator, and T, the associated failure time. Without loss of generality, we assume that a longer time to event (failure time) is more clinically desirable. Let $A \in \mathcal{A}$ represent the treatment assignment and $\boldsymbol{X} \in \mathcal{X}$ the vector of measured pre-treatment covariates for each subject. Denote by Δ, the event indicator and by C, the censoring time, where $\Delta = 1$ if an event of any type was observed and 0 otherwise. Note that neither K nor T are observed when $\Delta = 0$. Finally, let $R \in \{1, \ldots, r\}$ denote the cluster membership indicator. Although not immediately relevant for defining the optimal treatment rule, accounting for clustering in the outcome variable is important for efficient parameter estimation. This will be discussed further in Subsections 2.2 and 2.3. The observed data is the collection of random variates $\{(\boldsymbol{X}_i, A_i, \Delta_i, (1-\Delta_i)C_i, \Delta_i K_i, R_i, \Delta_i T_i)\}_{i=1}^n$. We define $T(a, k)$ as the counterfactual survival time for an individual if, possibly contrary to the fact, they had been assigned treatment a and failed from cause k. An ITR, in our

case, is a function $d : \mathcal{X} \to \mathcal{A}$ that maps a subject's covariate information to the set of possible treatments. Similarly, we can define $T(d, k)$ as the potential outcome of an individual with cause of failure k and treatment assigned according to d. For causal identification and estimation of these potential outcomes, we make the following assumptions:

1. Stable unit treatment value (SUTVA): The failure time of a subject is independent of other patients' treatment assignment (Rubin, 1980).

2. Consistency: Observed and counterfactual outcomes agree for $A = a$ and $K = k$, i.e., $T = \sum_{a \in \mathcal{A}, k=1,\ldots,\kappa} \mathbb{1}_a(A)\mathbb{1}_k(K)T(a, k)$.

3. No unmeasured confounders (NUC) (Gill and Robins, 2001): $T(a, k) \perp (A, K)|\boldsymbol{X}$ $\forall a \in \mathcal{A},\ \forall k = 1, \ldots, \kappa$.

4. Coarsening at random (Gill et al., 1997): $T(a, k) \perp \Delta|\boldsymbol{X}\ \forall a \in \mathcal{A},\ \forall k = 1, \ldots, \kappa$.

5. Conditional independence of failure type and joint distribution of treatment and censoring mechanisms: $K \perp (A, \Delta)|\boldsymbol{X}$.

The fundamental problem of ITR estimation in the competing risk setting is that the true future cause of failure for any patient is unknown at treatment assignment. As a result, optimal decision-making must account for this uncertainty as well as treatment effect heterogeneity. If the true cause of failure was known at treatment assignment, i.e., K was among the vector of measured covariates for each subject, the optimal ITR could be directly specified as

$$d_{\text{o}}^{\text{opt}}(\boldsymbol{X}, K) = \arg\max_{d} \mathbb{E}[f(T(d, K)) - \zeta(d, \boldsymbol{X})] \tag{1}$$

for some fixed monotonically increasing function $f : \mathbb{R}_+ \to \mathbb{R}$ and some known cost function $\zeta : \mathcal{A} \times \mathcal{X} \to \mathbb{R}_+$, which represents some measure of financial or other cost associated to the

allocation of treatment on the transformed survival scale. We will refer to this regime as the *oracle* regime. The cost function can also be used to specify a minimal effectiveness threshold, with treatment only being assigned when the treatment effect is larger than this threshold. This interpretation is made more concrete in Subsection 2.2. Alternatively, in resource-limited settings, ζ can be specified as a data-dependent constant which determines the proportion of the study population that will receive treatment (Luedtke and van der Laan, 2016).

A natural definition of the optimal ITR for unknown K is one that marginalizes over the distribution of the cause of failure K in equation (1):

$$\begin{aligned} d_{\mathrm{w}}^{\mathrm{opt}}(\boldsymbol{X}) &= \arg\max_d \mathbb{E}\Big[f(T(d,K)) - \zeta(d,\boldsymbol{X})\Big], \\ &= \arg\max_d \mathbb{E}\Big[\mathbb{E}[f(T(d,K))|\boldsymbol{X}] - \zeta(d,\boldsymbol{X})\Big], \\ &= \arg\max_d \mathbb{E}\left[\sum_{k=1}^{\kappa} P(K=k|\boldsymbol{X})\mathbb{E}[f(T(d,k))|\boldsymbol{X}] - \zeta(d,\boldsymbol{X})\right]. \end{aligned}$$

In the DTR literature, the optimal treatment rule is often expressed in terms of causal contrasts with respect to some reference treatment $A = 0$, leading to the following definition:

$$\gamma_k(\boldsymbol{x}, a) = \mathbb{E}[f(T(a,k)) - f(T(0,k))|\boldsymbol{X} = \boldsymbol{x}], \quad k = 1, \ldots, \kappa \text{ and } \boldsymbol{x} \in \mathcal{X}.$$

This quantity is called the *blip* function and represents a causal contrast (often a difference) in outcomes — here in the transformed survival times — of an individual with covariates $\boldsymbol{X} = \boldsymbol{x}$ when comparing treatment $A = a$ with the reference treatment (Robins, 2004). The expected counterfactual survival time can then be written in terms of the blip and a second component which does not depend on treatment:

$$\begin{aligned} \mathbb{E}[f(T(a,k))|\boldsymbol{X} = \boldsymbol{x}] &= \mathbb{E}[f(T(0,k))|\boldsymbol{X} = \boldsymbol{x}] + \mathbb{E}[f(T(a,k)) - f(T(0,k))|\boldsymbol{X} = \boldsymbol{x}], \\ &= m_k(\boldsymbol{x}) + \gamma_k(\boldsymbol{x}, a), \quad k = 1, \ldots, \kappa. \end{aligned}$$

The first term $m_k(\boldsymbol{x})$ is designated the *treatment-free* model as it only depends on the covariates $\boldsymbol{X}$ and the chosen reference treatment, but not a. Thus, the optimal ITR can be expressed in

terms of the blip functions for each failure type as

$$d_{\mathrm{w}}^{\mathrm{opt}}(\boldsymbol{X}) = \arg\max_{d} \mathbb{E}\left[\sum_{k=1}^{\kappa} \varphi_k(\boldsymbol{X})\gamma_k(\boldsymbol{X}, d) - \zeta(d, \boldsymbol{X})\right], \tag{2}$$

where, $\varphi_k(\boldsymbol{X}) = P(K = k|\boldsymbol{X})$. We will refer to this ITR as the *weighted* rule. As a result, determining the optimal regime relies on the identification of the blip functions for all K failure types. Under assumptions 2 and 3, the blip functions can be written in the following form:

$$\begin{aligned}\gamma_k(\boldsymbol{x}, a) &= \mathbb{E}[f(T(a,k)) - f(T(0,k))|\boldsymbol{X} = \boldsymbol{x}],\\ &= \mathbb{E}[f(T)|\boldsymbol{X} = \boldsymbol{x}, A = a, K = k] - \mathbb{E}[f(T)|\boldsymbol{X} = \boldsymbol{x}, A = 0, K = k] \quad k = 1, \ldots, \kappa.\end{aligned}$$

Given that that this contrast is expressed as the difference in conditional means, a natural approach would be to posit regression models for the survival times associated to each failure type, and estimate the blip directly. However, outcome regression-based methods are known to be sensitive to model misspecification; this considerable weakness will be addressed in Subection 2.3 when discussing doubly-robust estimation.

Equation (1) can be modified to obtain other definitions of an optimal ITR. For instance, we can consider a rule which optimally treats the most probable cause:

$$\begin{aligned}d_{\mathrm{g}}^{\mathrm{opt}}(\boldsymbol{X}) &= \arg\max_{d} \mathbb{E}[f(T(d, \arg\max_{k} \varphi_k(\boldsymbol{X}))) - \zeta(d, \boldsymbol{X})],\\ &\quad - \arg\max_{d} \mathbb{E}[\gamma_{k^*}(\boldsymbol{X}, d) - \zeta(d, \boldsymbol{X})], \quad k^* = \arg\max_{k} \varphi_k(\boldsymbol{X}),\end{aligned}$$

which we designate the *greedy* rule. Alternatively, the functions $\varphi_k(\boldsymbol{x})$ can be replaced by fixed constants which reflect the beliefs of the analyst concerning the relative importance of each cause in the decision making process.

2.2 Accelerated failure time specification

A common choice for f is $f(t) = \log t$, resulting in an AFT model associated to each cause. For the remaining sections, we assume binary treatment, i.e., $\mathcal{A} = \{0, 1\}$. Suppose the data

are comprised of r clusters, indexed by $i = 1, \ldots, r$, each of size r_i. Assume that the number and size of these clusters is fixed. For each cluster $i = 1, \ldots, r$, we consider the following specification for the mean of the response vector $\mathbf{T_i}$:

$$\log \mathbf{T_i} | \mathbf{K_i} = \mathbf{k_i} \sim \mathbf{X}_{\boldsymbol{\beta_k}} \boldsymbol{\beta_k} + \mathbf{A}\mathbf{X}_{\boldsymbol{\psi_k}} \boldsymbol{\psi_k} + \boldsymbol{\varepsilon_i},$$

where, $\mathbb{E}[\boldsymbol{\varepsilon_i}] = \mathbf{0}$ and $\text{var}[\boldsymbol{\varepsilon_i}] = \boldsymbol{V_i}$. The matrix $\mathbf{X}_{\boldsymbol{\psi_k}}$ contains all components of $\boldsymbol{X}$ that modify the effect of treatment for the k-th failure type, augmented with a leading column of 1s in order to account for the main effect of treatment. The matrix $\mathbf{X}_{\boldsymbol{\beta_k}}$ contains the same features as $\mathbf{X}_{\boldsymbol{\psi_k}}$, including the leading column of 1s, along with relevant confounders and other variables predictive of the outcome in the absence of treatment. Using the same identification arguments as in the previous subsection, we can connect this parametrization to the causal quantities defined above:

$$\begin{aligned} \mathbb{E}[\log T | \boldsymbol{X} = \boldsymbol{x}, A = a, K = k] &= m_k(\boldsymbol{x}_{\boldsymbol{\beta_k}}; \boldsymbol{\beta_k}) + \gamma_k(\boldsymbol{x}_{\boldsymbol{\psi_k}}, a; \boldsymbol{\psi_k}), \\ &= \boldsymbol{x}_{\boldsymbol{\beta_k}} \boldsymbol{\beta_k} + a \boldsymbol{x}_{\boldsymbol{\psi_k}} \boldsymbol{\psi_k}, \quad k = 1, \ldots, K. \end{aligned}$$

This allows us to directly write the regime $d_{\text{w}}^{\text{opt}}$ in terms of the parameters $\boldsymbol{\psi_k}$:

$$d_{\text{w}}^{\text{opt}}(\boldsymbol{X}) = \arg\max_d \mathbb{E}\left[d(\boldsymbol{X}) \sum_{k=1}^{K} \varphi_k(\boldsymbol{X}) \boldsymbol{X}_{\boldsymbol{\psi_k}} \boldsymbol{\psi_k} - \zeta(\boldsymbol{X}, d) \right]$$

Since the treatment is binary, we can write the cost function as $\zeta(\boldsymbol{X}, d) = \zeta_0(\boldsymbol{X}) + d(\boldsymbol{X})\zeta_1(\boldsymbol{X})$. As a result, $d_{\text{w}}^{\text{opt}}$ can be directly ascertained as

$$d_{\text{w}}(\boldsymbol{x}) = \mathbb{1}\left(\sum_{k=1}^{\kappa} \varphi_k(\boldsymbol{x}) \boldsymbol{x}_{\boldsymbol{\psi_k}} \boldsymbol{\psi_k} > \zeta_1(\boldsymbol{x}) \right), \quad \boldsymbol{x} \in \mathcal{X}.$$

The weighted rule has an intuitive interpretation: it prescribes treatment as long as the treatment has an overall effect larger than some minimal threshold $\zeta_1(\boldsymbol{x})$, where this global effect is a convex combination of the effects of treatment under different failure types, with weights

corresponding to the probability of the occurrence of each cause of failure. As a result, if treatment dramatically decreases survival times for any given failure type, the rule might possibly recommend withholding treatment, even if failures of that type are rare and treatment has a protective effect with respect to the other failure types; in that case, the risk outweighs the benefits.

Following the derivations above, we can write the oracle and greedy regimes in terms of the blip parameters:

$$d_{\mathrm{o}}(\boldsymbol{x}, k) = \mathbb{1}\left(\boldsymbol{x}_{\boldsymbol{\psi}_k}\boldsymbol{\psi}_k > \zeta_1(\boldsymbol{x})\right),$$

$$d_{\mathrm{g}}(\boldsymbol{x}) = \mathbb{1}\left(\boldsymbol{x}_{\boldsymbol{\psi}_{k^*}}\boldsymbol{\psi}_{k^*} > \zeta_1(\boldsymbol{x})\right), \quad k^* = \arg\max_k \varphi_k(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{X}.$$

All treatment rules d defined above can be written as $d(\boldsymbol{x}) = \mathbb{1}(\mathcal{B}_d(\boldsymbol{x}) > \zeta_1(\boldsymbol{x}))$ for some function $\mathcal{B}_d : \mathcal{X} \to \mathbb{R}$, which we term the *benefit* of the ITR d. In particular,

$$\mathcal{B}_{d_{\mathrm{w}}}(\boldsymbol{x}) = \sum_{k=1}^{\kappa} \varphi_k(\boldsymbol{x})\boldsymbol{x}_{\boldsymbol{\psi}_{\boldsymbol{k}}}\boldsymbol{\psi}_{\boldsymbol{k}},$$

$$\mathcal{B}_{d_{\mathrm{g}}}(\boldsymbol{x}) = \boldsymbol{x}_{\boldsymbol{\psi}_{k^*}}\boldsymbol{\psi}_{k^*}, \quad k^* = \arg\max_k \varphi_k(\boldsymbol{x}),$$

$$\mathcal{B}_{d_{\mathrm{o}}}(\boldsymbol{x}, k) = \boldsymbol{x}_{\boldsymbol{\psi}_k}\boldsymbol{\psi}_k.$$

The benefit for the oracle regime depends on the true cause of failure, k.

2.3 Estimation and inference

In Subsection 2.2, we derived the form of the optimal treatment rule in terms of the blip parameters. Consequently, determining the optimal ITR relies on consistent estimation of these parameters. For this purpose, we consider DWSurv, which provides doubly-robust semiparametric estimation of blip parameters in AFT models with weights accounting for non-randomized treatment assignment and covariate-dependent right-censoring (Simoneau et al., 2020). The

weighted GEE approach used in DWSurv can easily be modified to allow for non-independence of observations by specifying a working covariance matrix for each cluster (Liang and Zeger, 1986).

We propose estimating all κ sets of blip parameters by applying DWSurv to the observations of each of the failure types separately. The estimation procedure for an arbitrary treatment rule $d(\boldsymbol{x};\boldsymbol{\psi})$ is summarized by the following algorithm:

1. Posit models for the treatment, censoring, and cause of failure mechanisms: $P(A = 1|\boldsymbol{X} = \boldsymbol{x};\boldsymbol{\alpha})$, $P(\Delta = 1|A = a, \boldsymbol{X} = \boldsymbol{x};\boldsymbol{\xi})$, and $P(K = k|\boldsymbol{X} = \boldsymbol{x};\boldsymbol{\eta})$ respectively. Obtain estimates $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\xi}}$, and $\hat{\boldsymbol{\eta}}$ from the observed data.

2. For failure types $k = 1, \ldots, \kappa$,

   (a) Propose a linear model for the outcome of the form $\mu = \mathbb{E}[\log T|\boldsymbol{X} = \boldsymbol{x}, A = a, K = k;\boldsymbol{\beta_k}, \boldsymbol{\psi_k}] = \boldsymbol{x}_{\boldsymbol{\beta_k}}\boldsymbol{\beta_k} + a\boldsymbol{x}_{\boldsymbol{\psi_k}}\boldsymbol{\psi_k}$, and propose a working covariance matrix $\mathbf{V}_i(\boldsymbol{\lambda})$ for all clusters $i = 1, \ldots, r$.

   (b) Choose a weight function $w_k(\delta, a, \boldsymbol{x};\alpha)$ and compute the associated weights $\hat{w}_k$ with the estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\eta}}$. See below for details.

   (c) Given initial estimate $\hat{\boldsymbol{\lambda}}^{(0)}$, set $t = 0$ and alternate between the following two steps until convergence is achieved:

      i. Obtain estimate of $\hat{\boldsymbol{\theta}}^{(t+1)} = (\hat{\boldsymbol{\beta}_{\boldsymbol{k}}}^{(t+1)}, \hat{\boldsymbol{\psi}_{\boldsymbol{k}}}^{(t+1)})$ by solving the weighted estimating equation

$$\sum_{i=1}^{r} \mathbf{D}_i^\top [\mathbf{V}_i(\hat{\boldsymbol{\lambda}}^{(t)})]^{-1} \mathbf{W_i}\Big(\log(\mathbf{T}_i) - \boldsymbol{\mu}_i\Big) = 0, \tag{3}$$

      where $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\theta}} = [\mathbf{X}^{(i)}_{\boldsymbol{\beta_k}} \;\; \mathbf{A}^{(i)}\mathbf{X}^{(i)}_{\boldsymbol{\psi_k}}]$, $\mathbf{W}_i = \mathrm{diag}\Big(\delta_{ij}\hat{w}_{ijk}\Big)_{j=1,\ldots,r_i}$.

ii. Estimate $\hat{\boldsymbol{\lambda}}^{(t+1)}$ via method of moments based on the estimated residuals $\hat{\boldsymbol{e}}_i(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}^{(t+1)}, \hat{\boldsymbol{\psi}}_{\boldsymbol{k}}^{(t+1)}) = \log(\mathbf{T}_i) - \hat{\boldsymbol{\mu}}_i^{(t+1)}$.

3. Obtain the estimated optimal treatment rule $d(\boldsymbol{x}; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\eta}})$ from the final estimates of the blip parameters and the parameters of the failure type model.

The weights are chosen to satisfy the balance condition stated in the following theorem (proof in Appendix C).

Theorem 1 *Under assumptions* 1-5 *(see Subsection 2.1), solving the weighted GEE (3) yields consistent estimators for* $\{\boldsymbol{\psi}_k\}_{k=1}^{\kappa}$ *if the weights satisfy the following balancing property:*

$$\begin{aligned}[1 - c(0, \boldsymbol{x})][1 - b(\boldsymbol{x})]w_k(0, 0, \boldsymbol{x}) &= c(0, \boldsymbol{x})[1 - b(\boldsymbol{x})]w_k(1, 0, \boldsymbol{x}) \\ &= [1 - c(1, \boldsymbol{x})]b(\boldsymbol{x})w_k(1, 0, \boldsymbol{x}) \\ &= c(1, \boldsymbol{x})b(\boldsymbol{x})w_k(1, 1, \boldsymbol{x}), \qquad \textit{for } k = 1, \ldots, \kappa,\end{aligned}$$

where $b(\boldsymbol{x}) = P(A = 1|\boldsymbol{X} = \boldsymbol{x})$ *and* $c(a, \boldsymbol{x}) = P(\Delta = 1|A = a, \boldsymbol{X} = \boldsymbol{x})$.

For example, weights of the form

$$w(\delta, a, x) = \frac{|a - P(A = 1|\boldsymbol{X} = \boldsymbol{x})|}{P(\Delta = \delta|A = a, \boldsymbol{X} = \boldsymbol{x})} \tag{4}$$

satisfy the above equation (Simoneau et al., 2020).

Estimating the blip parameters $\boldsymbol{\psi}_{\boldsymbol{k}}$ via DWSurv offers the advantage of double-robustness. For each of the failure types $k = 1, \ldots, \kappa$, under correct specification of the blip function γ_k, it suffices to correctly specify either the treatment-free model or the weights in order to obtain consistent estimation of $\boldsymbol{\psi}_{\boldsymbol{k}}$. We note that correctly specifying the weights requires correct models for both the treatment and censoring mechanisms. Furthermore, under standard regularity conditions for GEEs, the procedure outlined above is robust to misspecification of the

working correlation structure: as long as either the outcome model or the weighting models are correctly specified, we have consistent estimation of the blip parameters, albeit with a possible loss of efficiency (Liang and Zeger, 1986).

If the nuisance models are estimated by maximum likelihood, the consistency and asymptotic normality of the blip parameter estimators follows directly from standard results on joint estimating equations (van der Vaart, 1998). Otherwise, similar conclusions hold under additional assumptions regarding the convergence of nuisance parameter estimators (Yuan and Jennrich, 2000). The asymptotic variance of $\hat{\boldsymbol{\psi}}$ can be obtained by considering a Taylor expansion of the estimating equation presented in step c(i) about the limiting distributions of the nuisance parameters (Simoneau et al., 2020). However, due to the large number of nuisance parameters, deriving an explicit formula is very cumbersome and consequently, of limited practical use. The nonparametric cluster bootstrap offers a conceptually simpler alternative to variance estimation at the cost of a larger computational burden (Davison and Hinkley, 1997). Alternatively, the sandwich errors that ignore the estimation of the weights may often provide reasonable performance (Lunceford and Davidian, 2004).

2.4 Regime evaluation

Although the treatment rules presented above target different criteria, a particular ITR might be preferable in a given situation based on overall in-sample performance. For a given regime d, we consider two main measures of performance, namely the proportion of optimal treatment (POT) and the value. These are defined as follows:

$$\mathrm{POT}(d) = \mathbb{E}[\mathbb{1}(d(\boldsymbol{X}) = d_{\mathrm{o}}(\boldsymbol{X}, K))],$$

$$\mathrm{V}(d) = \mathbb{E}\left[\log T(d, K)\right].$$

The POT for a decision rule d measures the proportion of individuals for whom the treatment allocated under d is equal to the treatment prescribed by the oracle regime. The value of d is the average survival time of a subject within the study population whose treatment is assigned according to d. The ideal ITR targets the specified criteria while also maintaining reasonably high POT and value. The following two unbiased estimating equations in p and μ yields estimators of the POT and value of an ITR d, respectively:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{P(\Delta=1|\boldsymbol{X}_i,A_i)}\Big(\mathbb{1}(d(\boldsymbol{X}_i)=d_{\text{o}}(\boldsymbol{X}_i,K_i))-p\Big)=0,$$
$$\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{P(\Delta=1|\boldsymbol{X}_i,A_i)}\Big(\log T_i+(d-A_i)\boldsymbol{X}_{\psi_{K_i}}\boldsymbol{\psi}_{K_i}-\mu\Big)=0.$$

Proofs of unbiasedness can be found in Appendix C. Since both the true censoring probabilities and the true values of the blip parameters (and thus, the oracle regime) are unknown, they must be estimated from the data. Estimates are readily available following the estimation procedure detailed in Subsection 2.3. Under correct specification of the censoring model and of the blip functions for each cause, as well as standard regularity conditions, the POT and value estimators are consistent and asymptotically normal. In practice, the estimating equations above can be solved by fitting intercept-only linear models to the outcomes $\mathbb{1}(d(\boldsymbol{X}_i)=\hat{d}_{\text{o}}(\boldsymbol{X}_i,K_i))$ and $\log T_i+(d-A_i)\boldsymbol{X}_i\hat{\boldsymbol{\psi}}_{K_i}$, with estimated inverse probability of censoring weights $\hat{w}_i=\Delta_i/\hat{P}(\Delta=1|\boldsymbol{X}_i,A_i)$. The following Subsection highlights the connection between ITR estimation with competing risks and ITR estimation for outcomes generated by a finite mixture model.

2.5 Connection to finite mixture models ITRs

We consider a binary treatment, continuous outcome setting without any censoring. Let $(Y,\boldsymbol{X},A)$ denote the triplet of observed variables, with Y the outcome variable. Suppose

Y is generated from a finite mixture model with κ components, each with the same linear decomposition as in Subsection 2.2. In particular, for $k = 1, \ldots, \kappa$, we have that the conditional distribution of Y given $K = k$ is

$$Y|K = k \sim \boldsymbol{X}_{\beta_k}\boldsymbol{\beta}_k + A\boldsymbol{X}_{\psi_k}\boldsymbol{\psi_k} + \boldsymbol{\varepsilon},$$

where, K denotes the latent mixture component indicator, and $\mathbb{E}[\boldsymbol{\varepsilon}] = 0$.

Let $Y(d, k)$ denote the counterfactual for a subject with treatment assigned via regime d and whose outcome was generated according to the k-th mixture component. Assuming the equivalent versions of Assumptions 1, 2, 3, and 5, we can derive ITRs which take into account the uncertainty associated with the unknown mixture membership indicator K:

$$d_w = \arg\max_d \mathbb{E}[Y(d, K)] = \mathbb{1}\left(\sum_{k=1}^{\kappa} \varphi_k(\boldsymbol{x})\boldsymbol{x}_{\boldsymbol{\psi_k}}\boldsymbol{\psi_k} > 0\right),$$

$$d_g = \arg\max_d \mathbb{E}[Y(d, \arg\max_k \varphi_k(\boldsymbol{X}))] = \mathbb{1}\left(\boldsymbol{x}_{\boldsymbol{\psi}_{k^*}}\boldsymbol{\psi}_{k^*} > 0\right), \quad k^* = \arg\max_k \varphi_k(\boldsymbol{x}),$$

where, all quantities are defined as before. These ITRs have the same interpretations as those introduced in Subsection 2.1; the weighted regime averages over all mixture components while the greedy regime targets the most likely component of the mixture distribution. However, in the finite mixture model case, K is not observed and thus blips cannot be estimated by considering observations from each mixture component separately as in the competing risk setting. The expectation-maximization (EM) algorithm, Bayesian data augmentation, or other standard estimation methods for finite mixture models can instead be used to estimate the blips under missingness of the mixture membership indicator (McLachlan et al., 2019). However, these approaches do not possess the same robustness properties as the method detailed above.

3 Simulations

We conducted several simulation studies to evaluate the finite-sample properties of the proposed estimation procedure and to compare the different treatment rules presented in Subsection 2.2. As in Subsection 2.3, we assume binary treatment; without loss of generality, we also assume $\zeta(\boldsymbol{x}) = 0$, i.e., no costs associated with treatment allocation.

In the simulations, blip parameters were estimated from a training set, with the resulting ITRs then evaluated on a fixed independent uncensored test set under the same data-generating mechanism. Default samples sizes of $n_{\text{train}} = 1000$ and $n_{\text{test}} = 10,000$ were used for the training and testing sets, respectively, unless specified otherwise. The variability of the estimating procedure was assessed by repeating parameter estimation over ($n_{\text{rep}} = 1000$) replicate training sets. Estimated standard errors (SE) and biases were reported for all blip parameter estimators. The various ITRs defined above were compared based on their proportion of optimal treatment and their value, as defined in Subsection 2.4. In this case, the expectation is taken with respect to the test set over all replications. Since the test set is uncensored, the POT and value were estimated via sample averages.

For the sake of comparison, we also define the *uniform* regime, which allocates treatment by sampling uniformly on $\mathcal{A} = \{0, 1\}$: $d_{\text{u}}(\boldsymbol{x}) \sim$ Bernoulli(0.5). Across all simulations scenarios, the estimated weighted and greedy ITRs were compared to the oracle and uniform ITRs, which served as benchmark regimes. The oracle regime was defined in terms of the true data-generating values of the blip parameters in order to provide an upper bound on overall performance.

3.1 Data generating mechanism

Let $\text{expit}(v) = \exp(v)/(1+\exp(v))$ for $v \in \mathbb{R}$. To allow for dependence between observations, survival times for each individual were generated via an AFT model with a random intercept, corresponding to an exchangeable within-cluster correlation structure on the log-survival scale. For cluster $i = 1, \ldots, r$ and subject $j = 1, \ldots, r_i$, we have the following model specification:

$$
\begin{aligned}
X_{ij1} \sim \mathcal{N}(0,1), \quad A_{ij} &\sim \text{Bernoulli}\Big(\text{expit}(0.5 + X_{ij1} + X_{ij2})\Big), \\
X_{ij2} \sim \mathcal{N}(0,4), \quad \Delta_{ij} &\sim \text{Bernoulli}\Big(\text{expit}(\delta_0 - X_{ij1} - 0.3X_{ij2})\Big), \\
K_{ij} &\sim \text{Bernoulli}\Big(\text{expit}(0.5 + X_{i1})\Big) + 1, \\
U_i \sim \mathcal{N}(0,\tau^2), \quad \varepsilon_{ij1}, \varepsilon_{ij2} &\sim \mathcal{N}(0,\sigma^2), \\
T_{ij}|K_{ij} = 1 &\sim \exp\Big(1 + 0.5X_{ij1} - 0.3X_{ij2} + U_i + A_{ij}(\psi_{11} + \psi_{12}X_{i1}) + \varepsilon_{ij1}\Big), \\
T_{ij}|K_{ij} = 2 &\sim \exp\Big(2 - 0.1X_{ij1} + 0.2X_{ij2} + U_i + A_{ij}(\psi_{21} + \psi_{22}X_{i1}) + \varepsilon_{ij2}\Big).
\end{aligned}
$$

Cluster membership was determined by drawing from an independent uniform distribution on $\{1, \ldots, r\}$. The following default values were used for all simulation settings, unless specified otherwise: $r = 50$, $\delta_0 = 1.73$, $\tau^2 + \sigma^2 = 0.5$ and $\text{ICC} = \tau^2/\sigma^2 + \tau^2 = 0.5$.

This data generating mechanism yields a censoring rate of approximately 25%, with 40% of individuals failing from cause 1. Estimation for the weighting models and the cause of failure model was accomplished using logistic regression, while estimation for the AFT-GEE model was done via the `geem` package in R. Weights as in (4) were used for fitting the weighted GEE. Unless otherwise stated, the exchangeable structure was used for the working covariance matrix associated to each cluster. Both weighting models and outcome models were correctly specified for all simulation scenarios except for those in Subsection 3.2. The models for the cause of failure K were correctly specified for all simulations. Additional simulations that explore the sensitivity of the proposed estimation procedure to variations in the data generating process

and other modelling choices can be found in Appendix A.

3.2 Parameter estimation

The first set of simulations showcases the consistency and double robustness of the estimation procedure outlined in Subsection 2.3. For this purpose, we consider two baseline settings, one with a moderate sample size, and one with a large sample size. For both settings, we conducted parameter estimation under four different model specifications:

(i) All models misspecified:

$$P(A_{ij} = 1) = \text{logit}(\alpha_1 + \alpha_2 X_{ij1}), \quad P(\Delta_{ij} = 1) = \text{logit}(\alpha_1^* + \alpha_2^* X_{ij1}),$$
$$\mathbb{E}[\log T | \boldsymbol{X}, A = a, K = k] = \beta_{k1} + \beta_{k2} X_{ij1} + a(\psi_{k1} + \psi_{k2} X_{i1}), \quad k = 1, 2.$$

(ii) Outcome model correct, weighting models incorrect:

$$P(A_{ij} = 1) = \text{logit}(\alpha_1 + \alpha_2 X_{ij1}), \quad P(\Delta_{ij} = 1) = \text{logit}(\alpha_1^* + \alpha_2^* X_{ij1}),$$
$$\mathbb{E}[\log T | \boldsymbol{X}, A = a, K = k] = \beta_{k1} + \beta_{k2} X_{ij1} + \beta_{k3} X_{ij2} + a(\psi_{k1} + \psi_{k2} X_{i1}), \quad k = 1, 2.$$

(iii) Weighting models correct, outcome model incorrect:

$$P(A_{ij} = 1) = \text{logit}(\alpha_1 + \alpha_2 X_{ij1} + \alpha_3 X_{ij2}), \; P(\Delta_{ij} = 1) = \text{logit}(\alpha_1^* + \alpha_2^* X_{ij1} + \alpha_3^* X_{ij2}),$$
$$\mathbb{E}[\log T | \boldsymbol{X}, A = a, K = k] = \beta_{k1} + \beta_{k2} X_{ij1} + a(\psi_{k1} + \psi_{k2} X_{i1}), \quad k = 1, 2.$$

(iv) All models correctly specified:

$$P(A_{ij} = 1) = \text{logit}(\alpha_1 + \alpha_2 X_{ij1} + \alpha_3 X_{ij2}), \; P(\Delta_{ij} = 1) = \text{logit}(\alpha_1^* + \alpha_2^* X_{ij1} + \alpha_3^* X_{ij2}),$$
$$\mathbb{E}[\log T | \boldsymbol{X}, A = a, K = k] = \beta_{k1} + \beta_{k2} X_{ij1} + \beta_{k3} X_{ij2} + a(\psi_{k1} + \psi_{k2} X_{i1}), \quad k = 1, 2.$$

Data-generating parameter values for setting 1 were (a) $(\psi_{11}, \psi_{12}) = (0.2, -0.2)$, $(\psi_{21}, \psi_{22}) = (0.2, 0.2)$, with $n_{\text{train}} = 1000$ and $r = 50$, and (b) $\boldsymbol{\psi}$ as before, with $n_{\text{train}} = 5000$ and $r = 250$.

Table 1 displays root-n adjusted biases and standard errors of the blip parameter estimators across the four model specifications and sample sizes. Figure 1 shows the empirical distributions of the blip parameter estimates across replications for $n_{\text{train}} = 1000$. The figure for the larger sample size can be found in Appendix A. As expected, we have consistent estimation of the blip parameters when either the treatment-free or weighting models are correctly specified. However, misspecifying the treatment-free component incurs a larger penalty in terms of efficiency than misspecifying the weighting models, as is evidenced by the larger variances and small sample bias of the estimators under model (iii), as opposed to those under model (ii).

Table 1 shows the estimated POTs and values for the estimated weighted and greedy ITRs as well as the values for the uniform and oracle regimes across models for both sample sizes. For these data generating mechanisms, models (ii)-(iv) all behave very similarly in terms of POT and value, with values that are very close to that of the oracle regime. As expected, model (i), for which the blip estimators are inconsistent, performs poorly.

In Figure 2, we present the median estimated benefit of the greedy ITR and the weighted ITR for subjects in the test set, along with pointwise 95% CIs, for model (i) in setting 1(a). The plot is stratified by true cause of failure and subjects are ordered by their benefit in increasing order. Within the same figure, we plot the benefit for the oracle regime, obtained from the true value of the blip parameters. Interpreting the benefit plots is relatively straightforward: a given ITR d prescribes treatment whenever $\mathcal{B}_d(\boldsymbol{x}) > 0$, i.e., when the plotted line is above the horizontal line at 0. An estimated ITR prescribes the wrong treatment whenever its benefit curve has the opposite sign of the benefit of the oracle regime, i.e., when the black and red lines are on opposite sides of the horizontal line at 0. Generally, we expect a given ITR to perform

better in terms of POT and value, if its estimated benefit function is closer to the benefit of the oracle regime. This can be easily deduced from the proximity of the lines in the benefit plot. As previously noted, model (i) performs worse than model (iv) in terms of value and POT, with the benefit curves of the estimated regimes being far from the benefit of the oracle regime whenever all models are misspecified. Benefit plots for all other simulation settings can be found in Appendix A.

3.3 Comparison of optimality criteria and resulting rules

The second set of simulations was conducted in order to compare the behaviors of the greedy rule and the weighted rule under different data-generating mechanisms. We considered the settings defined by the following values of the blip parameters, with $n_{\text{train}} = 1000$ and $r = 50$:

$$\text{Setting 2} : (\psi_{11}, \psi_{12}) = (3, -0.5), \quad (\psi_{21}, \psi_{22}) = (-1, 0.2).$$

$$\text{Setting 3} : (\psi_{11}, \psi_{12}) = (-0.5, -0.7), \quad (\psi_{21}, \psi_{22}) = (0.1, 0.08).$$

Table 1 showcases the biases and SEs of the blip parameter estimators for the two settings as well as the values and POTs of the estimated regimes. In setting 2, the weighted regime yields notably a higher value than the greedy regime, while having a moderately smaller value of POT. Since the blip parameters for both causes have relatively similar magnitudes and opposite signs, the weighted ITR's averaging of the two blips allows the benefit to be close to zero when the failure type probability is near 0.5. This allows the weighted regime to switch from one cause to the other much more quickly than the greedy regime, yielding the superior performance that we observe.

In the setting 3, the weighted regime performs slightly better than the greedy regime in terms of value, at the cost of a much lower POT. As a result, the greedy ITR might be preferable in this setting on the basis of overall performance. In this case, the magnitudes of the blip

parameters for the second cause are smaller than those for the first cause, and so averaging the two blips yields a benefit that is much closer to the benefit associated with the first cause. In other words, the weighted regime is "stuck" on the first cause, a problem that the greedy regime does not share.

In the additional simulation settings, we observe that the performance of both the greedy and the weighted ITRs is robust to variations in the data generating mechanism, with performance that matches the results presented above.

3.4 Comparison with standard methods

The final set of simulations compares our proposed method to standard approaches used in the competing risks literature.

3.4.1 Cause-specific ITR

The first approach, used in *cause-specific* analyses, censors the competing event and estimates the optimal ITR based on one cause of interest (Choi and Cho, 2019). Without loss of generality, we assume that the targeted cause of failure is cause 1. The cause-specific ITR is then estimated via DWSurv with outcome variable the time to failure from cause 1, and failures due to cause 2 treated as censored. We generate data according to the data generating mechanism of Subsection 3.1, with the following blip parameters and cause model:

$$\text{Setting 10}: \quad (\psi_{11}, \psi_{12}) = (1, -0.5), \quad (\psi_{21}, \psi_{22}) = (-3, 0.2),$$
$$K_{ij} \sim \text{Bernoulli}\Big(\text{expit}(2.5 + X_{i1})\Big) + 1,$$

which corresponds to a study population with approximately 10% of patients failing from cause 1. In this setting, the two causes of failure have vastly different associated ITRs. Allocating treatment based on only one cause of failure can then yield disastrous performance when the

treatment is harmful for those failing from the second cause, especially when the second cause is much more prevalent.

Blip parameter estimates and measures of uncertainty, as well as the metrics for our proposed ITRs and the cause-specific ITR are presented in Appendix A. The cause-specific ITR yields POT and value estimates of 0.19 and -0.65, respectively. As expected, the cause-specific ITR is nearly optimal for those failing from the targeted cause, but is far from optimal for those who experience the alternative cause of failure. Since the second cause is more prevalent, the approach yields overall performance worse than even the uniform regime, which has an estimated value of 0.61. The weighted and greedy ITRs perform very well in this setting, both with POT and value estimates of 0.90 and 1.87, respectively.

3.4.2 Composite outcome ITR

The second approach constructs an ITR based on a composite endpoint which includes all event types. This amounts to ignoring distinct causes of failure and estimating one set of blip parameters based on all observed data. Heuristically, this corresponds to a form of weighting, with the combined set of blip parameters being a weighted average of the blip parameters associated to each cause, weighted by the prevalence of each cause. The composite ITR is estimated via DWSurv, with the outcome being the time to any event.

To illustrate this, we consider Setting 4, where the greedy regime outperforms the weighted regime. In this setting, the composite ITR yields POT and value estimates of 0.49 and 1.54 respectively, nearly identical to those of the weighted regime. The composite ITR shares the same flaws as the weighted regime and is similarly surpassed by the greedy regime in this setting. In such situations, a composite ITR provides sensible results; however, the added flexibility of being able to choose between the weighted and greedy regimes based on the objective of the

analysis is an advantage of our proposed approach.

4 Application: OPTN kidney transplant data

4.1 Background: Donor HCV status

In this section, we illustrate our proposed method via application to the kidney transplant data from the OPTN database. We analyze data from patients enrolled on the kidney transplantation waiting list from January 1, 2001 to December 31, 2022. The main exposure of interest is the donor's HCV status, with 0 denoting an HCV-negative donor and 1 denoting a donor living with HCV. We seek to maximize time to either graft failure (cause 1), or death with a functioning graft (cause 2). We include donor age, donor type (living or dead), and recipient HCV status as possible tailoring variables, with donor age scaled by 10 years. Other key variables, including those used for the treatment-free component of the outcome model, are presented in Appendix B. For the sake of simplicity, we conducted a complete-case analysis, removing patients with missing event indicators, missing survival times or missing covariates. We emphasize that the goal of this analysis is to illustrate our methodology rather than to provide accurate treatment recommendations.

4.2 Data and analytic models

Observations in our dataset were clustered by transplant centers, with a total of 251 distinct centers. We used an exchangeable working correlation structure. Due to large cluster sizes, we used a single iteration of the algorithm in Subsection 2.3 with initial correlation estimates obtained by fitting a weighted linear regression under independence. For the exchangeable structure, the inverse of the working correlation matrix can be derived analytically, leading to

greater computational efficiency (Lipsitz et al., 2017). Weights as in (4) were used for fitting the weighted GEE. Confidence intervals for blip parameters were obtained via nonparametric cluster bootstrap with $B = 1000$ replications. Logistic regression was used for the treatment, censoring and cause models; details of model specification are located in Appendix B. All outcome models were fitted in R using the `geem` package.

Patient characteristics are presented in Appendix B. Our final dataset was made up of 311,474 patients, with 15.9% suffering from graft failure, 16.1% dying with a functioning graft and 68% with censored outcomes. Transplants with HCV positive donors were rare within the overall study population, constituting only 3.6% of all transplants. This proportion was larger within HCV positive recipients, at 31%.

4.3 Results

Table 2 shows blip parameter estimates for each cause of failure along with associated measures of uncertainty. For both causes, the one-step procedure yielded an estimated within-cluster correlation of around 5%. For patients having suffered from graft failure, receiving a kidney from an HCV positive donor over an HCV negative donor has a negative effect on survival time; however, the recipient being HCV positive and the transplant coming from a living donor both significantly attenuate the negative effect. Effect estimates were similar for patients who died with a functioning graft, with the only difference being the statistically significant negative effect of donor HCV status on survival time. Donor age had no significant tailoring effect, either statistically or clinically.

The two regimes provided nearly identical performance, with an estimated POT of around 0.94 and an estimated value of 6.96. Since the blip parameters are fairly similar for both causes, either regime appears to be a good choice. Benefit plots for the weighted and greedy ITRs can be

found in Appendix B. For most patients, the estimated ITRs recommend transplants with HCV-negative donors over transplants with HCV-positive donors, which is consistent with previous literature. However, for both regimes, a relatively large portion of patients have estimated benefits for which the 95% bootstrap CI include 0: 20% and 28% of subjects respectively. For these patients, there is no apparent harm to using an HCV-positive donor over an HCV-negative donor as both yield similar outcomes. We note that the observed treatment allocation also results in an estimated value of 6.96; however, only 3.63% of patients received kidneys from an HCV-positive donor. Under either of our proposed regimes, a much larger proportion of the study population could have been allocated kidneys from an HCV-positive donor without impacting their overall survival.

Additionally, we conducted the same analysis using the cause-specific and composite outcome approaches. Parameter estimates and benefit plots for these two analyses can be found in Appendix B. The composite outcome analysis yielded estimates consistent with those obtained in the main analysis and the resulting ITR exhibited performance similar to that of the weighted regime, as was observed in Subsection 3.4.2 of the simulations. The cause-specific analysis also yielded reasonable point estimates; however, the associated standard errors were much larger than the equivalent standard errors in the main analysis. In this case, the artificial censoring triggered by the cause-specific analysis leads to greater variability in the estimated weights. An unweighted analysis could be performed to remedy this issue, at the cost of increased sensitivity to model misspecification. Furthermore, the cause-specific ITR performs worse than all the previously mentioned regimes, with an estimated POT and value of 0.79 and 6.79, respectively. Although the estimated blip parameters are similar across causes, they are not identical; in this case, only allocating treatment based on one failure type results in sub-optimal performance.

5 Conclusion

In this paper, we proposed a framework for optimal ITR estimation in the competing risk setting which accounts for both treatment effect heterogeneity and the uncertainty related to the unknown failure type. Estimates of treatment effects are obtained by solving doubly-robust GEEs, with weights adjusted for non-randomized treatment assignment and censoring. Treatment effect estimates can then be combined in various ways to yield optimal ITRs which target different criteria. For example, we introduced the weighted and greedy ITRs, which target overall survival integrated over all failure types and survival tied to the most likely failure type, respectively. The choice of an optimal ITR should mainly be driven by subject matter expertise and the nature of the research question, but metrics such as the value and the POT may be used to compare the performance of the chosen treatment rule to that of alternative regimes.

The main advantage of our approach is that it explicitly incorporates the uncertainty regarding the failure type into the definition of the optimal ITR, rather than implicitly assuming that the cause of failure is known at treatment allocation. Although one can repeat a naive cause-specific analysis for every event type separately, there is limited formal methodology for combining these results into a single optimal treatment rule. Alternatively, as discussed in Subsection 3.4.2, the use of a composite outcome can yield similar performance to that of the weighted ITR, at least in the case $K = 2$; consequently, it also performs poorly where the weighted regime does. In these cases, one could instead use the greedy rule, which was shown to yield better performance in certain scenarios where the weighted rule is less effective. Another strength of our approach is the AFT specification, which allows for more intuitive modelling, as treatment effects can be directly interpreted in terms of the survival time, which is not the case for hazard-based models.

One weakness of our proposed approach is that it requires the relatively strong assumption that the cause of failure K is conditionally independent of the joint distribution of the treatment and censoring mechanism given observed covariates, i.e., Assumptions 3 and 5. Conditional independence of the failure type and treatment assignment may not hold if, for example, the treatment induces side-effects that can cause patients to fail before they experience the event type of interest. To remedy this particular problem, we can derive alternative definitions of the optimal ITR under the weaker assumption of sequential ignorability, leading to the optimal rule $d_{\mathrm{w}}(x) = \arg\max_a \sum_{k=1}^{\kappa} \varphi_k(\boldsymbol{x}, a) Q_k(\boldsymbol{x}, a)$, $\boldsymbol{x} \in \mathcal{X}$, where $\varphi_k(\boldsymbol{x}, a) = P(K = k | \boldsymbol{X} = \boldsymbol{x}, A = a)$ and $Q_k(\boldsymbol{x}, a) = \mathbb{E}[f(T) | \boldsymbol{X} = \boldsymbol{x}, A = a, K = k]$. See Appendix C for details. The optimal ITR can then be directly estimated from the data by positing regression models for the transformed survival times and the failure type. However, this approach is singly-robust and thus is intrinsically more sensitive to model misspecification. Strategies to improve the robustness of this approach will be explored in future work. Furthermore, if we do not assume that the censoring mechanism is conditionally independent of the failure type, the algorithm outlined in Subsection 2.3 would require the specification of the censoring probability for each subject conditional on their failure type. Since the failure type is unobserved for censored subjects, this would require leveraging the observed data to impute missing failure types. As noted in Subsection 2.5, several different methods could be used in order to achieve this goal.

We leveraged our methodological developments to assess the important question of kidney allocation with HCV-positive donors using data from the OPTN. We found statistically significant tailoring effects for the recipient HCV status and the donor type. The resulting ITRs were then used to identify patients for whom kidney transplants with HCV-positive donors did not yield inferior outcomes to transplants with HCV-negative donors. Our analysis suggests that a large portion of patients in the OPTN data who received HCV-negative kidneys would see their

overall survival unchanged if they were instead given a kidney from an HCV-positive donor. The estimated treatment rules could then be used to allocate kidneys for future transplant candidates, allowing for a more efficient use of kidneys from HCV-positive donors. However, our analysis was subject to certain limitations. First, we performed a complete case analysis, which could lead to bias if the variables concerned were not missing completely at random. Furthermore, our choice of tailoring variables was mainly based on substantive clinical knowledge and limited pre-screening of effect sizes. A more principled data-driven variable selection procedure, such as the one utilized in penalized dWOLS (Bian et al., 2021), could be used to narrow down the set of candidate predictors.

In spite of these limitations, our analysis had several strengths including a large sample size, appropriate accounting of confounding, censoring, and clustering in addition to robustness against model misspecification. The work shown here represents the first ITR approach to formalizing decision-making for kidney-allocation with a goal to minimizing harm in the context of potentially sub-optimal donor organs, and has revealed that a much larger group of recipients could be offered organs from donors living with HCV without evidence that this would result in recipient harm.

Table 1: Root-n adjusted bias and standard errors for blip parameter estimators, as well as ITR metrics for settings 1-3. Estimates and metrics for setting 1(a) and 1(b) were computed across 4 model specifications: (i) outcome and weighting models misspecified, (ii) outcome model correctly specified, (iii) weighting models correctly specified, (iv) all models correctly specified. Estimates were computed using 1000 replicate datasets. Regimes are identified with d_w, d_g, d_o, and d_u representing the weighted, greedy, oracle and uniform regimes, respectively.

	$\boldsymbol{\psi}$	Setting									
		1(a)				1(b)				2	3
		i	ii	iii	iv	i	ii	iii	iv	-	-
$\sqrt{n} \times$ Bias	ψ_{11}	-22.41	-0.04	-0.27	-0.07	-50.09	0.09	0.09	0.13	0.10	-0.07
	ψ_{12}	0.30	-0.08	-0.12	-0.10	1.00	0.10	0.12	0.12	-0.00	-0.10
	ψ_{21}	14.96	0.06	0.11	0.05	33.58	0.01	0.05	-0.01	0.07	0.05
	ψ_{22}	0.06	0.08	0.12	0.08	0.08	0.13	0.25	0.19	0.14	0.08
$\sqrt{n} \times$ SE	ψ_{11}	3.23	2.79	3.78	3.29	3.17	2.88	4.14	3.53	3.49	3.29
	ψ_{12}	3.30	2.52	3.98	3.37	3.38	2.59	4.18	3.50	3.49	3.37
	ψ_{21}	2.00	2.21	2.67	2.54	1.99	2.18	2.72	2.58	2.60	2.54
	ψ_{22}	2.72	2.43	3.32	3.08	2.61	2.34	3.29	2.98	3.12	3.08
POT($\hat{d}_{\mathrm{w}}$)		0.84	0.93	0.93	0.93	0.83	0.93	0.93	0.93	0.56	0.51
POT($\hat{d}_{\mathrm{g}}$)		0.73	0.92	0.92	0.92	0.72	0.93	0.93	0.93	0.69	0.70
V($\hat{d}_{\mathrm{w}}$)		1.73	1.81	1.81	1.81	1.70	1.76	1.76	1.76	2.30	1.61
V($\hat{d}_{\mathrm{g}}$)		1.72	1.81	1.79	1.81	1.68	1.76	1.76	1.76	2.17	1.55
V(d_{o})		1.81	1.81	1.81	1.81	1.77	1.77	1.77	1.77	2.81	1.68
V(d_{u})		1.67	1.67	1.67	1.67	1.62	1.62	1.62	1.77	1.89	1.54

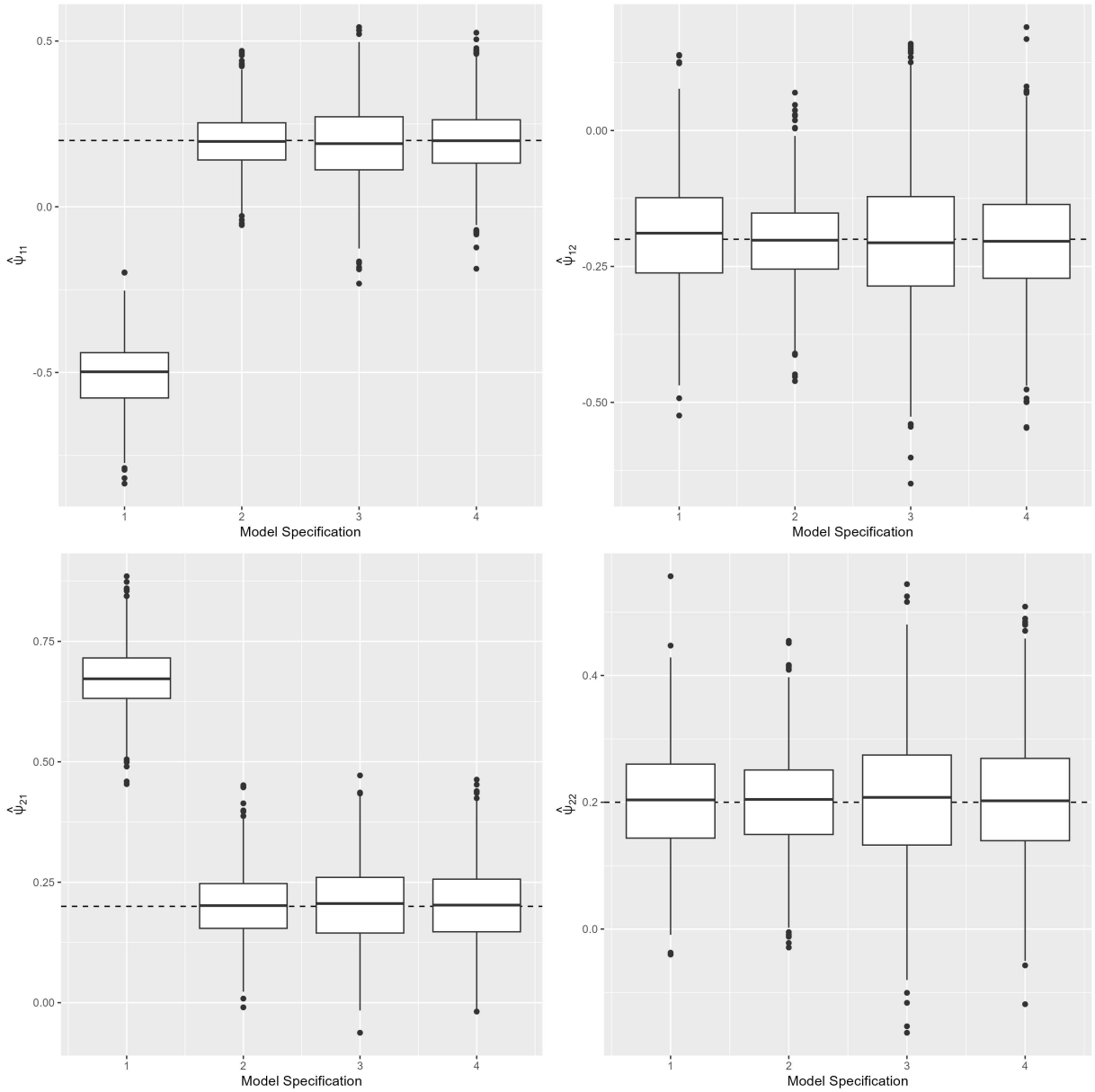

Figure 1: Empirical distribution of blip parameter estimates obtained via the weighted AFT-GEE model with sample size $n_{\text{train}} = 1000$ in setting 1(a) across 4 scenarios: (i) outcome and weighting models misspecified, (ii) outcome model correctly specified, (iii) weighting models correctly specified, (iv) all models correctly specified. Estimates were computed over 1000 replicate datasets.

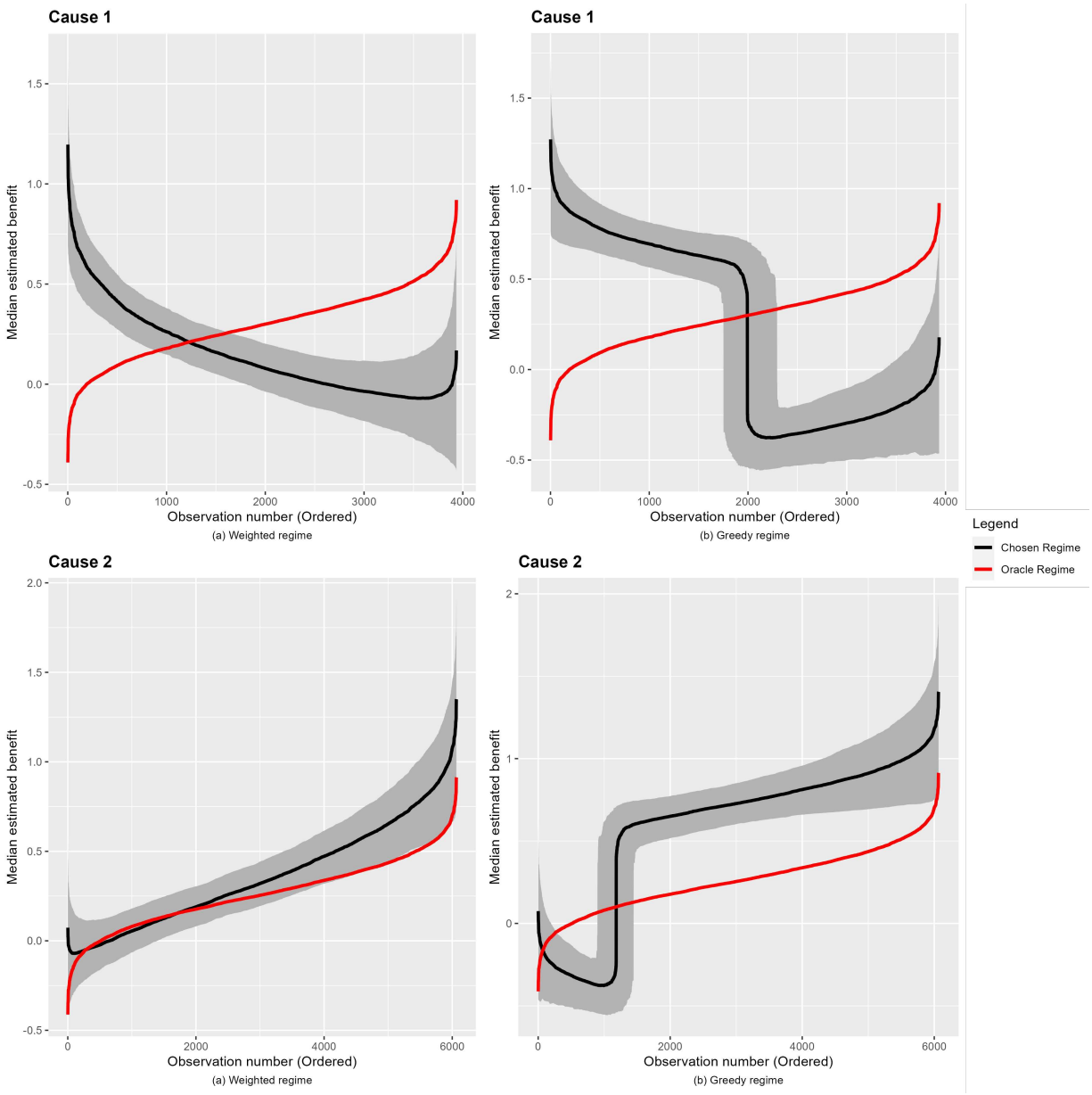

Figure 2: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 1(a) and model (i), for causes $K = 1$ (top row) and $K = 2$ (bottom row), $n_{\text{train}} = 1000$. Benefit curves were evaluated on the test set, with subjects ordered by their increasing benefit. Estimates were computed over 1000 replicate datasets.

Table 2: Blip parameter estimates and associated 95% bootstrap confidence intervals (CI) for both causes of failure in the OPTN data on $n = 311,474$ individuals from 251 centers, based on kidney transplantations carried out from January 1, 2001 to December 31, 2022. Nonparametric cluster bootstrap CIs were computed using $B = 1000$ bootstrap replicates. Donor and recipient variables are prefixed by Don and Rec, respectively.

Parameters	Graft Failure		Death with functioning graft	
	Estimate	95% CI	Estimate	95% CI
DonHCV	-0.86	$(-2.04, 0.04)$	-1.44	$(-2.30, -0.51)$
DonHCVxRecHCV	0.90	$(0.55, 1.32)$	1.21	$(0.97, 1.47)$
DonHCVxDonType	1.14	$(0.61, 1.71)$	0.40	$(-0.21, 1.01)$
DonHCVxDonAge	$1.55\text{x}10^{-5}$	$(-1.36\text{x}10^{-3}, 1.82\text{x}10^{-3})$	$4.32\text{x}10^{-4}$	$(-1.00\text{x}10^{-3}, 1.73\text{x}10^{-3})$

SUPPLEMENTARY MATERIAL

Appendix A, B, and C: Appendix A contains additional simulations and figures. Appendix B provides supplementary information regarding the data analysis. Appendix C presents proofs for the main results of the paper. (.pdf files)

References

Z. Bian, E. E. M. Moodie, S. Shortreed, and S. Bhatnagar. Variable selection in regression-based estimation of dynamic treatment regimes. *Biometrics*, 79, 11 2021.

L.-W. Chen, I. Yavuz, Y. Cheng, and A. S. Wahed. Cumulative incidence regression for dynamic treatment regimens. *Biostatistics*, 21(2):113–130, 2018.

S. Choi and H. Cho. Accelerated failure time models for the analysis of competing risks. *Journal of the Korean Statistical Society*, 48(3):315–326, 2019.

J. Clifton and E. Laber. Q-learning: Theory and applications. *Annual Review of Statistics and Its Application*, 7(1):279–301, 2020.

A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, Cambridge, UK, 1997.

J. P. Fine and R. J. Gray. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

R. D. Gill and J. M. Robins. Causal inference for complex longitudinal data: the continuous case. *The Annals of Statistics*, 29(6):1785–1811, 2001.

R. D. Gill, M. J. van der Laan, and J. M. Robins. Coarsening at Random: Characterizations, Conjectures, Counter-Examples. In D. Lin and T. R. Fleming, editors, *Proceedings of the First Seattle Symposium in Biostatistics*, pages 255–294. Springer, 1997.

G. Gupta, L. Kang, J. W. Yu, A. J. Limkemann, V. Garcia, D. Bandyopadhyay, D. Kumar, H. Fattah, M. Levy, A. H. Cotterell, A. Sharma, C. Bhati, T. Reichman, A. L. King, and R. Sterling. Long-term outcomes and transmission rates in hepatitis C virus-positive donor to hepatitis C virus-negative kidney transplant recipients: analysis of United States national data. *Clinical Transplantation*, 31(10):e13055, 2017.

E. B. Laber, D. J. Lizotte, and B. Ferguson. Set-valued dynamic treatment regimes for competing outcomes. *Biometrics*, 70(1):53–61, 2014.

K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

S. Lipsitz, G. Fitzmaurice, D. Sinha, N. Hevelone, J. Hu, and L. Nguyen. One-step generalized estimating equations with large cluster sizes. *Journal of Computational and Graphical Statistics*, 26:734–737, 2017.

A. R. Luedtke and M. J. van der Laan. Optimal individualized treatments in resource-limited settings. *The International Journal of Biostatistics*, 12(1):283–303, 2016.

J. K. Lunceford and M. Davidian. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*, 23(19): 2937–2960, 2004.

G. J. McLachlan, S. X. Lee, and S. I. Rathnayake. Finite mixture models. *Annual Review of Statistics and Its Application*, 6:355–378, 2019.

P. Morzywołek, J. Steen, W. Van Biesen, J. Decruyenaere, and S. Vansteelandt. On estimation and cross-validation of dynamic treatment regimes with competing risks. *Statistics in Medicine*, 41(26):5258–5275, 2022.

S. A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.

J. Orbe and V. Núñez-Antón. Alternative approaches to study lifetime data under different scenarios: from the PH to the modified semiparametric AFT model. *Computational Statistics & Data Analysis*, 50(6):1565–1582, 2006.

B. J. Pereira, E. L. Milford, R. L. Kirkman, and A. S. Levey. Transmission of hepatitis C virus by organ transplantation. *New England Journal of Medicine*, 325(7):454–460, 1991.

R. L. Prentice, J. D. Kalbfleisch, A. V. Peterson, N. Flournoy, V. T. Farewell, and N. E.

Breslow. The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4): 541–554, 1978.

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

D. Rubin. Discussion of "Randomization analysis of experimental data in the Fisher randomization test" by D. Basu. *Journal of the American Statistical Association*, 75(371):591–593, 1980.

G. Simoneau, E. E. M. Moodie, J. S. Nijjar, R. W. Platt, and the Scottish Early Rheumatoid Arthritis Inception Cohort Investigators. Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association*, 115(531):1531–1539, 2020.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

M. P. Wallace and E. E. M. Moodie. Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71:636–664, 2015.

I. Yavuz, Y. Cheng, and A. S. Wahed. Estimating the cumulative incidence function of dynamic treatment regimes. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 181(1):85–106, 2016.

K.-H. Yuan and R. Jennrich. Estimating Equations with Nuisance Parameters: Theory and Applications. *Annals of the Institute of Statistical Mathematics*, 52:343–350, 2000.

Appendix A

Plots for simulations

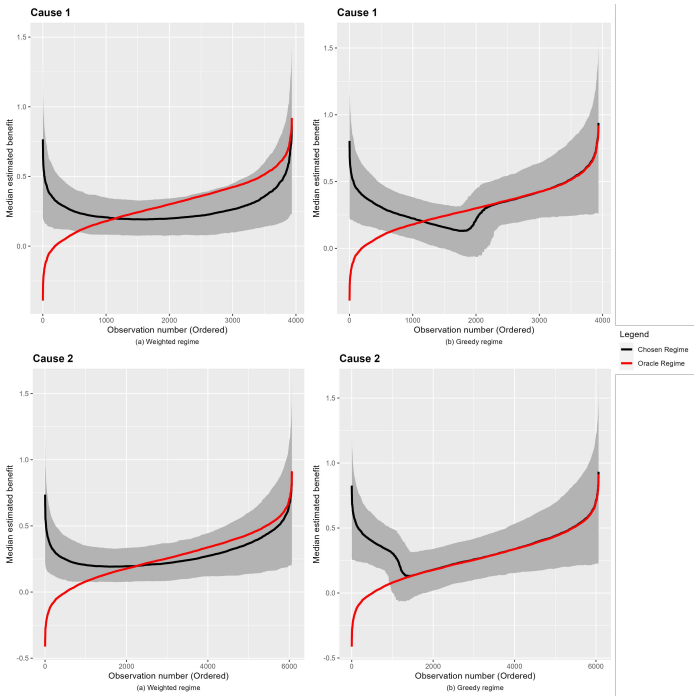

Figure A1: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 1(a) and model (iv), for causes $K = 1$ (top row) and $K = 2$ (bottom row), $n_{\text{train}} = 1000$. Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

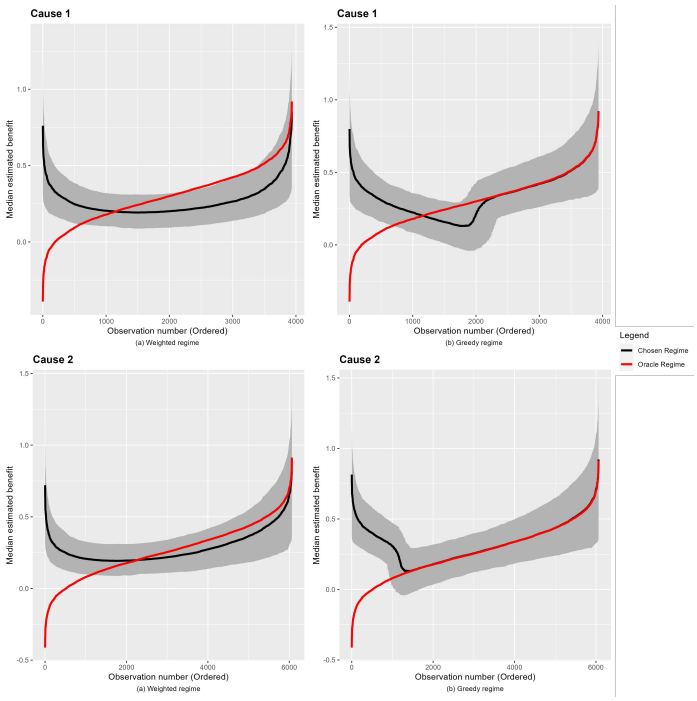

Figure A2: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for model setting 1(a) and model (ii), for causes $K = 1$ (top row) and $K = 2$ (bottom row), $n_{\text{train}} = 1000$. Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

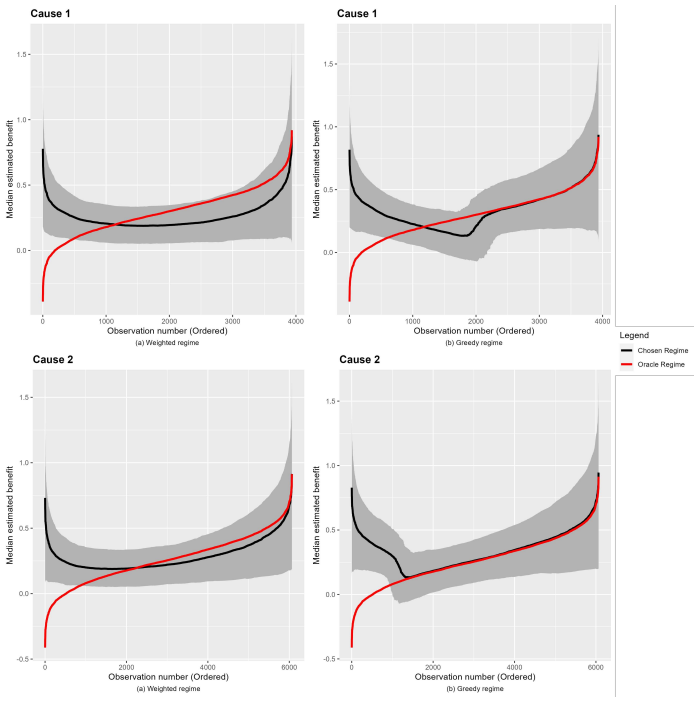

Figure A3: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 1(a) and model (iii), for causes $K = 1$ (top row) and $K = 2$ (bottom row), $n_{\text{train}} = 1000$. Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

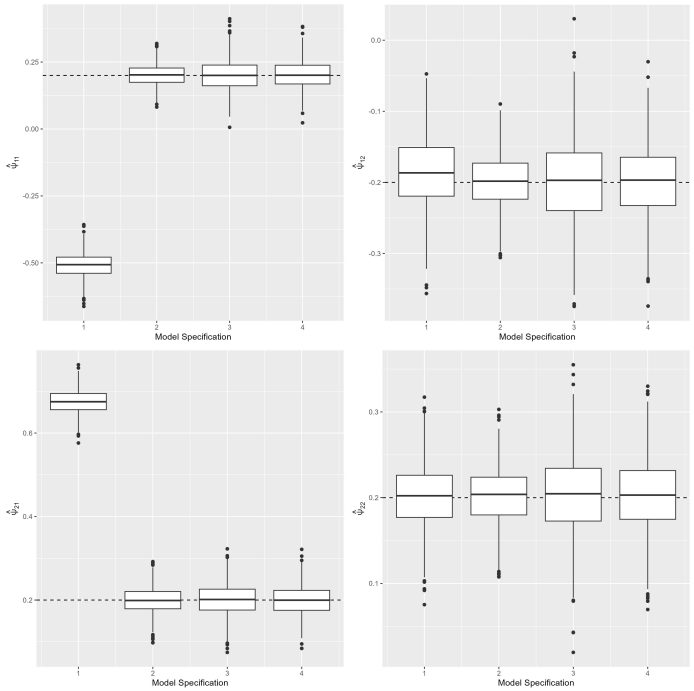

Figure A4: Empirical distribution of blip parameter estimates obtained via the weighted AFT-GEE model with sample size $n_{\text{train}} = 5000$ in setting 1(b) across 4 scenarios: (i) outcome and weighting models misspecified, (ii) outcome model correctly specified, (iii) weighting models correctly specified, (iv) all models correctly specified. Estimates were computed over 1000 replicate datasets.

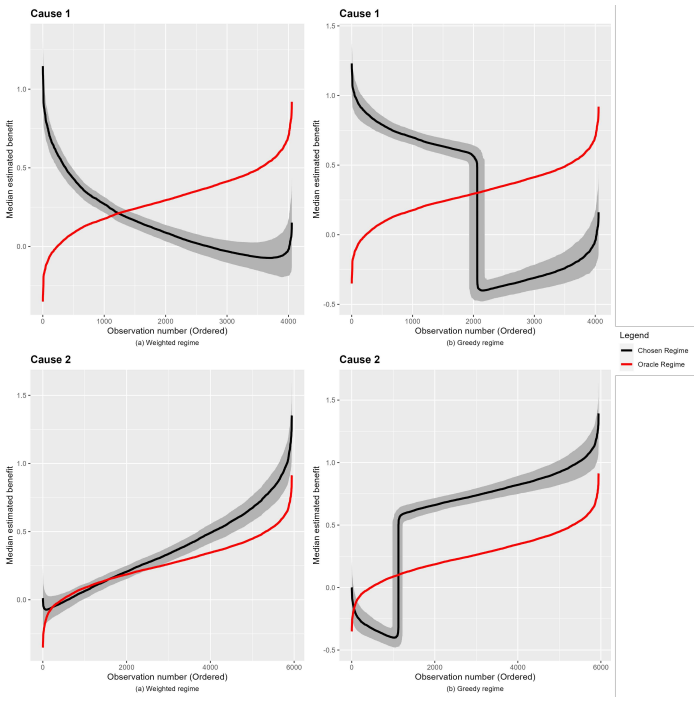

Figure A5: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 1(b) and model (i), for causes $K = 1$ (top row) and $K = 2$ (bottom row), $n_{\text{train}} = 5000$. Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

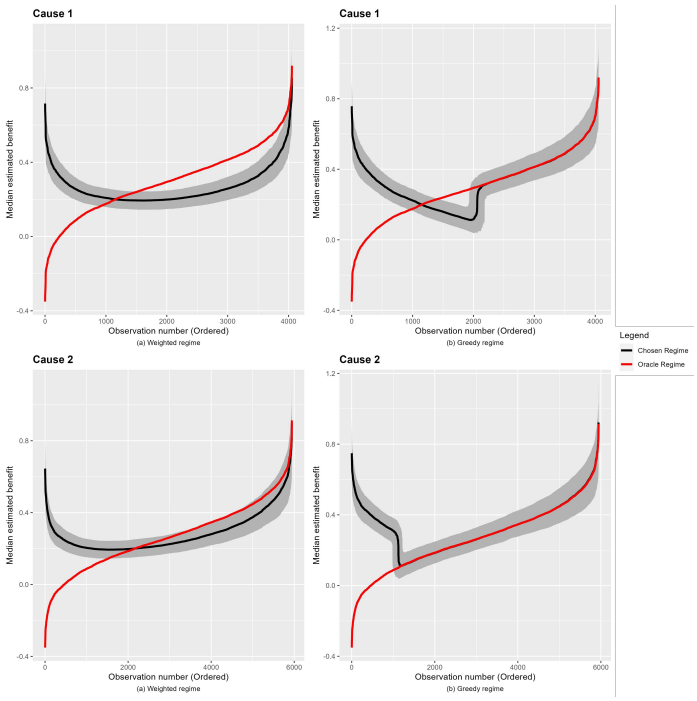

Figure A6: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 1(b) and model (ii), for causes $K = 1$ (top row) and $K = 2$ (bottom row), $n_{\text{train}} = 5000$. Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

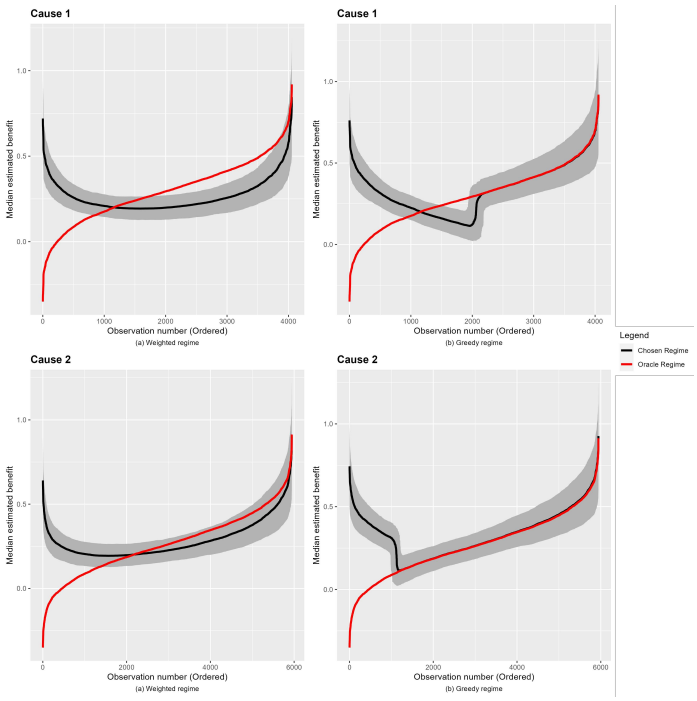

Figure A7: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 1(b) and model (iii), for causes $K = 1$ (top row) and $K = 2$ (bottom row), $n_{\text{train}} = 5000$. Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

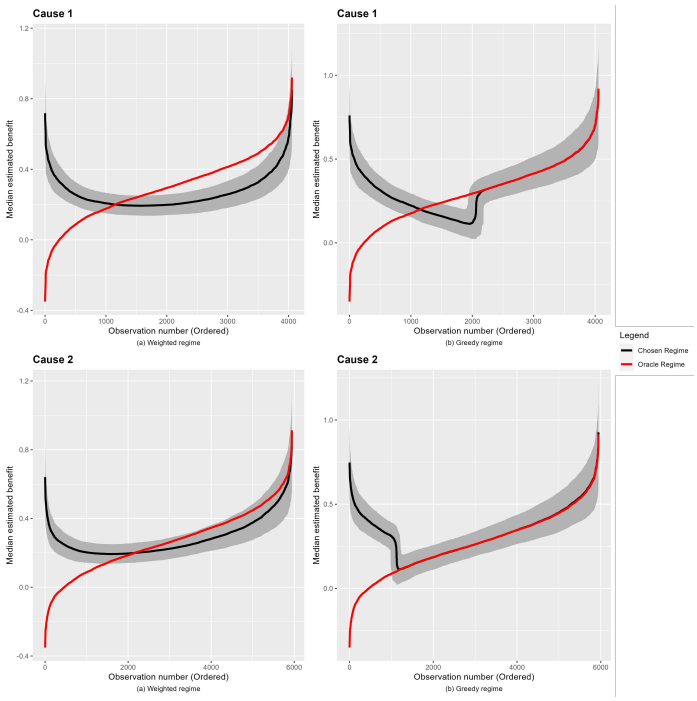

Figure A8: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 1(b) and model (iv), for causes $K = 1$ (top row) and $K = 2$ (bottom row), $n_{\text{train}} = 5000$. Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

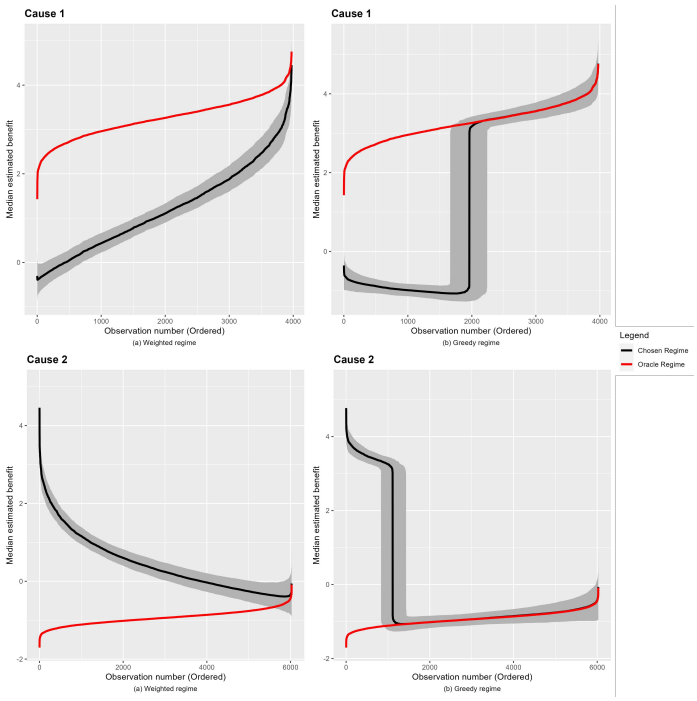

Figure A9: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 2, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

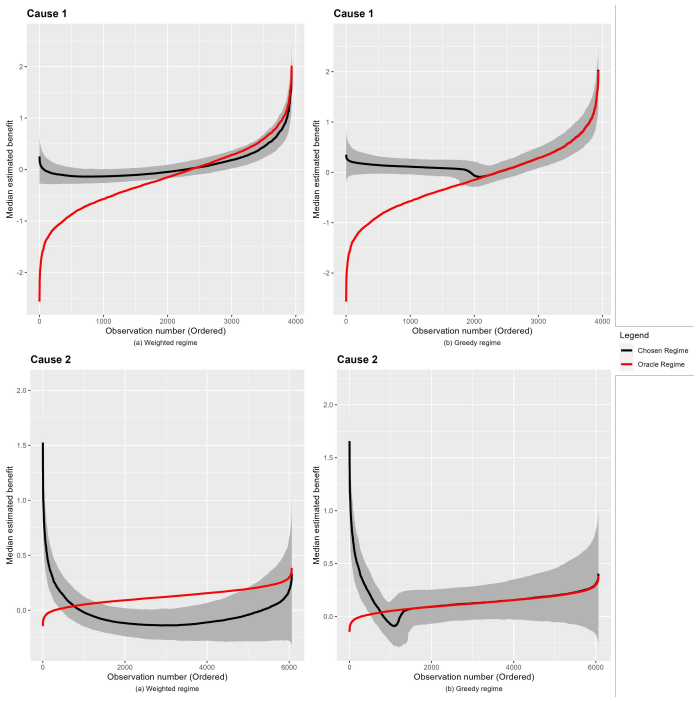

Figure A10: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 3, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

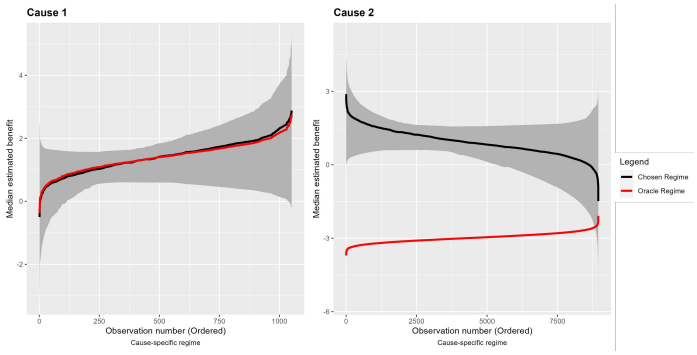

Figure A11: Median estimated benefit for the cause-specific regime along with the benefit of the oracle regime for setting 10, for causes $K = 1$ (left) and $K = 2$ (right). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

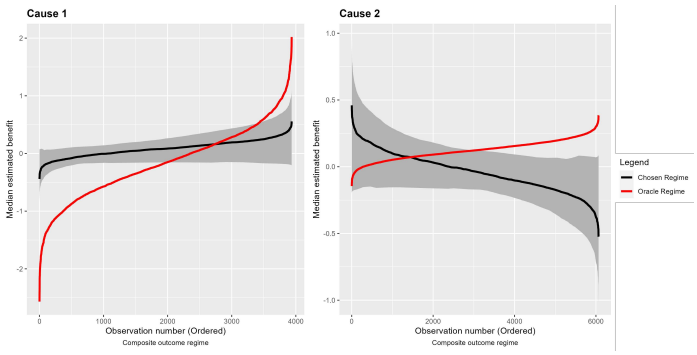

Figure A12: Median estimated benefit for the composite outcome regime along with the benefit of the oracle regime for setting 4, for causes $K = 1$ (left) and $K = 2$ (right). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

Additional simulations

This set of simulations explores the sensitivity of the proposed estimation procedure to variations in the data generating process and other modelling choices. For settings 4-9, we use the following parameter values:

$$(\psi_{11}, \psi_{12}) = (0.6, -0.6), \quad (\psi_{21}, \psi_{22}) = (-0.6, -0.6), \quad \tau^2 + \sigma^2 = 1,$$

and consider the following settings:

$$\begin{aligned}
&\text{Setting 4 : As in section 3.1, with } \delta_0 = 1.73 \text{ and ICC} = 0.5,\\
&\text{Setting 5.1 : ICC} = 0.9,\\
&\text{Setting 5.2 : ICC} = 0.1,\\
&\text{Setting 6 : } U_i \sim \text{Gamma}(\tau^2, 1) - \tau^2,\\
&\text{Setting 7 : } \delta_0 = 0, \text{corresponding to 50\% censoring,}\\
&\text{Setting 8 : Independence working correlation matrix, i.e., } \mathbf{V}_i = \text{diag}(\lambda)_{j=1,\ldots,r_i}.
\end{aligned}$$

For the next three simulation settings, we use the default specification of setting 4 and consider clustering in the treatment assignment, with the same cluster indices as for the outcome model. To induce dependence in treatment values, we add a random intercept to the logistic regression model for A_{ij}. For cluster $i = 1, \ldots, r$ and subject $j = 1, \ldots, r_i$, we generate values of the treatment according to

$$E_i \sim \mathcal{N}(0, \xi^2), \quad A_{ij} \sim \text{Bernoulli}\Big(\text{expit}(0.5 + X_{ij1} + X_{ij2} + E_i)\Big).$$

The following values for the variance of the random intercept were considered, each corresponding to a different degree of within-cluster dependence of treatments:

$$\text{Setting 9.1, 9.2, 9.3 :} \quad \xi^2 = (0.01, 0.25, 1).$$

Standard errors, biases, and metrics for the derived ITRs are found in table S1. As expected, the standard errors for the blip estimators were larger when the degree of censoring was more substantial (setting 7); SEs were also larger when the working correlation structure was misspecified (setting 8). For fixed overall variance $\tau^2 + \sigma^2$, SEs were larger for smaller values of ICC, corresponding to larger within-cluster variability (settings 5.1 and 5.2). Standard errors were generally unaffected by the specification of the random effect distribution (setting 6) and by clustering within treatment (settings 9.1, 9.2 and 9.3). As for the derived ITRs, both greedy and weighted rules performed very similarly across all settings, with reasonable POTs and values close to those of the oracle regime. Benefit plots are presented below.

Table S1: Root-n adjusted bias and standard errors for blip parameter estimators, as well as ITR metrics for settings 4-10. Estimates were computed using 1000 replicate datasets. Regimes are identified with d_w, d_g, d_{cs}, d_o, and d_u representing the weighted, greedy, cause-specific, oracle and uniform regimes, respectively.

	$\boldsymbol{\psi}$	Setting									
		4	5.1	5.2	6	7	8	9.1	9.2	9.3	10
$\sqrt{n} \times$ Bias	ψ_{11}	-0.10	-0.16	-0.07	0.09	0.01	-0.13	0.25	0.31	0.11	-0.28
	ψ_{12}	-0.14	-0.07	-0.19	-0.07	0.05	-0.26	-0.00	0.05	-0.13	0.67
	ψ_{21}	0.08	0.07	0.05	0.03	-0.28	0.01	0.02	0.08	-0.07	0.03
	ψ_{22}	0.11	0.05	0.16	0.25	-0.17	0.14	0.07	-0.14	0.02	0.22
$\sqrt{n} \times$ SE	ψ_{11}	4.65	3.88	5.22	4.90	7.04	5.20	4.90	4.68	4.89	8.14
	ψ_{12}	4.77	4.12	5.24	4.89	7.16	5.16	4.76	4.54	4.88	7.49
	ψ_{21}	3.59	2.96	3.99	3.58	4.96	3.98	3.32	3.59	3.32	1.95
	ψ_{22}	4.35	3.86	4.66	4.18	6.46	4.81	4.29	4.10	4.41	2.38
POT($\hat{d}_{\text{w}}$)		0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.90
POT($\hat{d}_{\text{g}}$)		0.75	0.75	0.75	0.75	0.75	0.75	0.76	0.76	0.76	0.90
POT($\hat{d}_{\text{cs}}$)		-	-	-	-	-	-	-	-	-	0.19
V($\hat{d}_{\text{w}}$)		1.82	1.84	1.78	1.82	1.81	1.82	1.74	1.73	1.74	1.87
V($\hat{d}_{\text{g}}$)		1.82	1.82	1.78	1.79	1.79	1.81	1.73	1.73	1.73	1.87
V($\hat{d}_{\text{cs}}$)		-	-	-	-	-	-	-	-	-	-0.65
V(d_{o})		1.92	1.93	1.89	1.92	1.91	1.92	1.83	1.83	1.83	2.02
V(d_{u})		1.48	1.50	1.45	1.48	1.48	1.48	1.40	1.40	1.40	0.61

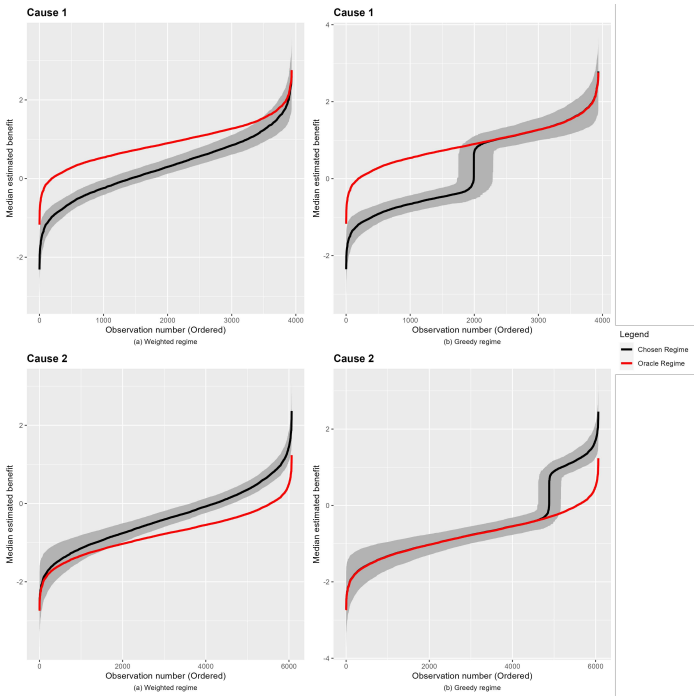

Figure A13: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 4, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

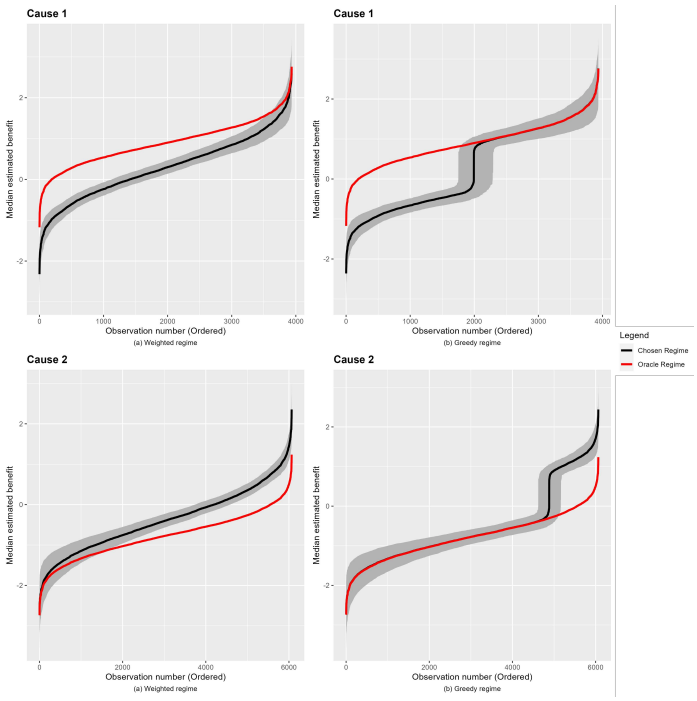

Figure A14: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 5.1, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

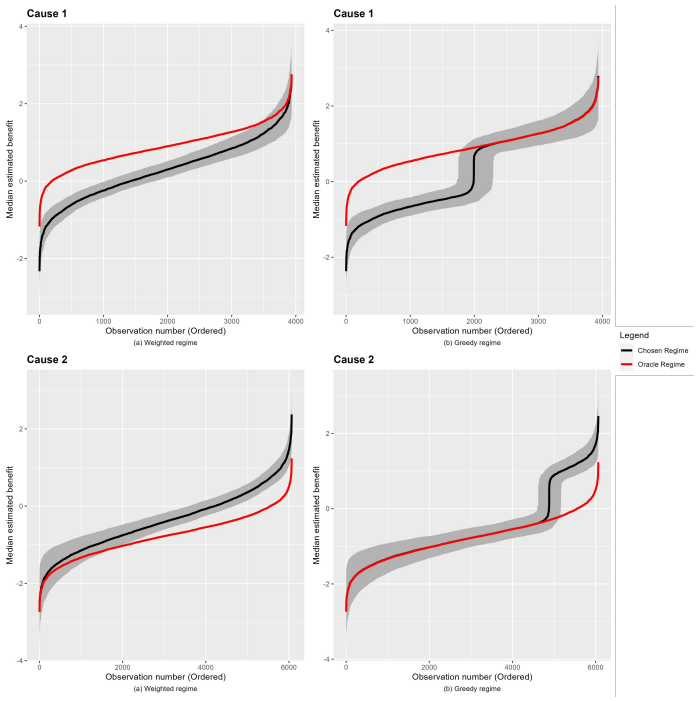

Figure A15: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 5.2, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

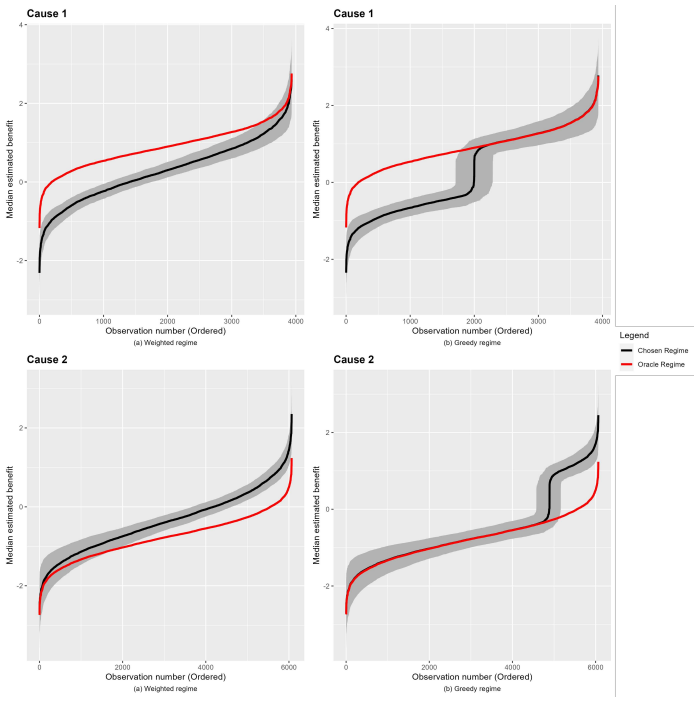

Figure A16: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 6, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

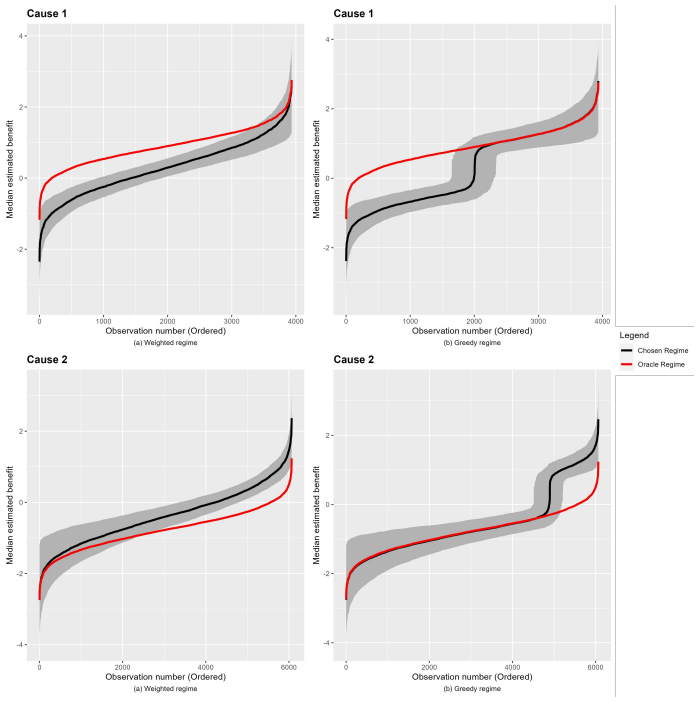

Figure A17: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 7, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

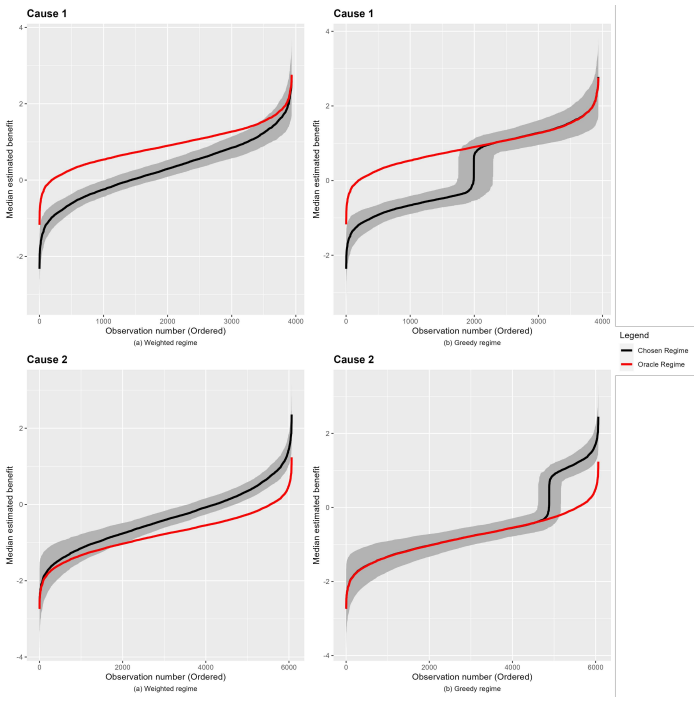

Figure A18: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 8, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

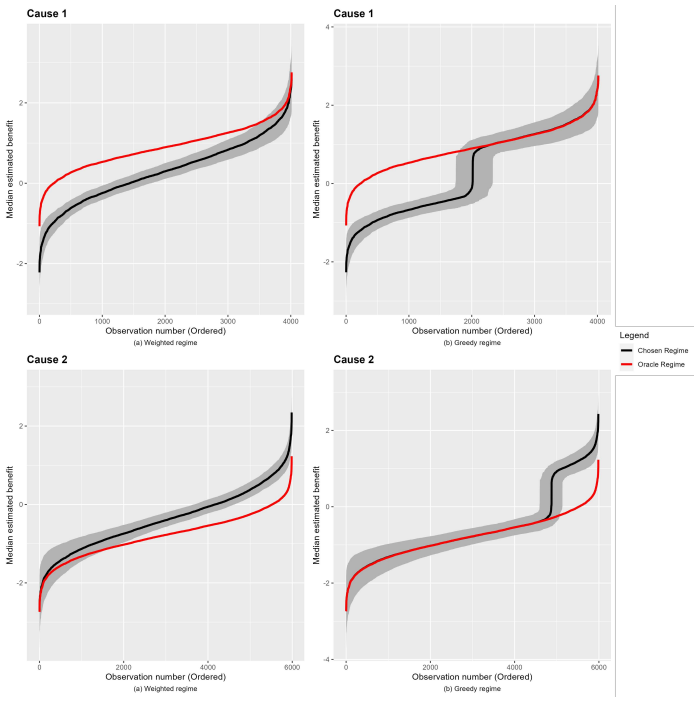

Figure A19: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 9.1, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

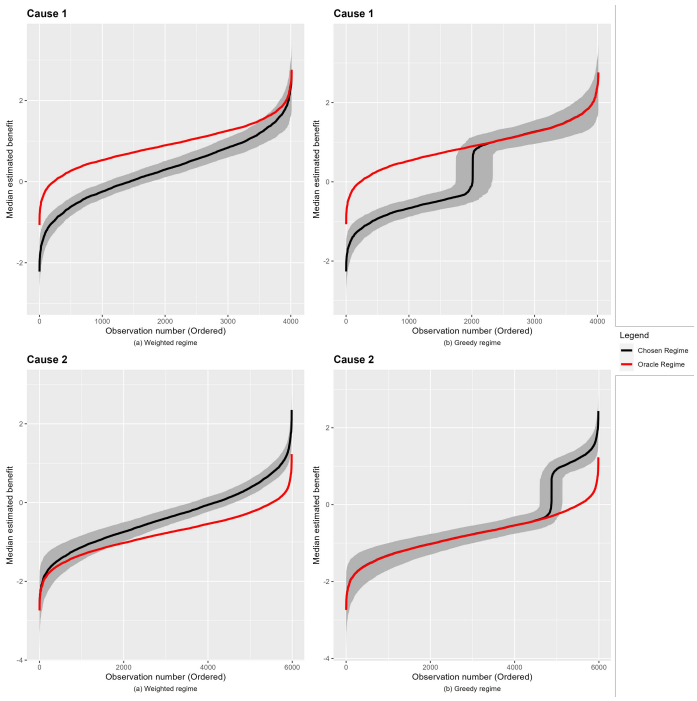

Figure A20: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 9.2, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

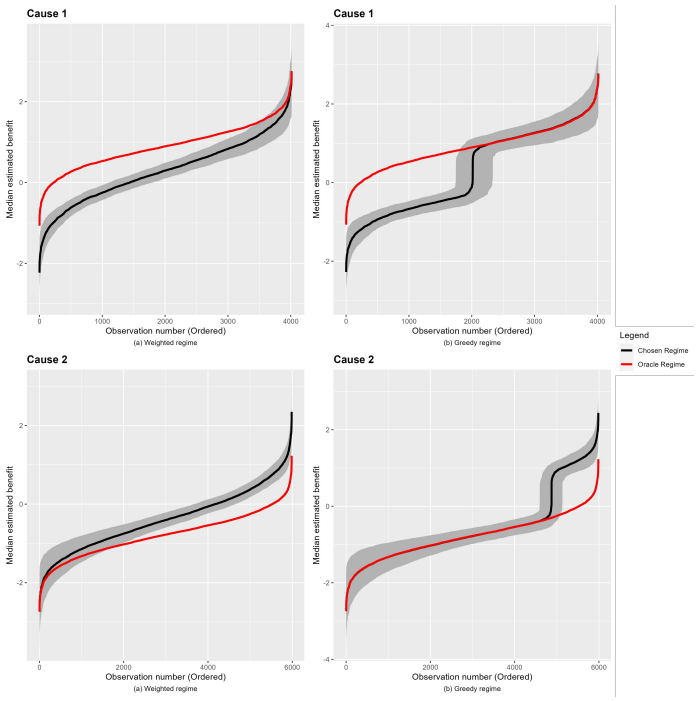

Figure A21: Median estimated benefit for the weighted (left column) and greedy (right column) regimes along with the benefit of the oracle regime for setting 9.3, for causes $K = 1$ (top row) and $K = 2$ (bottom row). Benefit curves were evaluated on the test set, with subjects ordered by their benefit in increasing order. Estimates were computed over 1000 replicate datasets.

Appendix B

The following models were used for the data analysis:

$$
\begin{aligned}
P(\text{DonHCV} = 1) = \text{expit}\Big(&\text{RecAge} + \text{RecGender} + \text{RecAge} + \text{RecEthnicity} + \text{RecDiabetes} + \text{RecHCV} + \\
&\text{RecImmunoGroup} + \log \text{CenterSize}\Big) \\
P(\Delta = 1) = \text{expit}\Big(&\text{RecAge} + \text{RecGender} + \text{RecAge} + \text{RecEthnicity} + \text{RecDiabetes} + \text{RecHCV} + \\
&\text{RecImmunoGroup} + \log \text{CenterSize} + \text{DonorHCV} + \text{DonorType} + \text{DonorAge} + \\
&\text{DonorEthnicity} + \text{DonorGender} + \text{DonorCigUse} + \text{DonorStroke} + \\
&\text{KidneyIschemicTime} + \text{OrganAllocationType}\Big) \\
E[T|K = k] = &\text{RecAge} + \text{RecGender} + \text{RecAge} + \text{RecEthnicity} + \text{RecDiabetes} + \text{RecHCV} + \\
&\text{RecImmunoGroup} + \log \text{CenterSize} + \text{DonorHCV} + \text{DonorType} + \text{DonorAge} + \\
&\text{DonorEthnicity} + \text{DonorGender} + \text{DonorCigUse} + \text{DonorStroke} + \\
&\text{KidneyIschemicTime} + \text{OrganAllocationType} + \\
&\text{DonorHCV} \times \Big(\text{RecHCV} + \text{DonorType} + \text{RecAge}\Big), \quad k = 1, 2 \\
P(K = 1) = \text{expit}\Big(&\text{RecAge} + \text{RecGender} + \text{RecAge} + \text{RecEthnicity} + \text{RecDiabetes} + \text{RecHCV} + \\
&\text{RecImmunoGroup} + \log \text{CenterSize} + \text{DonorHCV} + \text{DonorType} + \text{DonorAge} + \\
&\text{DonorEthnicity} + \text{DonorGender} + \text{DonorCigUse} + \text{DonorStroke} + \\
&\text{KidneyIschemicTime} + \text{OrganAllocationType}\Big).
\end{aligned}
$$

Donor and recipient variables are prefixed by Don and Rec, respectively.

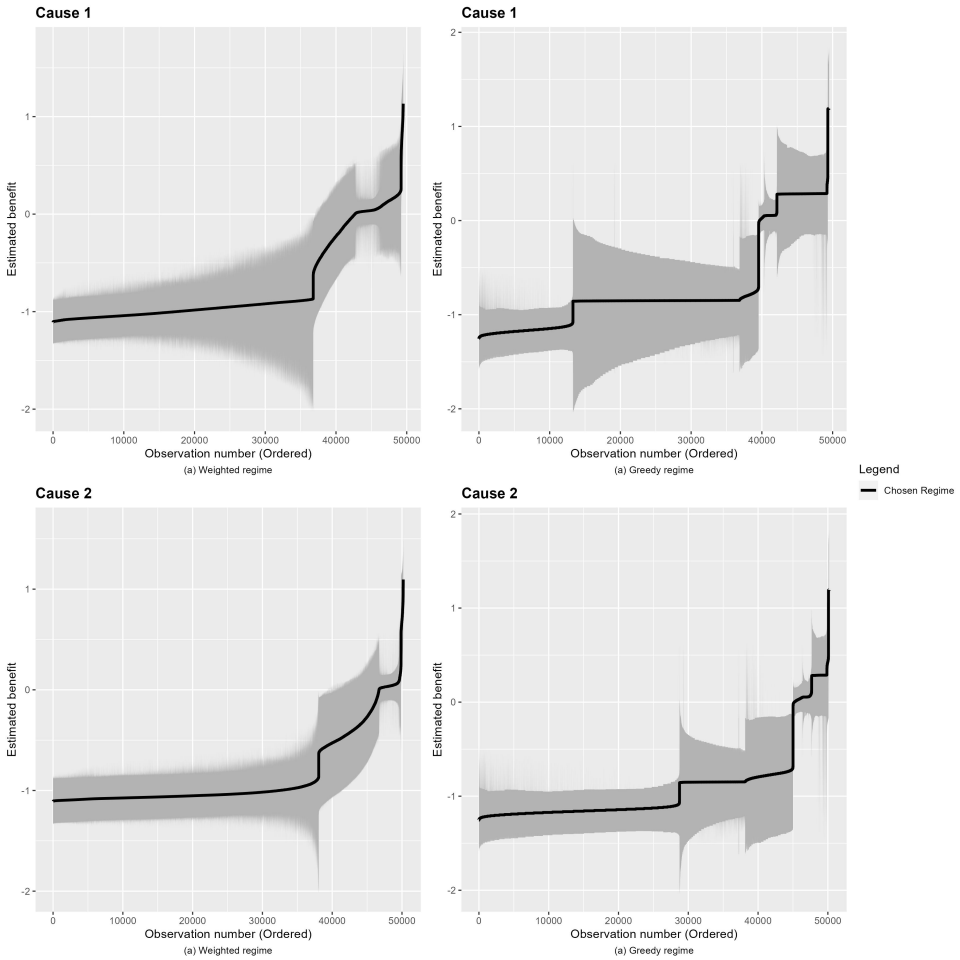

Figure B1: Estimated benefit for the weighted (left column) and greedy (right column) regimes for causes $K = 1$ (top row) and $K = 2$ (bottom row), along with associated pointwise 95% bootstrap confidence intervals. The analysis was conducted using the OPTN data on $n = 311,474$ individuals from 251 centers, based on kidney transplantations carried out in the period from January 1, 2001 to December 31, 2022. Benefit curves were evaluated for all non-censored individuals, with subjects ordered by their increasing benefit. Bootstrap intervals were computed using $B = 1000$ replicate datasets.

Table S1: OPTN kidney transplant characteristics stratified by donor HCV status along with standardized mean differences (SMD).

Characteristics	HCV-	HCV+	Overall	SMD
	(n=300,161)	(n=11,313)	(n=311,474)	
Recipient gender				0.249
Female	119857 (39.9%)	3190 (28.2%)	123047 (39.5%)	
Male	180304 (60.1%)	8123 (71.8%)	188427 (60.5%)	
Recipient ethnicity				0.399
Non-African American	220256 (73.4%)	6179 (54.6%)	226435 (72.7%)	
African American	79905 (26.6%)	5134 (45.4%)	85039 (27.3%)	
Immuno-group				0.099
1	186856 (62.3%)	7518 (66.5%)	194374 (62.4%)	
2	63193 (21.1%)	2012 (17.8%)	65205 (20.9%)	
3	9908 (3.3%)	410 (3.6%)	10318 (3.3%)	
4	40204 (13.4%)	1373 (12.1%)	41577 (13.3%)	
Recipient HCV status				0.922
HCV-	290861 (96.9%)	7191 (63.6%)	298052 (95.7%)	
HCV+	9300 (3.1%)	4122 (36.4%)	13422 (4.3%)	
Donor type				0.726
Deceased	214903 (71.6%)	10925 (96.6%)	225828 (72.5%)	
Living	85258 (28.4%)	388 (3.4%)	85646 (27.5%)	
Recipient age (years)				0.547
Mean (SD)	49.8 (15.6)	57.1 (10.9)	50.0 (15.5)	
Donor age (years)				0.467
0-18	22956 (7.6%)	44 (0.4%)	23000 (7.4%)	
19-49	197611 (65.8%)	9314 (82.3%)	206925 (66.4%)	
$>$ 49	79594 (26.5%)	1955 (17.3%)	81549 (26.2%)	
Donor ethnicity				0.176
Non-African American	261997 (87.3%)	10473 (92.6%)	272470 (87.5%)	
African American	38164 (12.7%)	840 (7.4%)	39004 (12.5%)	
Donor gender				0.151
Female	137046 (45.7%)	4324 (38.2%)	141370 (45.4%)	
Male	163115 (54.3%)	6989 (61.8%)	170104 (54.6%)	
Donor history of cigarette use				0.338
No	232353 (77.4%)	7027 (62.1%)	239380 (76.9%)	
Yes	67808 (22.6%)	4286 (37.9%)	72094 (23.1%)	
Donor cerebrovascular/stroke				0.113
No	237035 (79.0%)	9431 (83.4%)	246466 (79.1%)	
Yes	63126 (21.0%)	1882 (16.6%)	65008 (20.9%)	
Kidney cold ischemic time (hours)				0.507
Mean (SD)	13.6 (10.5)	18.5 (8.49)	13.8 (10.5)	
Organ allocation type				0.835
Local	38210 (12.7%)	3585 (31.7%)	41795 (13.4%)	
Regional	0 (0%)	0 (0%)	0 (0%)	
National	236221 (78.7%)	4642 (41.0%)	240863 (77.3%)	
Foreign	25730 (8.6%)	3086 (27.3%)	28816 (9.3%)	
Log of center size				0.094
Mean (SD)	7.47 (0.811)	7.55 (0.718)	7.48 (0.808)	
Follow-up time (days)				0.614
Median [IQR]	1526 [695, 2900]	739 [355, 1469]	1480 [676, 2875]	
Event type				0.154
Censored	203410 (67.8%)	8374 (74.0%)	211784 (68.0%)	
Graft failure	48227 (16.1%)	1281 (11.3%)	49508 (15.9%)	
Death with functioning graft	48524 (16.2%)	1658 (14.7%)	50182 (16.1%)	

Figure B2: Blip parameter estimates and associated 95% bootstrap confidence intervals (CI) for the cause-specific and composite outcome analyses using the OPTN data on $n = 311,774$ individuals from 251 centers, based on kidney transplantations carried out in the period from January 1, 2001 to December 31, 2022. Nonparametric cluster bootstrap CIs were computed using $B = 1000$ bootstrap replicates.

Parameters	Cause-specific		Composite outcome	
	Estimate	95% CI	Estimate	95% CI
DonHCV	-0.87	$(-3.31, 6.26)$	-0.97	$(-3.00, -0.53)$
DonHCVxRecHCV	0.93	$(-0.38, 1.80)$	1.08	$(0.96, 1.55)$
DonHCVxDonType	1.22	$(-0.41, 1.89)$	0.82	$(0.55, 1.23)$
DonHCVxDonAge	$3.77\text{x}10^{-5}$	$(-8.62\text{x}10^{-3}, 3.85\text{x}10^{-3})$	$1.36\text{x}10^{-4}$	$(-7.02\text{x}10^{-4}, 2.49\text{x}10^{-3})$

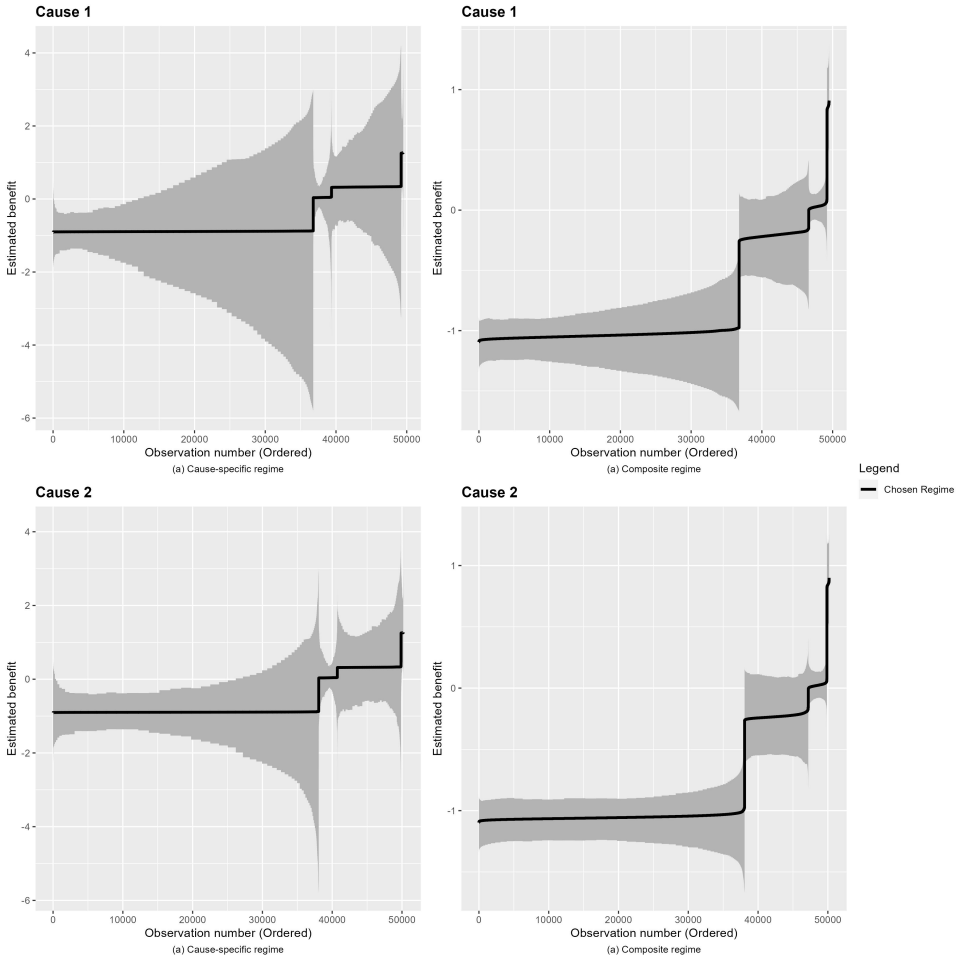

Figure B3: Estimated benefit for the cause-specific (left column) and composite (right column) regimes for causes $K = 1$ (top row) and $K = 2$ (bottom row), along with associated pointwise 95% bootstrap confidence intervals. The analysis was conducted using the OPTN data on $n = 311,474$ individuals from 251 centers, based on kidney transplantations carried out in the period from January 1, 2001 to December 31, 2022. Benefit curves were evaluated for all non-censored individuals, with subjects ordered by their increasing benefit. Bootstrap intervals were computed using $B = 1000$ replicate datasets.

Appendix C

Proof of consistency of blip estimators

Suppose we are estimating the blip parameters for cause k and let $t_{\max}$ denote the index of the last iteration of the AFT-GEE fitting procedure. The estimating equation used to estimate the blip parameters can be written in the following way:

$$\begin{aligned}
&\sum_{i=1}^{r} \mathbf{D}_i^\top [\mathbf{V}_i(\hat{\boldsymbol{\lambda}}^{(t_{\max})})]^{-1} \, \mathbf{W_i} \Big(\log(\mathbf{T}_i) - \mathbf{X}_{i,\boldsymbol{\beta_k}} \boldsymbol{\beta_k} - \mathbf{A} \mathbf{X}_{i,\boldsymbol{\psi_k}} \boldsymbol{\psi_k} \Big), \\
&= \sum_{i=1}^{r} \mathbf{D}_i^\top [\mathbf{V}_i(\hat{\boldsymbol{\lambda}}^{(t_{\max})})]^{-1} \, \Big(\log(\mathbf{T}_i^w) - \mathbf{X}^w_{i,\boldsymbol{\beta_k}} \boldsymbol{\beta_k} - \mathbf{A}^w \mathbf{X}^w_{i,\boldsymbol{\psi_k}} \boldsymbol{\psi_k} \Big),
\end{aligned}$$

which, by DWSurv theory, is equivalent to estimating $(\boldsymbol{\beta}_k, \boldsymbol{\psi}_K)$ by generalized least squares using the weighted dataset $(T^w, \boldsymbol{X}^w, A^w)$, where the covariates $\boldsymbol{X}^w$ are independent of the treatment and censoring mechanism (A^w, Δ^w). By standard linear regression results, this is sufficient to account for possible omitted variable bias in $\boldsymbol{\psi}_k$ due to misspecification of the treatment-free model for any fixed variance-covariance matrices $\mathbf{V}_i(\hat{\boldsymbol{\lambda}}^{(t_{\max})}), \quad i = 1, \ldots, r$. Thus, the procedure yields consistent estimators of the blip parameters $\boldsymbol{\psi}_k$ when the chosen weight function w_k satisfies the DWSurv balancing property (Simoneau et al., 2020).

Estimation of ITR metrics

For a given regime regime d, we estimate its POT and value via the following unbiased estimating equations in p, μ:

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{P(\Delta = 1 | \boldsymbol{X}_i, A_i)} \Big(\mathbb{1}(d(\boldsymbol{X}_i) = d_{\mathrm{o}}(\boldsymbol{X}_i, K_i)) - p \Big) = 0, \\
\frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{P(\Delta = 1 | \boldsymbol{X}_i, A_i)} \Big(\log T_i + (d(\boldsymbol{X}_i) - A_i) \boldsymbol{X}_{\psi_{K_i}} \psi_{K_i} - \mu \Big) = 0.
\end{aligned}$$

We now show that the two equations above are unbiased for the parameters of interest. For the POT of a given ITR d, we have that

$$\begin{aligned}
&\mathbb{E}\Big[\frac{\Delta}{P(\Delta = 1 | \boldsymbol{X}, A)} \Big(\mathbb{1}(d(\boldsymbol{X}) = d_{\mathrm{o}}(\boldsymbol{X}, K)) - p \Big) \Big] \\
&= \mathbb{E}\left[\mathbb{E}\Big[\frac{\Delta}{P(\Delta = 1 | \boldsymbol{X}, A)} \Big(\mathbb{1}(d(\boldsymbol{X}) = d_{\mathrm{o}}(\boldsymbol{X}, K)) - p \Big) \Big| \boldsymbol{X}, A, K \Big] \right] \\
&= \mathbb{E}\left[\Big(\mathbb{1}(d(\boldsymbol{X}) = d_{\mathrm{o}}(\boldsymbol{X}, K)) - p \Big) \mathbb{E}\Big[\frac{\Delta}{P(\Delta = 1 | \boldsymbol{X}, A)} \Big| \boldsymbol{X}, A, K \Big] \right] \\
&= \mathbb{E}\left[\Big(\mathbb{1}(d(\boldsymbol{X}) = d_{\mathrm{o}}(\boldsymbol{X}, K)) - p \Big) \frac{P(\Delta = 1 | \boldsymbol{X}, A, K)}{P(\Delta = 1 | \boldsymbol{X}, A)} \right] \\
&= \mathbb{E}\left[\Big(\mathbb{1}(d(\boldsymbol{X}) = d_{\mathrm{o}}(\boldsymbol{X}, K)) - p \Big) \frac{P(\Delta = 1 | \boldsymbol{X}, A)}{P(\Delta = 1 | \boldsymbol{X}, A)} \right] \qquad \text{by assumption 5} \\
&= \mathbb{E}\left[\mathbb{1}(d(\boldsymbol{X}) = d_{\mathrm{o}}(\boldsymbol{X}, K)) \right] - p \\
&= 0,
\end{aligned}$$

for $p = \mathrm{POT}(d)$.

Similarly, for the value of a regime d,

$$\begin{aligned}
&\mathbb{E}\left[\frac{\Delta}{P(\Delta=1|\boldsymbol{X},A)}\Big(\log T+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu\Big)\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\frac{\Delta}{P(\Delta=1|\boldsymbol{X},A)}\Big(\log T+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu\Big)\Big|\boldsymbol{X},A,K,\Delta\right]\right]\\
&=\mathbb{E}\left[\frac{\Delta}{P(\Delta=1|\boldsymbol{X},A)}\mathbb{E}\left[\Big(\log T+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu\Big)\Big|\boldsymbol{X},A,K,\Delta\right]\right].
\end{aligned}$$

Then, we have the following identity:

$$\begin{aligned}
&\mathbb{E}\left[\Big(\log T+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu\Big)\Big|\boldsymbol{X},A,K,\Delta\right]\\
&=\mathbb{E}\left[\log T|\boldsymbol{X},A,K,\Delta\right]+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu\\
&=\mathbb{E}\left[\log T(A,K)|\boldsymbol{X},A,K,\Delta\right]+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu && \text{by assumptions } 2,3\\
&=\mathbb{E}\left[\log T(A,K)|\boldsymbol{X},A,K\right]+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu && \text{by assumption } 4\\
&=\mathbb{E}\left[\log T|\boldsymbol{X},A,K\right]+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu\\
&=\boldsymbol{X}_{\beta_K}\boldsymbol{\beta}_K+A\boldsymbol{X}\boldsymbol{\psi}_K+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu\\
&=\boldsymbol{X}_{\beta_K}\boldsymbol{\beta}_K+d(\boldsymbol{X})\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu\\
&=\mathbb{E}\left[\log T\Big|\boldsymbol{X},A=d,K\right]-\mu\\
&=\mathbb{E}\left[\log T(d,K)|\boldsymbol{X}\right]-\mu.
\end{aligned}$$

Substituting this component back into the original equation gives:

$$\begin{aligned}
&\mathbb{E}\left[\frac{\Delta}{P(\Delta=1|\boldsymbol{X},A)}\mathbb{E}\left[\Big(\log T+(d(\boldsymbol{X})-A)\boldsymbol{X}_{\psi_K}\boldsymbol{\psi}_K-\mu\Big)\Big|\boldsymbol{X},A,K,\Delta\right]\right]\\
&=\mathbb{E}\left[\frac{\Delta}{P(\Delta=1|\boldsymbol{X},A)}\left(\mathbb{E}\left[\log T(d,K)|\boldsymbol{X}\right]-\mu\right)\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\frac{\Delta}{P(\Delta=1|\boldsymbol{X},A)}\left(\mathbb{E}\left[\log T(d,K)|\boldsymbol{X}\right]-\mu\right)\Big|\boldsymbol{X},A\right]\right]\\
&=\mathbb{E}\left[\left(\mathbb{E}\left[\log T(d,K)|\boldsymbol{X}\right]-\mu\right)\mathbb{E}\left[\frac{\Delta}{P(\Delta=1|\boldsymbol{X},A)}\Big|\boldsymbol{X},A\right]\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\log T(d,K)|\boldsymbol{X}\right]\right]-\mu\\
&=0,
\end{aligned}$$

for $\mu=\mathbb{E}[\log T(d,K)]$.

Extension to treatment-dependent causes

We can derive alternative definitions of the optimal ITR if we replace assumptions 3 and 5 of section 2.1 by the following weaker assumption:

*3. Sequential ignorability (Imai et al., 2010): for all $a\in\mathcal{A}$ and all $k=1,\ldots,\kappa$,

$$\begin{aligned}
T(a,k)&\perp A|\boldsymbol{X},\\
T(a,k)&\perp K|\boldsymbol{X},A.
\end{aligned}$$

Under this assumption, the mediation formula of Pearl (2012) yields the following definition for the weighted ITR:

$$\begin{aligned} d_{\mathrm{w}}^{\mathrm{opt}} &= \arg\max_{d} \mathbb{E}\Big[f(T(d,K))\Big] \\ &= \arg\max_{d} \mathbb{E}\Big[\mathbb{E}[f(T(d,K))|\boldsymbol{X}]\Big] \\ &= \arg\max_{d} \mathbb{E}\left[\sum_{k=1}^{\kappa} P\left(K=k|\boldsymbol{X}, A=d(\boldsymbol{X})\right) \mathbb{E}[f(T)|\boldsymbol{X}, A=d(\boldsymbol{X}), K=k]\right]. \end{aligned}$$

Letting $\varphi_k(\boldsymbol{x}, a) = P(K=k|\boldsymbol{X}=\boldsymbol{x}, A=a)$ and $Q_k(\boldsymbol{x}, a) = \mathbb{E}[f(T)|\boldsymbol{X}=\boldsymbol{x}, A=a, K=k]$, we have the equivalent form:

$$d_{\mathrm{w}}^{\mathrm{opt}} = \arg\max_{d} \mathbb{E}\left[\sum_{k=1}^{\kappa} \varphi_k(\boldsymbol{X}, d(\boldsymbol{X})) Q_k(\boldsymbol{X}, d(\boldsymbol{X}))\right],$$

which gives rise to the explicit rule

$$d_{\mathrm{w}}(\boldsymbol{x}) = \arg\max_{a} \sum_{k=1}^{\kappa} \varphi_k(\boldsymbol{x}, a) Q_k(\boldsymbol{x}, a), \quad \boldsymbol{x} \in \mathcal{X}.$$

References

K. Imai, L. Keele, and D. Tingley. A general approach to causal mediation analysis. *Psychological Methods*, 15:309–34, 2010.

J. Pearl. The causal mediation formula–a guide to the assessment of pathways and mechanisms. *Prevention Science*, 13:426–36, 2012.

G. Simoneau, E. E. M. Moodie, J. S. Nijjar, R. W. Platt, and the Scottish Early Rheumatoid Arthritis Inception Cohort Investigators. Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association*, 115(531):1531–1539, 2020.