Estimating optimal individualized treatment rules with multistate processes

Giorgos Bakoyannis
Department of Biostatistics and Health Data Science, Indiana University, Indianapolis, Indiana, USA

Abstract
Multistate process data are common in studies of chronic diseases such as cancer. These data are ideal for precision medicine purposes as they can be leveraged to improve more refined health outcomes, compared to standard survival outcomes, as well as incorporate patient preferences regarding quantity versus quality of life. However, there are currently no methods for the estimation of optimal individualized treatment rules with such data. In this paper, we propose a nonparametric outcome weighted learning approach for this problem in randomized clinical trial settings. The theoretical properties of the proposed methods, including Fisher consistency and asymptotic normality of the estimated expected outcome under the estimated optimal individualized treatment rule, are rigorously established. A consistent closed-form variance estimator is provided and methodology for the calculation of simultaneous confidence intervals is proposed. Simulation studies show that the proposed methodology and inference procedures work well even with small-sample sizes and high rates of right censoring. The methodology is illustrated using data from a randomized clinical trial on the treatment of metastatic squamous-cell carcinoma of the head and neck.

Keywords
multistate model; outcome-weighted learning; patient preferences; precision medicine; response

1 | INTRODUCTION
Modern precision medicine acknowledges the heterogeneity of the majority of human diseases and aims to develop and deliver therapies that are tailored to the individual patient. At the heart of these efforts is the development of data-driven individualized treatment assignment rules (Kosorok & Laber, 2019). The purpose of such rules is to provide the right treatment to a given patient (Lipkovich et al., 2017) and, thereby, to improve health
outcomes among patients overall. Over the past decade, there has been a large number of new statistical methods for the estimation of optimal individualized treatment rules (ITRs) with various types of outcomes, including continuous, binary, survival, and competing risks outcomes. However, to the best of our knowledge, there are currently no methods for the estimation of optimal ITRs with stochastic processes that evolve through multiple discrete states over (continuous) time, also known as multistate processes. Such processes are commonly encountered in studies of chronic diseases, such as cancer and HIV infection, where disease evolution is often characterized by multiple discrete health states. An example is the SPECTRUM trial (Vermorken et al., 2013), a phase III randomized clinical trial on recurrent or metastatic squamous-cell carcinoma of the head and neck, where patient event history was characterized by the states “tumor response” (i.e., tumor shrinkage per Therasse et al. 2000), “disease progression”, and “death.” This paper addresses the issue of optimal ITR estimation with right-censored multistate processes data in randomized clinical trial settings.

Multistate process data are ideal for precision medicine purposes as they can be leveraged to improve more refined health outcomes (compared to, e.g., overall or progression-free survival) as well as incorporate patient preferences. Typically, such data contain information about one or more transient health states (e.g., tumor response), that is not fully captured by standard survival or competing risks data. Therefore, they provide more comprehensive information about more refined health outcomes that may reflect both quantity and quality of life. Improvement of such outcomes may be more desirable than mere life extension to both patients and clinicians. In oncology, for example, such a desirable health outcome is sustained tumor response, which is associated with better quality of life, extended progression-free survival (PFS) time, and a prolonged treatment-free interval (Kaufman et al., 2017). Tumor response was also a desirable outcome in the SPECTRUM trial. Another important outcome is extended quality-adjusted lifetime, which can be defined as the weighted sum of the times spent in a set of desirable health states (Oza et al., 2020; Pelzer et al., 2017). Quality-adjusted lifetime can be defined in multiple different ways according to patient preferences by using different weighting schemes that reflect the needs of different groups of patients. In this way, individual priorities toward quality versus quantity of life can be taken into account.

The methods for optimal ITR estimation can be classified into two broad categories: (i) backward induction methods such as Q-learning (Murphy, 2005; Zhao et al., 2009) and A-learning (Murphy, 2003; Robins, 2004) and (ii) direct-search methods, also known as policy-search methods, such as outcome weighted learning (Zhao et al., 2012) and value search estimation (Zhang et al., 2012). The first class of methods typically estimates the optimal ITR by modeling either the conditional expectation of the outcome given the covariates (Q-learning) or the interactions between treatment and the covariates (A-learning). The second class of methods estimates the optimal ITR directly by optimizing an appropriate objective function. Given that imposing a realistic model for the conditional mean outcome or the treatment interactions is difficult for general multistate processes, the focus in this paper is on direct-search methods. The issue of ITR estimation with survival outcomes has been well addressed in the literature. Zhao et al. (2015) extended the outcome-weighted learning framework of Zhao et al. (2012) to situations, where the outcome of interest
is a right-censored survival time by proposing (i) an inverse censoring weighting (ICO) and (ii) a doubly robust (DR) outcome-weighted learning approach. Estimation in these approaches relies on a weighted version of support vector machines (Cortes & Vapnik, 1995). A similar method for ITR estimation with survival outcomes was proposed by Bai et al. (2017). A value search estimation approach for maximizing the $t$-year survival probability was developed by Jiang et al. (2017). This method approximates the nonsmooth objective function of the problem by a smooth one using kernel smoothing. Recently, there has also been some interest in the issue of ITR estimation with competing risk outcomes (He et al., 2021; Zhou et al., 2021). Nevertheless, the methods for survival and competing risk outcomes cannot be used for the estimation of optimal ITRs with right-censored multistate processes.

In this work, we extend the outcome-weighted learning framework (Zhao et al., 2012, 2015) to deal with situations, where the outcome of interest is an arbitrary right-censored multistate process, incorporating also patient preferences. The proposed method does not impose Markov assumptions or model assumptions on the multistate process of interest. The novelty of this paper is twofold. First, we devise a novel objective function which, in contrast to the ICO approach by Zhao et al. (2015), utilizes information from the censored cases. Importantly, this is achieved without imposing and estimating a model for the conditional expectation of the outcome given treatment and the covariates, as opposed to the DR approach by Zhao et al. (2015). Second, in addition to showing Fisher and universal consistency of the proposed method, we establish the asymptotic distribution of the proposed estimator for the expected outcome under the estimated optimal ITR and derive a consistent closed-form variance estimator. Based on our theoretical results, we also propose a method for the calculation of simultaneous confidence intervals over a set of patient preferences to account for multiplicity. The simulation studies provide evidence that the proposed estimator and inference procedures work well even with small-sample sizes and under high rates of right censoring. Furthermore, the simulation studies reveal that the proposed method performs better than the previously proposed ICO and DR approaches for censored failure times.

2 | METHODOLOGY

2.1 | Notation and data

Consider a multistate process $\{X(t) : t \in [0, \tau]\}$ with (finite) state space $S = \{1, \ldots, S\}$, where $\tau < \infty$ is the maximum observation time. Let $Z \in Z \subset \mathbb{R}^p$ denote a vector of baseline covariates that may be related to the effect of treatment on the multistate process of interest. For simplicity, the treatment $A$ is considered to be a binary variable taking its values in the treatment set $\{-1, 1\}$. With multistate processes, the goal of treatment is typically to prolong the time spent in a set of desirable health states. Given that some health states may be more desirable than others and that patient preferences may vary, we define the benefit processes

$$\{Y_w(t) = w'X(t) : t \in [0, \tau], w \in \mathbb{W}\},$$
where \( \bar{X}(t) = (I\{X(t) = 1\}, ..., I\{X(t) = S\})' \), \( w \) is an \( S \)-dimensional vector of preference weights that satisfies \( 0 \leq w \leq 1 \) and the sum of its elements is positive and less than \( S \), and \( \mathcal{W} = \{w_1, ..., w_u\} \) is a prespecified finite set of preference weight vectors that reflect different patient preferences/priorities. Based on the latter processes, we define the utilities based on the latter processes, we define the utilities

\[
\int_0^\tau Y_{w}(t)dm(t), \; w \in \mathcal{W},
\]

(1)

where the integrator function is \( m(t) = t \) and induces the Lebesgue measure on the Borel \( \sigma \)-algebra \( \mathcal{B}([0, r]) \). Since \( \int_0^\tau I\{X(t) = j\}dm(t) \) is the time spent in the \( j \)th state during the time interval \([0, r]\), the utilities (1) represent weighted sums of the (restricted) times spent in each state. For example, consider the states “initial disease state” (state 1), “tumor response” (state 2), and “disease progression or death” (state 3), in the setting of the SPECTRUM trial mentioned in Section 1. A potential choice for the set of preference weights in this example is \( \mathcal{W} = \{0, 1, 0', 0.5, 1, 0', 1, 1, 0'\} \). When \( w = (0, 1, 0') \), the utility \( \int_0^\tau Y_{w}(t)dm(t) = \int_0^\tau I\{X(t) = 2\}dm(t) \) corresponds to the restricted time spent in the tumor response state. The utility under the choice \( w = (0.5, 1, 0') \) corresponds to the restricted quality-adjusted lifetime \( \int_0^\tau Y_{w}(t)dm(t) = 0.5 \times \int_0^\tau I\{X(t) = 1\}dm(t) + \int_0^\tau I\{X(t) = 2\}dm(t) \). This utility represents the restricted (quality-adjusted) time lived, where the time spent in the initial disease state (i.e., without tumor response) is reduced by 50% to reflect a quality of life loss due to disease symptoms and/or side effects due to treatment continuation. When \( w = (1, 1, 0') \), the utility is the (restricted) PFS time, in which the times lived in the initial disease state and the tumor response state are equally important. Different choices of \( w \) reflect different patient priorities towards quality versus quantity of life.

In practice, the multistate and benefit processes are not fully observed for all individuals due to the usual right censoring. Let \( C \) denote the right censoring time and \( T^* \) the time of arrival at an absorbing state (i.e., death). Letting \( T = T^* \land \tau \), with \( a \land b = \min(a, b) \) for any \( a, b \in \mathbb{R} \), \( \bar{T} = T \lor C \), and \( \Delta_s = I\{T \leq C\} \), the observed data consist of independent and identically distributed copies of \( D_i = (Z_i, A_i, \bar{T}_i, \Delta_{s}, \{X_i(t)I(C_i \geq T \land t) : t \in [0, r]\}) \), \( i = 1, ..., n \), where \( X_i(t)I(C_i \geq T \land t) \), \( t \in [0, r] \), is the censored version of the multistate process which is equal to 0 for the censored individuals after their censoring time. Note that \( Y_{w}(t) = w' \{I\{X(t) = 1\}, ..., I\{X(t) = S\}\}' \) is the benefit process for the \( i \)th individual.

### 2.2 Optimal individualized treatment rule estimation with multistate processes

An ITR is a deterministic function \( d : \mathcal{X} \mapsto \{-1, 1\} \) which suggests treatment choice \( d(z) \in \{-1, 1\} \) for a patient with covariates \( Z = z \). Since the estimation of an optimal ITR is essentially a causal inference problem, we utilize the potential outcomes causal framework (Rubin, 2005; Tsiatis et al., 2019). Let \( Y_{w}(t; 1) \) and \( Y_{w}(t; -1) \), \( w \in \mathcal{W} \), denote the potential benefit processes under treatment choices 1 and −1, respectively, at time \( t \in [0, r] \). Since the potential outcomes cannot be directly observed in real-world settings, we need to impose the following causal assumptions.
A1. Stable unit treatment value assumption: $Y_a(t) = Y_a(t; 1)I(A = 1) + Y_a(t; -1)I(A = -1)$, $t \in [0, \tau]$.

A2. Independent treatment assignment assumption: \{$Y_a(t; 1), Y_a(t; -1), Z$\} \iid $A$, $t \in [0, \tau]$.

A3. Positivity assumption: $\pi_0 = P(A = 1) \in [c_0, c_1]$, with $0 < c_1 < c_2 < 1$.

Assumptions A2 and A3 are automatically satisfied in randomized clinical trials. It must be highlighted that assumption A1 implies that the observed benefit process \{$Y_a(t)$ : $t \in [0, \tau]$\}, unlike the potential benefit processes \{$Y_a(t; 1), Y_a(t; -1)$ : $t \in [0, \tau]$\}, is associated with treatment $A$. Now, we can define the potential benefit processes under an ITR $d$ as

$$Y_a(t; d) = Y_a(t; 1)I\{d(Z) = 1\} + Y_a(t; -1)I\{d(Z) = -1\},$$

$w \in \mathcal{W}, t \in [0, \tau]$.

These processes are essential for defining potential benefit under an ITR for multistate processes. For such processes, we define the value functions as

$$\mathcal{V}_a(d) = E\left\{\int_0^\tau Y_a(t; d)dm(t)\right\}, w \in \mathcal{W},$$

where $\mathcal{V}_a(d)$ is the expected sum of the weighted (under $w$) times spent in each state of the process during the time interval $[0, \tau]$, under the ITR $d$. Under assumptions A1–A3, the independent right censoring assumption, and the fact that

$$\frac{Y_a(t)I(C \geq T \land t)}{\exp\{-\Lambda_0(T \land t)\}} = \frac{Y_a(t)I(C \geq T \land t)}{\exp\{-\Lambda_0(T \land t)\}}, t \in [0, \tau], w \in \mathcal{W},$$

it can be shown that the value functions can be expressed in terms of the observable data as

$$\mathcal{V}_a(d) = E\left\{\int_0^\tau \frac{Y_a(t)I(C \geq T \land t)}{\exp\{-\Lambda_0(T \land t)\}}dm(t)\left\{\frac{I(A = d(Z))}{\Lambda_0 + (1 - A)/2}\right\}\right\}, w \in \mathcal{W},$$

where $\Lambda_0(t)$ is the cumulative hazard function of the right censoring variable $C$ at time $t$. An optimal ITR $d^*_a, w \in \mathcal{W}$, is a maximizer of the corresponding value function, that is,

$$d^*_a \in \arg\max_d \mathcal{V}_a(d).$$

Given that any rule $d : \mathcal{X} \rightarrow \{-1, 1\}$ can be expressed as $d(z) = \text{sgn}\{f(z)\}$, where $\text{sgn}(x) = I(x \geq 0) - I(x < 0)$, for some measurable function $f : \mathcal{X} \rightarrow \mathbb{R}$, and since

$$\mathcal{V}_a(d) = E\left\{\int_0^\tau \frac{Y_a(t)I(C \geq T \land t)}{\exp\{-\Lambda_0(T \land t)\}}dm(t)\left\{\frac{1}{\Lambda_0 + (1 - A)/2}\right\}\right\} - E\left\{\int_0^\tau \frac{Y_a(t)I(C \geq T \land t)}{\exp\{-\Lambda_0(T \land t)\}}dm(t)\left\{\frac{I(A \neq d(Z))}{\Lambda_0 + (1 - A)/2}\right\}\right\},$$
an optimal ITR is \( d_u(z) = \text{sgn}(f'_u(z)) \), where \( f'_u \) is a minimizer of the risk function

\[
R_u(f) = \mathbb{E} \left[ \int_0^\tau Y_u(t) I(C \geq T \land t) \exp(-A(T \land t)) \, dm(t) \right]
\]

\[
= \mathbb{E} \left[ \int_0^\tau Y_u(t) I(C \geq T \land t) \exp(-A(T \land t)) \, dm(t) \right] \mathbb{P}(A < 0)
\]

over all measurable functions \( f : \mathcal{X} \to \mathbb{R} \). It is not hard to see that

\[
\mathcal{V}_u(\text{sgn}(f'_u)) - \mathcal{V}_u(\text{sgn}(f)) = R_u(f) - R_u(f'_u), \quad w \in \mathcal{W},
\]

for any measurable \( f \). Minimizing \( R_u(f) \) over all measurable functions is clearly not feasible. Therefore, we will consider minimization over a subset of the class of all measurable functions \( f : \mathcal{X} \to \mathbb{R} \). In particular, we consider either of the following subsets.

- Class of linear functions \( \{ f(\cdot) = \beta_0 + \langle \beta, \cdot \rangle : \beta_0 \in \mathbb{R}, \beta_i \in \mathbb{R}^p \} \), where \( \langle \beta_i, z \rangle = \beta_i z \).

- Reproducing kernel Hilbert space (RKHS) \( \mathcal{H} \) with kernel \( k \), which is the completion of the space \( \{ f(\cdot) = \sum_{j=1}^m a_j k(\cdot, z_j) : m \in \mathbb{N}, z_j \in \mathcal{X}, a_j \in \mathbb{R} \} \).

Here, we consider the RKHS with the Gaussian kernel, \( k(z_i, z) = \exp(-\sigma \| z_i - z_j \|^2) \), \( z_i, z_j \in \mathcal{X} \).

Minimizing the empirical version of \( R_u(f) \), over the chosen class \( \mathcal{H} \), is challenging because it involves a discontinuous and nonconvex function of \( f \). To alleviate, we follow the paradigm of outcome weighting learning (Zhao et al., 2012, 2015) and support vector machines (Steinwart & Christmann, 2008) and utilize the surrogate risk

\[
\hat{R}_{\text{sgn}}(f) = \mathbb{E} \left[ \int_0^\tau Y_u(t) I(C \geq T \land t) \exp(-A(T \land t)) \phi(A(Z)) \, dm(t) \right] \mathbb{P}(A < 0) + (1 - A)^2
\]

where \( \phi(x) = \max(0, 1 - x) \) is the hinge loss, which is convex in \( f \). The cumulative hazard function \( \Lambda_n \) of the right censoring distribution can be estimated using the nonparametric Nelson–Aalen estimator \( \hat{\Lambda}_n(t) = \sum_{i=1}^n \int_0^t \{ \sum_{u=1}^n Y_i(u) \}^{-1} dN_i(u), \quad t \in [0, \infty) \),

where \( N_i(t) = (1 - \Delta_i)I(T_i \leq t) \) and \( Y_i(t) = I(T_i > t) \). An obvious estimator of \( \pi_0 \) is \( \hat{\pi}_n = n^{-1} \sum_{i=1}^n I(A_i = 1) \). Thus, the empirical version of the surrogate risk \( \hat{R}_{\text{sgn}} \) is

\[
\hat{R}_{\text{sgn}}(f) = \frac{1}{n} \sum_{i=1}^n \left[ \mathbb{E} \left[ \int_0^\tau Y_u(t) I(C_i \geq T_i \land t) \exp(-A(T_i \land t)) \phi(A(f(Z))) \, dm(t) \right] \mathbb{P}(A < 0) + (1 - A)^2 \right]
\]

Note that, even though \( \pi_0 \) is known in clinical trials, the estimate \( \hat{\pi}_n \) is used in Equation (2) as this typically leads to some efficiency gain in inverse probability weighting type estimators.
The empirical surrogate risk for the ICO estimator for censored failure times (e.g., $\bar{T}$) by Zhao et al. (2015) for the setting considered here is

$$\frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\Delta_i \bar{T}_i}{\exp[-\Lambda_\hat{n}(\bar{T}_i)]} \phi(A_i f(Z_i)) \right] A_i \pi_\hat{n} + \left(1 - A_i\right)/2.$$

The latter incorporates inverse censoring weighting in the censored time of interest and discards the censored observations ($\Delta_i = 0$). In contrast, the proposed empirical surrogate risk (2) utilizes information from both uncensored and censored cases by incorporating inverse censoring weighting in the underlying stochastic process \(\{Y_w(t) : t \in [0, \tau]\}\). Importantly, this is achieved without imposing and estimating a model for the conditional expectation of the time of interest given \(A\) and \(Z\), as opposed to the DR approach (Zhao et al., 2015).

Simulation studies summarized in Section 4 reveal that these characteristics of the proposed method lead to a better performance compared to the ICO and DR estimators (Zhao et al., 2015) for censored failure times.

Similarly to Zhao et al. (2012, 2015), the estimators of the optimal decision functions within the class \(\mathcal{F}\) are obtained using penalized minimization of \(\mathcal{R}_{\phi, w}\) as

$$\hat{f}_{\lambda, w} = \arg \min_{f \in \mathcal{F}} \{ \mathcal{R}_{\phi, w}(f) + \lambda_n \| f \|^2 \}, \ w \in \mathcal{W},$$

where \(\lambda_n\) is a positive tuning parameter that controls the complexity of \(f\) and \(\| \cdot \|\) is a norm on the chosen class \(\mathcal{F}\). For example, \(\| \cdot \|\) is the Euclidean norm if \(\mathcal{F}\) is the class of linear functions. For notational simplicity, we omit the subscript \(\lambda_n\) from the estimated decision functions and use the more compact notation \(\hat{f}_{\lambda, w}\). Based on the estimated decision functions \(\{\hat{f}_{\lambda, w} : w \in \mathcal{W}\}\) the estimated optimal ITRs are

$$\hat{d}_{\lambda, w}(z) = \text{sgn}(\hat{f}_{\lambda, w}(z)), \ w \in \mathcal{W}, \ z \in \mathcal{Z}.$$

Based on the class of estimated ITRs \(\{\hat{d}_{\lambda, w} : w \in \mathcal{W}\}\), treatment assignment for a given patient utilizes \(\hat{d}_{\lambda, w}\) for the weight vector \(w\) that is closest, with respect to the Euclidean norm, to the preference weight \(w_0\) that reflects the patient’s own preferences/priorities.

### 2.3 Estimation of the value function

The value function of an ITR \(d\) can be estimated as

$$\mathcal{F}_{\lambda, w}(d) = \frac{1}{n} \sum_{i=1}^{n} \left[ \int_0^\tau Y_w(t) I(C_i \geq T_i \wedge t) \frac{d\mu(t)}{\exp[-\Lambda_\hat{n}(T_i \wedge t)]} \right] \times I(A_i = d(Z_i)) \left( \frac{A_i \pi_\hat{n}}{A_i \pi_\hat{n} + (1 - A_i)/2} \right), \ w \in \mathcal{W}.$$

*Biometrics. Author manuscript; available in PMC 2024 February 16.*
The value $\tilde{V}_{n,w}(d)$ can be seen as the (estimated) performance of the ITR $d$ under the preference weight vector $w$. The estimated value of the estimated optimal ITR $\hat{V}_{n,w}(d_{n,w})$ is expected to be an over-optimistic estimate of the performance of $\hat{d}_{n,w}$ in a future (out of sample) patient when sample size is small relatively to $p$ or relatively to the complexity of $\mathcal{F}$. This is because the estimator $\hat{V}_{n,w}(d_{n,w})$ uses the same set of data $\{D_i : i = 1, \ldots, n\}$, for both the estimation of the optimal ITR and the evaluation of its performance. This phenomenon is similar to the behavior of the training error in support vector machines (Hastie et al., 2009). A better estimate of $V_{w}(\hat{d}_{n,w})$ in finite samples is expected to be the jackknife or leave-one-out cross-validation value estimator

$$\hat{V}_{n,w}^{\text{jk}}(d_{n,w}) = \frac{1}{n} \sum_{i=1}^{n} \left[ \int_{0}^{\tau} Y_{n,w}(t) I(C_i \geq T_i \wedge t) dm(t) \right] \times \frac{I(A_i = d_{n,w}^{-}(Z_i))}{A_i x_n + (1 - A_i)^2}, \ w \in \mathcal{W},$$

where $d_{n,w}^{-} = \text{sgn}(\hat{f}_{n,w})$ is the optimal ITR estimated under the preference weight vector $w$ using all but the data of the $i$th individual.

### 3 | THEORETICAL PROPERTIES

The first theorem justifies the use of the surrogate risk $\mathcal{R}_{\phi,w}$ instead of the original risk $\mathcal{R}_{w}$, for the estimation of optimal ITR $d_{w}^*$, $w \in \mathcal{W}$.

**Theorem 1 (Fisher consistency).**

Suppose that assumptions A1–A3 and condition C1 in Web Appendix A hold. Then, for any $w \in \mathcal{W}$, if $\hat{f}_{w}$ minimizes $\mathcal{R}_{\phi,w}$, $d_{w}(z) = \text{sgn}(\hat{f}_{w}(z))$ for all $z \in \mathcal{X}$.

The proof of Theorem 1 can be found in Web Appendix A.1. The next theorem ensures that $\mathcal{R}_{\phi,w}(\hat{f}_{n,w})$, which is the true surrogate risk of the estimated decision function $\hat{f}_{n,w}$, converges (in probability) to the minimal surrogate risk over the chosen class $\mathcal{F}$. It also asserts that, if the chosen class $\mathcal{F}$ is appropriate, then the proposed estimator $\hat{d}_{n,w}$ is universally consistent, that is, its value converges (in probability) to the optimal value $V_{w}(d_{w}^*)$.

**Theorem 2.**

Suppose that assumptions A1–A3 and conditions C1, C3, and C4 in Web Appendix A hold. Then, for $\lambda_n > 0$ with $\lambda_n \rightarrow 0$, and $n \lambda_n \rightarrow \infty$,

$$\left| \mathcal{R}_{\phi,w}(\hat{f}_{n,w}) - \inf_{f \in \mathcal{F}} \mathcal{R}_{\phi,w}(f) \right|_{p} \rightarrow 0, \ w \in \mathcal{W}.$$
as \( n \to \infty \), for any distribution \( P \) of the data. Moreover, if (i) \( \mathcal{F} \) is the space of linear functions and \( f'_{w} \in \mathcal{F} \) or (ii) \( \mathcal{F} \) is the RKHS with the Gaussian kernel and the marginal distribution \( \mu \) of \( Z \) is regular, then

\[
\| \mathcal{V}_w(\hat{d}_n) - \mathcal{V}_w(d'_w) \| \to 0, \quad \text{as} \quad n \to \infty .
\]

Theorem 2 is proved in Web Appendix A.2. When \( \mathcal{F} \) is the space of linear functions, universal consistency requires that the optimal decision function \( f'_{w} \) is linear, that is, \( f'_{w} \in \mathcal{F} \). This requirement can be made more plausible by considering an enlarged covariate space \( \mathcal{Z} \) that includes polynomial terms and/or two-way interaction terms between the original covariates \( Z \). If \( f'_{w} \notin \mathcal{F} \), \( \mathcal{V}_w(\hat{d}_n) \) is expected to converge to a value that is lower than the optimal value \( \mathcal{V}_w(d'_w) \). Nevertheless, the limit of \( \mathcal{V}_w(\hat{d}_n) \) can be seen as an approximation to the optimal value \( \mathcal{V}_w(d'_w) \) by the first statement of Theorem 2 and the fact that \( \mathcal{R}_w(f') \leq \mathcal{R}_w(f) \leq \mathcal{R}_w(f) \) for any \( f \in \mathcal{F} \), since the hinge loss satisfies \( \phi(x) \geq I(x < 0) \) for all \( x \in \mathbb{R} \).

Interestingly, when \( \mathcal{F} \) is the RKHS with the Gaussian kernel and the marginal distribution of \( Z \) is regular, the estimated ITR is always universally consistent. However, the so-called no-free-lunch theorem (Steinwart & Christmann, 2008) implies that the corresponding rate of convergence can be extremely slow for at least some distributions of the data. This means that an extremely large sample size may be required in practice in order to obtain an ITR \( \hat{d}_n \) with a value reasonably close to the optimal value. For this reason, we will restrict our attention to the case where \( \mathcal{F} \) is the space of linear functions in the remainder of this paper.

The next theorem characterizes the asymptotic distribution of the estimated value function of a fixed decision function \( f \).

**Theorem 3.**

Under assumptions A1–A3 and conditions C1, C2, and C4 in Web Appendix A, we have that

\[
G_{w}(f) : = \sqrt{n}(\hat{\mathcal{P}}_{w}(\text{sgn}(f)) - \mathcal{V}_w(\text{sgn}(f)))
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{w}(f) + \epsilon_{w}(f), \quad w \in \mathcal{W},
\]

for any given measurable decision function \( f \), where the explicit formula for the influence function \( \psi_{w}(f) \) is provided in Web Appendix B and \( \epsilon_{w}(f) = o(1) \). Moreover, if \( \mathcal{F} \) is the space of linear functions, then \( \sup_{f \in \mathcal{F}} | \epsilon_{w}(f) | = o(1) \) and the class of influence functions \( \{ \psi_{w}(f) : f \in \mathcal{F} \} \) is \( P \)-Donsker.

The proof of Theorem 3 is given in Web Appendix A.3. This theorem implies that, for any measurable decision function \( f \), \( G_{w}(f) \) is asymptotically normal with mean zero and variance \( \sigma_{w}^{2}(f) = E\psi_{w}^{2}(f) \). This asymptotic variance can be consistently (in probability)
Theorem 4.

Suppose that \( F \) is the space of linear functions. Then, under assumptions A1–A3, conditions C1–C5 in Web Appendix A, the additional assumption that the optimal linear decision function \( \tilde{f}_{n,w}(z) \) satisfies \( P(\tilde{f}_{n,w}(Z) = 0) = 0, w \in W \), and for \( \lambda_n > 0 \) with \( \lambda_n \to 0 \) and \( n\lambda_n \to \infty \), we have that

\[
\max_{w \in W} \left| G_{n,w}(\tilde{f}_{n,w}) - G_{n,w}(\tilde{f}_n) \right| = o(1).
\]

The proof of Theorem 4 is given in Web Appendix A.4. The last theorem and Theorem 3 imply that \( G_{n,w}(\tilde{f}_{n,w}) \) is asymptotically normal with zero mean and variance \( \hat{\sigma}_n^2(\tilde{f}_n) \). This variance can be consistently (in probability) estimated by \( \hat{\sigma}_n^2(\tilde{f}_{n,w}) = n^{-1} \sum_{i=1}^n \hat{\psi}_{i,w}^2(\tilde{f}_{n,w}) \). This result can be used for the calculation of (pointwise) confidence intervals and conducting hypothesis testing regarding the performance of the estimated ITR \( V_n(\text{sgn}(\tilde{f}_{n,w})) \) under the preference weight \( w \). Theorems 3 and 4 can also be used for simultaneous inference.

Define the vector \( G_n := (G_{n,w}(\tilde{f}_{n,w}), \ldots, G_{n,w}(\tilde{f}_{n,w}))^\prime \) and let \( \| x \|_\infty := \max_{1 \leq i \leq M} |x_i| \) denote the maximum norm of a vector \( x = (x_1, \ldots, x_M) \in \mathbb{R}^M \). Theorems 3 and 4, the Cramér–Wold theorem, and the continuous mapping theorem, imply that, for any fixed matrix \( Q \) (of appropriate dimension),

\[
\| QG_n \|_\infty \xrightarrow{d} \| QG \|_\infty ,
\]

where \( G \sim N(0, \Omega) \), with \( \Omega \) being a positive semidefinite matrix with elements \( \omega_{i,i} = \sigma_{i,i}(\tilde{f}_n) \), \( i = 1, \ldots, M \), and \( \omega_{i,l} = E[\psi_{i,w}(\tilde{f}_n)\psi_{l,w}(\tilde{f}_n)] \), \( i \neq l \). The covariance matrix \( \Omega \) can be consistently (in probability) estimated by the matrix \( \hat{\Omega} \), with elements \( \hat{\omega}_{i,i} = \hat{\sigma}_{i,i}(\tilde{f}_n) \), \( i = 1, \ldots, M \), and \( \hat{\omega}_{i,l} = n^{-1} \sum_{i=1}^n \hat{\psi}_{i,w}(\tilde{f}_{n,w})\psi_{l,w}(\tilde{f}_{n,w}) \), \( i \neq l \). Setting \( Q = \text{diag}(\sigma_{11}(\tilde{f}_n), \ldots, \sigma_{MM}(\tilde{f}_{n,w})) \) implies that

\[
\max_{w \in W} \| \sigma_{w,w}(\tilde{f}_n)G_{n,w}(\tilde{f}_{n,w}) \|_\infty \xrightarrow{d} \| QG \|_\infty .
\]

The last result implies that \( 1 - \alpha \) simultaneous confidence intervals over \( W \), which account for the multiplicity due to considering multiple preference weights, can be calculated as
where $c_\alpha$ is the $1 - \alpha$ percentile of the distribution of $\| Q G \|_\infty$. This percentile can be easily estimated using the following simulation procedure. First, choose a large number $B$ (say $B = 1,000$) and simulate vectors $G_b \sim N(0, \hat{\Omega}_z)$, for $b = 1, \ldots, B$. Then, $c_\alpha$ can be estimated as the empirical $1 - \alpha$ percentile of the sample $\| \hat{Q}_b G_1 \|_\infty, \ldots, \| \hat{Q}_b G_B \|_\infty$, where $\hat{Q}_b = \text{diag}(\hat{\sigma}_{\alpha,1}(\hat{f}_{a,n}), \ldots, \hat{\sigma}_{\alpha,n}(\hat{f}_{a,n}))$. Simultaneous confidence intervals for the differences $\mathcal{V}_\alpha(\text{sgn}(\hat{f}_{a,n})) - \mathcal{V}_\alpha(1)$ and $\mathcal{V}_\alpha(\text{sgn}(\hat{f}_{a,n})) - \mathcal{V}_\alpha(-1)$, $w \in \mathcal{W}$, where $\mathcal{V}_\alpha(1)$ and $\mathcal{V}_\alpha(-1)$ are the value functions for the fixed rules $d(z) = 1$ and $d(z) = -1$, can be calculated similarly by expanding the vector $G_z$ to include $G_{a,n}(1)$ and $G_{a,n}(-1)$, $w \in \mathcal{W}$, and using an appropriate matrix $Q$. This is illustrated in the data application presented in Section 5. We argue that the same inference procedures can also be used for the jackknife value estimator $\mathcal{V}_\alpha(\text{sgn}(\hat{f}_{a,n}))$, $w \in \mathcal{W}$. This statement is justified numerically in the simulation studies.

4 | SIMULATION STUDIES

The finite sample performance of the proposed methods was evaluated in a series of simulation experiments. Specifically, we assessed the performance of (i) the proposed ITR estimator $\hat{d}_{a,n}$ and (ii) the proposed inference methods for the (true) value of the estimated ITR $\mathcal{V}_\alpha(\hat{d}_{a,n})$. We considered a binary treatment variable $A \in \{-1,1\}$, a two-dimensional covariate vector $Z = (Z_i,Z_j)'$, and a multistate process $\{X(t) : t \in [0,\tau]\}$ under a progressive illness–death model with state space $S = \{1,2,3\}$. State 1 represented the initial disease state, state 2 the tumor response state, and state 3 the disease progression or death state, in a hypothetical oncology trial. We considered the preference weights $w_1 = (0,1,0)'$ and $w_2 = (1,1,0)'$, which correspond to the (restricted) mean duration of tumor response and the PFS time, respectively. Treatment was simulated with $P(A = 1) = 0.5$, while the covariates were simulated from the uniform distribution $U(-1,1)$. The multistate process was simulated, conditionally on $A$ and $Z$, based on the transition intensities $\alpha_z(A,Z) = \exp(-0.5Z_i + 0.5Z_i + A f_{a}(Z))$, $\alpha_y(A,Z) = \exp(-0.5Z_i + 0.5Z_j)/4$, and $\alpha_z(A,Z) = \exp(-0.5Z_i + 0.5Z_j - A f_{a}(Z))/2$, where $\alpha_{a}(a,z)$ represents the transition rate from state $h$ to state $j$ for a patient with $(A,Z') = (a,z')$ and $f_{a}, w \in \{w_1,w_2\}$, is the optimal decision function under $w$. Under the aforementioned choices, $\text{sgn}(f_{a}) = \text{sgn}(f_{a}) = \text{sgn}(f_{a})$ (more details on this equality are provided in Web Appendix C). In total, we considered four scenarios according to the form of the optimal decision function as follows:

- **Scenario 1:** $f_{a}(Z) = Z_i + Z_j$
- **Scenario 2:** $f_{a}(Z) = Z_i - Z_j$
- **Scenario 3:** $f_{a}(Z) = 1 + Z_i - \exp(-Z_j)$

*Biometrics. Author manuscript; available in PMC 2024 February 16.*
Scenario 4: \( f_w(Z) = 2 \log(2 - Z_1 - Z_2) - 1.4 \)

The right censoring time was simulated independently of the multistate process from the exponential distribution \( \text{Exp}(\theta) \), with \( \theta \in \{-1.6, -1, -0.4\} \), and the total duration of the study was set to \( \tau = 3 \). These choices for \( \theta \) and \( \tau \) led on average to 28.4%, 42.8%, and 59.5% right-censored observations, respectively.

For each scenario and censoring rate \( \theta \), we considered the training sample sizes \( n \in \{100, 200, 300, 400\} \) and repeated the simulation 1,000 times. In each training dataset, we applied the proposed method with the search space \( \mathcal{F} \) being the space of linear functions and \( \tau = 3 \). Note that, \( f_w \not\in \mathcal{F} \) in Scenarios 3 and 4. The tuning parameter was set to \( \lambda_n = n^{-1/2} \), which satisfies the requirements of Theorems 2 and 4. To evaluate the performance of the estimated ITR \( \hat{d}_{a,w} \), we considered two metrics: (i) the estimated ITR value ratio \( \mathcal{V}_w(\hat{d}_{a,w}) / \mathcal{V}_w(d^*_a) \), where \( d^*_a = \text{sgn}(f_w) \) and \( \mathcal{V}_w(d^*_a) \) is the maximum possible value, and (ii) the misclassification rate, defined as the proportion of patients which were assigned by \( \hat{d}_{a,w} \) to the wrong treatment. An estimated ITR value ratio close to 1 indicates that the performance of \( \hat{d}_{a,w} \) is close to optimal. For each simulated training dataset, the true value of the estimated ITR \( \mathcal{V}_w(\hat{d}_{a,w}) \) and the misclassification rate were calculated based on an independently simulated large testing dataset of size 10,000. To evaluate the validity of the proposed inference methods for \( \mathcal{V}_w(\hat{d}_{a,w}) \) we considered: (i) the average percent errors of the value function estimators \( \hat{\mathcal{V}}_{a,w}(\hat{d}_{a,w}) \) and \( \hat{\mathcal{V}}^{\text{est}}_{a,w}(\hat{d}_{a,w}) \), defined as

\[
\frac{1}{1000} \sum_{b=1}^{1000} \frac{\hat{\mathcal{V}}_{a,w}(\hat{d}_{a,w}) - \mathcal{V}_w(\hat{d}_{a,w})}{\mathcal{V}_w(d^*_a)} \times 100
\]

and

\[
\frac{1}{1000} \sum_{b=1}^{1000} \frac{\hat{\mathcal{V}}^{\text{est}}_{a,w}(\hat{d}_{a,w}) - \mathcal{V}_w(\hat{d}_{a,w})}{\mathcal{V}_w(d^*_a)} \times 100
\]

where \( \hat{\mathcal{V}}_{a,w}, \hat{\mathcal{V}}^{\text{est}}_{a,w}, \) and \( \hat{d}_{a,w} \) are estimates from the \( b \)th simulated training dataset, (ii) the average of the proposed standard error estimates relatively to the Monte Carlo standard deviation of the value estimates, and (iii) the coverage probability of the 95% confidence intervals calculated using the proposed standard error estimator under asymptotic normality.

The simulation results regarding the performance of the estimated ITR \( \hat{d}_{a,w} \) for the time in response (i.e., \( w = (0, 1, 0)' \)) are depicted in Figure 1. The estimated ITR value ratio was above 0.9 in all cases. Thus, even in Scenarios 3 and 4, where \( f_w \) is not linear, the performance of the estimated rule was close to optimal. The maximum fixed rule value ratio \( \max\{\mathcal{V}_w(1), \mathcal{V}_w(-1)\} / \mathcal{V}_w(d^*_a) \), where \( \mathcal{V}_w(1) \) and \( \mathcal{V}_w(-1) \) are the values for the fixed rules \( d(z) = 1 \) and \( d(z) = -1 \), was close to 0.8 in all cases. This indicates that the estimated ITR \( \hat{d}_{a,w} \) can lead to substantially better health outcomes on average compared to fixed, one-size-fits-all, rules. As expected, the estimated ITR value ratio was higher with larger training sample sizes and lower censoring rates. A similar pattern was observed for the misclassification rate of \( \hat{d}_{a,w} \), which was lower for larger training samples and lower censoring rates. The simulation results regarding the validity of the proposed inference
methods for $\mathcal{V}_w(\hat{d}_{n,w})$ are summarized in Tables 1 and 2. In all cases, the value estimator $\hat{\mathcal{V}}_{n,w}(\hat{d}_{n,w})$ provided slightly optimistic estimates of the true value $\mathcal{V}_w(\hat{d}_{n,w})$. The percent error was over 4% only in a few cases with $n = 100$ and a 60% censoring rate. As expected, the percent error of the jackknife value estimator $\hat{\mathcal{V}}_{n,w}^{\text{jk}}(\hat{d}_{n,w})$ was lower than that of $\hat{\mathcal{V}}_{n,w}(\hat{d}_{n,w})$. However, the difference between the two value estimators was smaller for larger training samples and lower censoring rates. The average standard error estimates were close to the corresponding Monte Carlo standard deviation of the estimates and the coverage probabilities close to the nominal level in all cases. These results indicate the consistency of the proposed standard error estimator and support the asymptotic normality result from Theorems 3 and 4.

Additional simulation results evaluating the effect of selecting different values of $\tau$ for the analysis are presented in Figures 1–3 in Web Appendix C.1. In these simulation studies, we considered the values $\tau \in \{1, 2, 3\}$, with 3 being equal to the length of the study. In these studies, larger values of $\tau$ led to a better performance in terms of both the median value function and the variability. This is to be expected as more information is incorporated in the proposed method with a larger value of $\tau$. However, the differences between the choices $\tau = 2$ and $\tau = 3$ were not pronounced in general. Further simulation results on the performance of the proposed ITR estimator when $\mathcal{F}$ is the RKHS with the Gaussian kernel with $\sigma = 1$ (less flexible kernel) and $\sigma = 5$ (more flexible kernel), for the duration of tumor response, are presented in Figures 4–6 in Web Appendix C.1. A larger training sample $n$ led to a better performance in all cases, which reflects the consistency of the proposed ITR estimator. Using a more flexible kernel led to an inferior performance with smaller training sample sizes $n$. In most cases with $n = 200$ or $n = 400$, the performance of a less flexible kernel was the best, while the use of a linear decision function led in general to a better performance compared to two kernel choices when $n = 100$. However, in many cases, the differences between a less flexible kernel and the linear decision function were not pronounced.

Simulation results regarding the performance of the estimated ITR $\hat{d}_{n,w}$ for the PFS time (i.e., $w = (1, 1, 0)'$) are depicted in Figures 7–9 in Web Appendix C.2. For comparison, these plots also illustrate the performance of the ICO and DR methods (Zhao et al., 2015). To apply the DR approach, we estimated $E(T \mid T > t, A, Z)$ based on the semiparametric Cox model for $T$ with $A$, $Z$, and the interactions between $A$ and $Z$ as covariates, according to Zhao et al. (2015). This model is misspecified due to the complexity of the multistate process considered in the simulation studies. The performance of all methods improved with a larger sample size. Furthermore, the proposed method provided estimated ITRs with a substantially lower variability and slightly larger median value (i.e., PFS time) compared to the ICO and DR methods. The improved performance of the proposed method over the ICO and DR methods was more pronounced with a higher censoring rate. The simulation results regarding the validity of the proposed inference methods for $\mathcal{V}_w(\hat{d}_{n,w})$ are summarized in Tables 1 and 2 in Web Appendix C.2. Results on the ICO and DR methods are not reported, there are no such inference procedures for these methods. Similarly to the simulations for the time in response, the performance of our inference methods was satisfactory in all cases.
with the exception of low coverage probabilities for the jackknife estimator when \( n = 100 \). The latter coverage probabilities were at the nominal level when \( n = 400 \). Further simulation results for evaluating the performance of the proposed ITR estimator when \( \mathcal{F} \) is the RKHS with the Gaussian kernel are illustrated in Figures 10–12 in Web Appendix C.2. These results revealed similar patterns to those for the duration of the tumor response.

5 | SPECTRUM Trial Data Analysis

The proposed methodology was applied to data from the SPECTRUM Trial (Vermorken et al., 2013), a phase III randomized trial on recurrent or metastatic squamous-cell carcinoma of the head and neck. The goal of this trial was to evaluate the effectiveness of the addition of panitumumab, a fully human monoclonal antibody which inhibits the epidermal growth factor receptor, to chemotherapy as a first-line treatment approach. Of the 520 patients included in this analysis, 260 patients were randomly assigned to the chemotherapy+panitumumab group \((A = 1)\), while the remaining patients were assigned to the chemotherapy alone group \((A = -1)\). In this trial, tumor response was defined as an at least 30% decrease in the sum of the longest diameter of target lesions according to the response evaluation criteria in solid tumors (Therasse et al., 2000). Throughout the follow-up period, 138 (26.5%) patients achieved tumor response, while 457 (87.9%) experienced a progression of their disease and/or died. Among the latter patients, 120 (26.3%) had achieved tumor response prior to their disease progression or death. The overall median (95% CI) PFS time was 5.59 (5.29, 5.85) months. The estimates of the treatment-specific cumulative transition intensities and state occupation probabilities are depicted in Figures 13 and 14 in Web Appendix D.

In this analysis, the value \( \tau \) was set to 18 months (90th percentile of the follow-up times; there were not many transitions to or from the tumor response state after this timepoint) and the preference weight set was \( \mathcal{W} = \{w_1, w_2, w_3\} = \{(0, 1, 0)', (0.5, 1, 0)', (1, 1, 0)\}'\). As mentioned in Section 2, in this multistate process setup, the utility under the preference weight \( w_1 \) corresponds to the restricted mean duration of tumor response, while the choices \( w_2 \) and \( w_3 \) provide a restricted quality-adjusted lifetime (where the time spent in the initial state is reduced by 50%) and the restricted PFS time, respectively. The covariates considered in this analysis were centered age (in years) at randomization \( Z_{\text{age.c}} \), indicator that the primary tumor site is hypopharynx \( Z_{\text{hyp}} \), indicator of history of the prior treatment for squamous-cell carcinoma of the head and neck \( Z_{\text{trt.hist}} \), and indicator that ECOG performance status at baseline indicates symptoms but the patient is ambulatory (vs. fully active; \( Z_{\text{ECOG}} \)). The tuning parameters \( \lambda_{u,v}, w \in \mathcal{W} \), were selected from the candidate set \( \{0.001, 0.005, 0.01, 0.1, 0.25, 0.5, 1, 2.5, 5, 10, 20, 50, 100\} \times n^{-1/2} \), where \( n = 520 \), using leave-one-out cross-validation. All these potential choices for \( \lambda_{u,v} \) satisfy the requirements of the theorems in Section 3. The estimated optimal decision functions were

\[
\hat{f}_{u,v}(z) = -1.32 - 0.54z_{\text{age.c}} + 0.90z_{\text{hyp}} + 1.17z_{\text{trt.hist}} + 0.83z_{\text{ECOG}}.
\]
\[ \hat{f}_{n,w}(z) \approx -1.40 + 2.00z_{\text{trt.hist}} \quad \text{and} \quad \hat{f}_{n,w}(z) \approx -1.47 + 2.00z_{\text{trt.hist}}, \]

where the absolute values of the estimated coefficients of \( z_{\text{age.c}}, z_{\text{hyp}}, \) and \( z_{\text{ECOG}} \), in \( \hat{f}_{n,w}(z) \) and \( \hat{f}_{n,w}(z) \) were all less than \( 10^{-6} \).

The more complicated form of \( \hat{f}_{n,w} \) may reflect that there is higher heterogeneity in the treatment effect on tumor response compared to the treatment effect on quality-adjusted lifetime and PFS time. Intuitively, one would expect that this higher heterogeneity with respect to tumor response would be carried over to the other outcomes. However, this may not be the case because response has typically a short duration for this tumor and might not have a substantial impact on disease progression and/or death. The class of estimated ITRs is \( \{\hat{d}_{n,w} = \text{sgn}(\hat{f}_{n,w}): w \in \{w_1, w_2, w_3\}\} \). Clearly, the estimated ITRs \( \hat{d}_{n,w} \) and \( \hat{d}_{n,w} \) are equivalent and assign chemotherapy+panitumumab (treatment 1) to the patients with a history of prior treatment, and chemotherapy alone (treatment −1) to those without prior treatment. In contrast, the rule \( \hat{d}_{n,w} \) is more complicated and accounts for more covariates. Next, we estimated the performances (i.e., value functions) of the estimated optimal ITRs and compared them to those of the fixed, one-size-fits-all, rules \( d(z) = 1 \) (everyone is assigned to chemotherapy+panitumumab) and \( d(z) = -1 \) (everyone is assigned to chemotherapy alone). To account for the multiplicity due to conducting inference about nine parameters in total, we calculated 95% simultaneous confidence intervals using the approach described in Section 3. The percentile \( c_{0.05} \) was estimated based on \( B = 1,000 \) simulation replications and the corresponding estimate was \( \hat{c}_{0.05} = 2.59 \). The results from this analysis are summarized in Table 3. The estimated restricted mean (95% CI) potential time in the tumor response state under \( \hat{d}_{n,w} \) was 2.57 (1.76, 3.38) months. This time was significantly longer than the corresponding time under the fixed rule \( d(z) = -1 \) [difference (95% CI): 0.86 (0.03, 1.70) months]. The estimated restricted mean (95% CI) potential quality-adjusted lifetime under \( \hat{d}_{n,w} \) was 5.00 (4.17, 5.83) months. This time was significantly longer than that under the fixed rule \( d(z) = -1 \) [difference (95% CI): 0.95 (0.00, 1.90) months]. Finally, the estimated restricted mean (95% CI) potential PFS time under \( \hat{d}_{n,w} \) was 7.43 (6.38, 8.49) months. There were no significant differences in the restricted mean PFS time between the optimal rule and \( d(z) = -1 \) [difference (95% CI): 1.04 (−0.25, 2.33) months] and between \( d(z) = 1 \) and \( d(z) = -1 \) [difference (95% CI): 0.97 (−0.27, 2.22) months]. Also, no significant differences were observed between the optimal rule and \( d(z) = 1 \). This might be due to a potentially small difference between the effects of chemotherapy alone and chemotherapy+panitumumab among patients for which the two rules assigned different treatment.

6 | DISCUSSION

This paper addressed the issue of optimal ITR estimation in randomized clinical trials with right-censored multistate processes. To achieve this, we devised a novel objective function that can handle general nonhomogeneous multistate processes and can easily incorporate patient preferences. A key feature of the proposed methodology is that it utilizes information from both uncensored and censored observations without positing and estimating a model for the conditional expectation of the outcome given \( A \) and \( Z \), as opposed to the methods by Zhao et al. (2015) for censored failure times. Optimization of this objective function was...
based on the outcome-weighted learning framework (Zhao et al., 2012). The simulation studies provided numerical evidence for the validity of the proposed ITR estimation approach and inference procedures. Also, these studies showed a better performance of the proposed method compared to the methods by Zhao et al. (2015) for censored failure times.

There are two important practical considerations when applying the proposed approach. First, one needs to specify the preference weight set \( \mathcal{W} \). This can be achieved by subject matter experts (e.g., clinicians) or via a survey in a sample of patients. For a given \( \mathcal{W} \), the choice of the most appropriate rule within the class of estimated ITRs \( \hat{d}_{n,w} \) for a given patient with a preference weight \( w_0 \) is \( \hat{d}_{n,w_0} \) with \( w = \arg \min_{w \in \mathcal{W}} \| w - w_0 \| \) (i.e., the closest preference weight in the set \( \mathcal{W} \)). Second, the class of decision functions \( \mathcal{F} \) needs to be chosen. Even though the proposed ITR estimator is universally consistent when \( \mathcal{F} \) is the RKHS with the Gaussian kernel, the rate of convergence under this choice may be extremely slow (Steinwart & Christmann, 2008). This means that an extremely large-sample size may be needed in practice in order to obtain an ITR \( \hat{d}_{n,w_0} \) whose value \( V_{w_0}(\hat{d}_{n,w_0}) \) is reasonably close to the optimal value \( V_{w_0}(d^*_w) \). For this reason, we mainly focused on the class of linear decision functions in most of this paper. Under this choice, the proposed estimator is universally consistent only if the true optimal decision function \( f_{w_0} \) is linear. However, even if \( f_{w_0} \) is not linear, the value of the estimated ITR \( V_{w_0}(\hat{d}_{n,w_0}) \) can be close to the optimal value. This was illustrated in the simulation studies presented in Section 4. In practice, one can further improve the performance of the estimated ITR when \( \mathcal{F} \) is the class of linear functions by considering two-way or three-way covariate interactions (Zhou et al., 2017).

A key assumption of the proposed approach is independent censoring. This assumption is realistic in many clinical trials, where accrual time is not associated with patient characteristics and censoring is mainly due to administrative reasons (Goldberg & Kosorok, 2017; Zhao et al., 2011). A plausible relaxation of this assumption is to allow censoring to depend on treatment \( A \), since censoring rate will likely be higher among those receiving the treatment with the greater toxicity (Templeton et al., 2020). A further relaxation of the independent censoring assumption is to allow censoring to depend on both \( A \) and \( Z \) by imposing a semiparametric Cox model (more details on both relaxations are provided in Web Appendix E). It must be noted that, for the case of censored failure times, the DR method (Zhao et al., 2015) allows the censoring model to be misspecified (unlike the proposed approach) provided that the failure time model is correctly specified. However, the true censoring model may be of a less complicated form than the true failure time model in clinical trial applications (Goldberg & Kosorok, 2017) and, thus, more likely to be correctly specified.

An interesting extension of this work is to allow for interval censoring. This could be achieved by utilizing an estimator of the state occupation probabilities with interval-censored data, and deriving an appropriate objective function using, potentially, calculations similar to those in Section 2. Also, extending the proposed approach for the single-decision problem to the multi-decision setting, such as a sequential multiple assignment randomized
trial (SMART) (Lavori & Dawson, 2000), is both practically and methodologically important.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**ACKNOWLEDGMENTS**

The author thanks the Associate Editor and the anonymous referee for their insightful comments that led to a significant improvement of this paper. This paper is based on research using data obtained from [https://www.projectdatasphere.org/](https://www.projectdatasphere.org/), which is maintained by Project Data Sphere. Neither Project Data Sphere nor the owner(s) of any information from the web site have contributed to, approved, or are in any way responsible for the contents of this article. This research was supported by grants R01AI140854 and R21AI145662 from the National Institutes of Health and the Indiana University Precision Health Initiative.

**Funding information**

National Institute of Allergy and Infectious Diseases, Grant/Award Numbers: R01AI140854, R21AI145662

**DATA AVAILABILITY STATEMENT**

The data that support the findings of this paper are available from Project Data Sphere. Restrictions apply to the availability of these data, which were used under license in this paper. Data are available at [https://www.projectdatasphere.org/](https://www.projectdatasphere.org/) with the permission of Project Data Sphere.

**REFERENCES**

Bai X, Tsiatis AA, Lu W & Song R (2017) Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. Lifetime Data Analysis, 23(4), 585–604. [PubMed: 27480339]

Cortes C & Vapnik V (1995) Support-vector networks. Machine Learning, 20(3), 273–297.

Goldberg Y & Kosorok MR (2017) Support vector regression for right censored data. Electronic Journal of Statistics, 11(1), 532–569.

Hastie T, Tibshirani R, Friedman JH & Friedman JH (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. Springer.

He Y, Kim S, Kim M-O, Saber W & Ahn KW (2021) Optimal treatment regimes for competing risk data using doubly robust outcome weighted learning with bi-level variable selection. Computational Statistics & Data Analysis, 158, 107167. [PubMed: 33994608]

Jiang R, Lu W, Song R & Davidian M (2017) On estimation of optimal treatment regimes for maximizing t-year survival probability. Journal of the Royal Statistical Society: Series B, 79(4), 1165–1185.

Kaufman HL, Andtbacka RH, Collichio FA, Wolf M, Zhao Z, Shilkrut M, et al. (2017) Durable response rate as an endpoint in cancer immunotherapy: insights from oncolytic virus clinical trials. Journal for Immunotherapy of Cancer, 5(1), 1–9. [PubMed: 28105368]

Kosorok MR & Laber EB (2019) Precision medicine. Annual Review of Statistics and its Application, 6, 263–286.

Lavori PW & Dawson R (2000) A design for testing clinical strategies: biased adaptive within-subject randomization. Journal of the Royal Statistical Society: Series A (Statistics in Society), 163(1), 29–38.

Lipkovich I, Dmitrienko A & B D’Agostino R Sr (2017) Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. Statistics in Medicine, 36(1), 136–196. [PubMed: 27488683]
Murphy SA (2003) Optimal dynamic treatment regimes. Journal of the Royal Statistical Society: Series B, 65(2), 331–355.

Murphy SA (2005) An experimental design for the development of adaptive treatment strategies. Statistics in Medicine, 24(10), 1455–1481. [PubMed: 15586395]

Oza AM, Lorusso D, Aghajanian C, Oaknin A, Dean A, Colombo N, et al. (2020) Patient-centered outcomes in ARIEL3, a phase III, randomized, placebo-controlled trial of rucaparib maintenance treatment in patients with recurrent ovarian carcinoma. Journal of Clinical Oncology, 38(30), 3494–3505. [PubMed: 32840418]

Pelzer U, Blanc J-F, Melisi D, Cubillo A, Von Hoff DD, Wang-Gillam A, et al. (2017) Quality-adjusted survival with combination nal-IRI+5-FU/LV vs 5-fu/lv alone in metastatic pancreatic cancer patients previously treated with gemcitabine-based therapy: a Q-TWiST analysis. British Journal of Cancer, 116(10), 1247–1253. [PubMed: 28350787]

Robins JM (2004) Optimal structural nested models for optimal sequential decisions. Proceedings of the second seattle Symposium in Biostatistics, Springer. pp. 189–326.

Rubin DB (2005) Causal inference using potential outcomes: design, modeling, decisions. Journal of the American Statistical Association, 100(469), 322–331.

Steinwart I & Christmann A (2008) Support Vector Machines. Springer Science & Business Media.

Templeton AJ, Amir E & Tannock IF (2020) Informative censoring—a neglected cause of bias in oncology trials. Nature Reviews Clinical Oncology, 17(6), 327–328.

Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. (2000) New guidelines to evaluate the response to treatment in solid tumors. Journal of the National Cancer Institute, 92(3), 205–216. [PubMed: 10655437]

Tsiatis AA, Davidian M, Holloway ST & Laber EB (2019) Dynamic treatment regimes: statistical methods for precision medicine. Chapman and Hall.

Vermorken JB, Stöhlmer-Williams J, Davidenko I, Licitra L, Winquist E, Villanueva C, et al. (2013) Cisplatin and fluorouracil with or without panitumumab in patients with recurrent or metastatic squamous-cell carcinoma of the head and neck (SPECTRUM): an open-label phase 3 randomised trial. The Lancet Oncology, 14(8), 697–710. [PubMed: 23746666]

Zhang B, Tsiatis AA, Laber EB & Davidian M (2012) A robust method for estimating optimal treatment regimes. Biometrics, 68(4), 1010–1018. [PubMed: 22550953]

Zhao Y, Kosorok MR & Zeng D (2009) Reinforcement learning design for cancer clinical trials. Statistics in Medicine, 28(26), 3294–3315. [PubMed: 19750510]

Zhao Y, Zeng D, Rush AJ & Kosorok MR (2012) Estimating individualized treatment rules using outcome weighted learning. Journal of the American Statistical Association, 107(499), 1106–1118. [PubMed: 23630406]

Zhao Y, Zeng D, Socinski MA & Kosorok MR (2011) Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. Biometrics, 67(4), 1422–1433. [PubMed: 21385164]

Zhao Y-Q, Zeng D, Laber EB, Song R, Yuan M & Kosorok MR (2015) Doubly robust learning for estimating individualized treatment with censored data. Biometrika, 102(1), 151–168. [PubMed: 25937641]

Zhou J, Zhang J, Lu W & Li X (2021) On restricted optimal treatment regime estimation for competing risks data. Biostatistics, 22(2), 217–232. [PubMed: 31373360]

Zhou X, Mayer-Hamblett N, Khan U & Kosorok MR (2017) Residual weighted learning for estimating individualized treatment rules. Journal of the American Statistical Association, 112(517), 169–187. [PubMed: 28943682]
FIGURE 1.
Simulation study: performance of \( \hat{d}_{n,w} \) for the duration of tumor response (i.e., \( w = (0, 1, 0)' \), in terms of the average estimated ITR value ratio \( V_w(\hat{d}_{n,w}) / V_w(d_w) \) (top row) and misclassification rate (bottom row). The solid grey horizontal lines (top row) correspond to an optimal performance while the dashed grey horizontal lines (top row) correspond to the maximum fixed rule value ratio \( \max\{V_w(1), V_w(-1)\} / V_w(d_w) \), where \( V_w(1) \) and \( V_w(-1) \) are the values for the fixed rules \( d(z) = 1 \) and \( d(z) = -1 \). ITR, individualized treatment rule.
TABLE 1

Simulation study: performance of the proposed inference methods for the true value of the estimated ITR $\mathcal{V}_w(\hat{d}_{n,w})$ for the duration of tumor response (i.e., $w = (0,1,0)^\prime$), under a linear optimal decision function $f_w$ (Scenarios 1 and 2).

| Scenario | Cens | $n$ | % error | MCSD | ASE | CP | % error | MCSD | ASE | CP |
|----------|------|-----|---------|------|-----|----|---------|------|-----|----|
| 1        | 28%  | 100 | 3.418   | 0.224| 0.215| 0.954| -1.402  | 0.226| 0.215| 0.940|
|          |      | 200 | 2.129   | 0.151| 0.155| 0.961| -0.540  | 0.153| 0.155| 0.953|
|          |      | 400 | 1.062   | 0.109| 0.110| 0.954| -0.511  | 0.112| 0.110| 0.936|
|          | 43%  | 100 | 4.707   | 0.249| 0.240| 0.955| -0.856  | 0.248| 0.240| 0.946|
|          |      | 200 | 2.731   | 0.171| 0.173| 0.951| -0.039  | 0.172| 0.173| 0.949|
|          |      | 400 | 1.330   | 0.117| 0.123| 0.966| -0.630  | 0.120| 0.123| 0.954|
|          | 60%  | 100 | 5.289   | 0.313| 0.303| 0.951| -0.602  | 0.301| 0.303| 0.951|
|          |      | 200 | 3.411   | 0.215| 0.218| 0.964| 0.016   | 0.218| 0.218| 0.947|
|          |      | 400 | 1.983   | 0.141| 0.154| 0.964| -0.681  | 0.143| 0.154| 0.962|
| 2        | 28%  | 100 | 2.393   | 0.265| 0.231| 0.947| -0.663  | 0.261| 0.231| 0.940|
|          |      | 200 | 1.359   | 0.172| 0.166| 0.953| -0.529  | 0.174| 0.166| 0.941|
|          |      | 400 | 0.807   | 0.123| 0.118| 0.964| -0.218  | 0.124| 0.118| 0.953|
|          | 42%  | 100 | 2.403   | 0.288| 0.258| 0.951| -1.570  | 0.286| 0.258| 0.939|
|          |      | 200 | 1.586   | 0.194| 0.185| 0.953| -0.692  | 0.198| 0.185| 0.944|
|          |      | 400 | 1.207   | 0.139| 0.132| 0.946| -0.078  | 0.140| 0.132| 0.945|
|          | 57%  | 100 | 1.164   | 0.353| 0.319| 0.935| -3.364  | 0.343| 0.319| 0.931|
|          |      | 200 | 1.953   | 0.242| 0.230| 0.953| -0.785  | 0.236| 0.230| 0.944|
|          |      | 400 | 1.217   | 0.162| 0.164| 0.964| -0.492  | 0.165| 0.164| 0.955|

Note: Cens, right censoring rate; $n$, training sample size; MCSD, Monte Carlo standard deviation of the estimates; ASE, average of the standard error estimates; CP, empirical coverage probability of the 95% confidence interval.
Simulation study: performance of the proposed inference methods for the true value of the estimated individualized treatment rule $\hat{\mathcal{V}}_{\mathcal{A}}(d_{n,w})$ for the duration of tumor response (i.e., $w = (0,1,0)'$), under a nonlinear optimal decision function $f_w^*$ (Scenarios 3 and 4).

| Scenario | Cens | $n$  | $\%$ error | MCSD | ASE | $\%$ error | MCSD | ASE | CP   |
|----------|------|------|-------------|------|-----|-------------|------|-----|------|
| 3        | 29%  | 100  | 3.706       | 0.240| 0.216| -0.415      | 0.236| 0.216| 0.942|
|          |      | 200  | 2.242       | 0.160| 0.156| -0.049      | 0.156| 0.156| 0.956|
|          |      | 400  | 1.136       | 0.113| 0.111| -0.331      | 0.115| 0.111| 0.943|
|          | 43%  | 100  | 4.483       | 0.258| 0.243| -0.999      | 0.265| 0.243| 0.935|
|          |      | 200  | 2.355       | 0.182| 0.174| -0.563      | 0.184| 0.174| 0.937|
|          |      | 400  | 1.505       | 0.120| 0.124| -0.264      | 0.124| 0.124| 0.949|
|          | 60%  | 100  | 4.719       | 0.317| 0.305| -1.154      | 0.306| 0.305| 0.938|
|          |      | 200  | 3.421       | 0.220| 0.219| 0.254       | 0.214| 0.219| 0.957|
|          |      | 400  | 1.397       | 0.146| 0.155| 0.964       | 0.143| 0.155| 0.961|
| 4        | 29%  | 100  | 3.957       | 0.229| 0.216| -0.411      | 0.234| 0.216| 0.945|
|          |      | 200  | 2.232       | 0.158| 0.155| -0.269      | 0.160| 0.155| 0.944|
|          |      | 400  | 1.128       | 0.113| 0.111| 0.030       | 0.115| 0.111| 0.936|
|          | 43%  | 100  | 4.419       | 0.262| 0.240| 0.260       | 0.268| 0.240| 0.927|
|          |      | 200  | 2.278       | 0.180| 0.173| 0.786       | 0.171| 0.173| 0.948|
|          |      | 400  | 1.197       | 0.128| 0.123| 0.664       | 0.129| 0.123| 0.932|
|          | 59%  | 100  | 5.189       | 0.303| 0.306| -1.175      | 0.313| 0.306| 0.927|
|          |      | 200  | 2.862       | 0.224| 0.218| -1.089      | 0.226| 0.218| 0.929|
|          |      | 400  | 1.526       | 0.161| 0.154| -0.809      | 0.161| 0.154| 0.931|

Note: Cens, right censoring rate; $n$, training sample size; MCSD, Monte Carlo standard deviation of the estimates; ASE, average of the standard error estimates; CP, empirical coverage probability of the 95% confidence interval.
TABLE 3
Analysis of the SPECTRUM trial: estimated mean potential utilities (i.e., value functions) under the estimated optimal individualized treatment rules $\hat{d}_{u,w}$ for the preference weights $w \in \{(0, 1, 0)', (0.5, 1, 0)', (1, 1, 0)\}$, and comparison with those under the fixed, one-size-fits-all, rules $d(z): = 1$ (everyone is assigned to chemotherapy+panitumumab) and $d(z): = −1$ (everyone is assigned to chemotherapy alone).

| $w$        | Parameter                  | Estimate | 95% CI          |
|------------|----------------------------|----------|-----------------|
| (0, 1, 0)' | $\mathcal{V}_w(\hat{d}_{u,w})$ | 2.570    | (1.763, 3.376)  |
|            | $\mathcal{V}_w(d_{u,w}) − \mathcal{V}_w(1)$ | 0.109    | (−0.481, 0.700) |
|            | $\mathcal{V}_w(\hat{d}_{u,w}) − \mathcal{V}_w(−1)$ | 0.862    | (0.025, 1.699)  |
| (0.5, 1, 0)' | $\mathcal{V}_w(\hat{d}_{u,w})$ | 4.999    | (4.168, 5.829)  |
|            | $\mathcal{V}_w(d_{u,w}) − \mathcal{V}_w(1)$ | 0.087    | (−0.539, 0.713) |
|            | $\mathcal{V}_w(\hat{d}_{u,w}) − \mathcal{V}_w(−1)$ | 0.949    | (0.003, 1.896)  |
| (1, 1, 0)' | $\mathcal{V}_w(\hat{d}_{u,w})$ | 7.434    | (6.377, 8.490)  |
|            | $\mathcal{V}_w(d_{u,w}) − \mathcal{V}_w(1)$ | 0.071    | (−0.739, 0.881) |
|            | $\mathcal{V}_w(\hat{d}_{u,w}) − \mathcal{V}_w(−1)$ | 1.043    | (−0.248, 2.334) |

Note: The corresponding 95% simultaneous confidence intervals, that adjust for multiplicity due to conducting inference about nine parameters in total, are also provided. ($\mathcal{V}_w(\hat{d}_{u,w})$, mean potential utility under $\hat{d}_{u,w}$ for the preference weight $w$; $\mathcal{V}_w(1)$, mean potential utility for the preference weight $w$ under the fixed rule $d(z): = 1$; $\mathcal{V}_w(−1)$, mean potential utility for the preference weight $w$ under the fixed rule $d(z): = −1$).