We consider the conditional treatment effect for competing risks data in observational studies. We derive the efficient score for the treatment effect using modern semiparametric theory, as well as two doubly robust scores with respect to (1) the assumed propensity score for treatment and the censoring model, and (2) the outcome models for the competing risks. An important property regarding the estimators is rate double robustness, in addition to the classical model double robustness. Rate double robustness enables the use of machine learning and nonparametric methods in order to estimate the nuisance parameters, while preserving the root-n asymptotic normality of the estimated treatment effect for inferential purposes. We study the performance of the estimators using simulation. The estimators are applied to the data from a cohort of Japanese men in Hawaii followed since 1960s in order to study the effect of mid-life drinking behavior on late life cognitive outcomes. The approaches developed in this article are implemented in the R package “HazardDiff”.

**KEYWORDS**
machine learning, nonparametric methods, orthogonal score, regression models, semiparametric efficiency, treatment effect

1 INTRODUCTION

Our work was motivated by data from the linked epidemiological projects Honolulu Hearth Program (HHP) and Honolulu-Asia Aging Study (HAAS). HHP was established in 1965 as an epidemiological study of rates and risk factors for heart disease and stroke in men of Japanese ancestry living in Oahu and born between 1900 and 1919. HAAS was established in 1991 as a continuation of the HHP with a shift of focus on brain aging, Alzheimer’s disease, vascular dementia, other causes of cognitive and motor impairment, stroke, and the common chronic conditions of late life. During the HAAS period neuropsychological assessments were performed every 2-3 years until 2012. In particular, we are interested in the effect of mid-life exposures captured during HHP on late life cognitive outcomes collected in HAAS. For this cohort of participants then, death presents a competing risk for the cognitive outcomes. This is commonly referred as truncation by death when the cognitive outcomes are analyzed as longitudinal repeated measures,1-4 and competing risk when the cognitive outcomes are time to events, which is the case here.

In observational studies such as the above, it is necessary to control for covariates in order to study the effects of exposures such as alcohol. One way to control for covariates is through including the covariates in a regression model. This leads to the so-called conditional treatment effect. For interpretation purposes, risk functions are often preferred; these are typically survival probabilities for time-to-events data, and cumulative incidence functions for competing risks data. Regression models have been proposed, including on the cause-specific hazard functions and on the subdistribution...
hazard functions. The conditional risk functions can then be obtained via transformations of the conditional hazards. The conditional risk functions can then be obtained via transformations of the conditional hazards.

The additive hazards model has recently been considered in this context. For a binary treatment, this conditional treatment effect is the hazard difference given the covariates under the additive hazards model. Note that misspecification of the functional form of the covariates in the hazard regression model can lead to bias in the estimation of the treatment effect of interest.

To alleviate the reliance on the correct specification of the covariate forms which are “nuisance” themselves, flexible modeling such as nonparametric approaches might be considered. However, they are often inefficient and lead to slower rates of convergence of the estimated treatment effect; this is the “curse of dimensionality” problem discussed in Robins and Ritov. Alternatively, there has been a growing literature on doubly robust estimators that protect against misspecification of the “nuisance” parts of the model. In the survival context Zhang and Schaubel, and derived doubly robust estimators for the treatment effect defined as, or equivalent to, a contrast between expectations of functions of the potential failure times, that is, the failure time that would be observed if the same subject were treated or untreated, respectively, regardless of the actual treatment received. Yang et al developed a doubly robust estimator for the structural accelerated failure time models. Petersen et al and Zheng et al derived targeted maximum likelihood estimators that are doubly robust after discretizing time and recasting the failure event as a binary outcome.

In the absence of competing risks, doubly robust estimators for the hazard difference have been proposed by Dukes et al and Hou et al. Dukes et al considered low dimensional setting, that is, all the nuisance parameters are estimated parametrically or semiparametrically at root- rate, but the treatment might be continuous. Hou et al considered high dimensional setting, using regularization methods with LASSO as a specific case for their theoretical as well as empirical investigation.

In the following we first derive the semiparametrically efficient score for the cause-specific hazard difference under competing risks. We then propose two doubly robust estimators with respect to two sets of models. The first set contains the treatment assignment model, also called the propensity score model, and the model for the censoring distribution. The second set contains the cause-specific hazard models for the competing risks. The proposed estimators are both model doubly robust and rate doubly robust. Model doubly robust refers to the property that the estimators of treatment effects are consistent and asymptotically normal, as long as any one of the two sets of the models are correctly specified, and that the correctly specified models are estimated at root- rate. This is doubly robust in the classical sense.

Rate doubly robust refers to the property that the estimators of treatment effects are consistent and asymptotically normal, when both sets of the models estimate the truth but at possibly slower than root- rate, as long as their product rate is faster than root-. Rate doubly robust property enables the use of modern machine learning or other nonparametric methods, which substantially broadens the range of estimators to be used for the nuisance parameters. With these methods model specification might become much less an issue than previously, so that one might be less concerned about having at least one of the two sets of models correctly specified. We note that rate double robustness was not considered in Dukes et al. Meanwhile, although rate double robustness was established in Hou et al for the LASSO estimators, our results in this paper can be much more broadly applied to potentially many machine learning and other nonparametric methods. In the process we also weaken the censoring assumption as required in Dukes et al and Hou et al.

The rest of the article is organized as follows: after formally defining the parameter of interest, in Section 2 we describe the semiparametrically efficient and Section 3 the two doubly robust scores. In Section 4 we discuss the implementation of the doubly robust estimators, and describe the asymptotic properties of the estimated treatment effects. We study the finite sample performance of the estimators through extensive simulations in Section 5, and apply them to the HHP-HAAS data in order to estimate the effect of alcohol exposure on cognitive impairment in Section 6. We conclude with discussion in the last section.

1.1 Model and notation

Denote \( T \) time to failure, and \( \epsilon = 1, \ldots, J \) the type of failure. Let \( C \) be the censoring random variable, \( X = \min(T, C) \) be the observed (and possibly censored) failure time, and \( \delta = \mathbb{I}\{T \leq C\} \) the event indicator. Let \( A = 0, 1 \) be a binary treatment, and \( Z \) be a vector of baseline covariates. We assume \( \tau < \infty \) to be an upper limit of follow-up time.

A commonly used approach for competing risks data is to model the cause-specific hazard function for each type of failure. The cause-specific hazard functions are the quantities “just identified” by such data, in the sense that any other quantity that can be identified from the competing risks data, can be expressed as a function of the cause-specific
hazard.28 We assume that the conditional cause-specific hazard function, \( h_j(t|A, Z) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} P(t \leq t + \Delta t, \epsilon = j|T \geq t, A, Z) \), for \( j = 1, \ldots, J \), satisfies:

\[
h_j(t|A, Z) = \beta_j A + \lambda_j(t, Z),
\]

where \( \lambda_j(t, Z) \), representing the effect of the covariates on the hazard, is left unspecified. This is a key difference from the more traditional cause-specific additive hazards model that assumes linear effects of both \( A \) and \( Z \); see for example, Shen and Cheng.29 From model (1) then, \( \beta_j = h_j(t|A = 1, Z) - h_j(t|A = 0, Z) \) is the difference between the conditional cause-specific hazard functions of the two treatment groups.

In the following we assume that \( C \perp T|\{A, Z\} \), where “\( \perp \)” indicates statistical independence. This is a standard assumption in the analysis of time-to-event data, and it relaxes the stricter \( C \perp (A, T)|Z \) assumption imposed by both Hou et al8 and Dukes et al.7 We will also use the counting process and the at-risk process notation: \( N_j(t) = \mathbb{1} \{X \leq t, \delta = 1, \epsilon = j\} \) and \( Y(t) = \mathbb{1} \{X \geq t\} \). Under model (1), \( M_j(t) = N_j(t) - \int_0^t Y(u)h_j(u|A, Z)du \) is a local square-integrable martingale with respect to the filtration \( F_t = \sigma \{N_j(s), Y(s+), A, Z : j = 1, \ldots, J, 0 < s < t\} \). In addition, the predictable covariation process \( \langle M_l, M_j \rangle(i) = 0 \) for \( l \neq j \) because with absolutely continuous distributions of the event times, the probability that the competing events happen at the same time equals zero.

2 | SEMIPARAMETRICALLY EFFICIENT SCORE FOR \( \beta \)

The derivation below follows the modern semiparametric theory as described in Tsiatis.19 We provide a sketch for the readers here and leave all the details to the Supplementary Material.

Under the semiparametric model (1), the parameter of interest is \( \beta = [\beta_1, \ldots, \beta_J]^T \) and the nuisance parameter is \( \eta = [\lambda_1(t, z), \ldots, \lambda_J(t, z), \lambda_c(t|a, z), P(a|z), f(z)]^T \), where \( \lambda_c(t|a, z) \) is the conditional hazard function for \( C \) given \( A \) and \( Z \), \( P(a|z) \) is the conditional probability of \( A \) given \( Z \), and \( f(z) \) is the density or probability function of the covariates \( Z \). The likelihood for a single copy of the data takes the following form:

\[
L = \prod_{j=1}^J \left( \beta_j A + \lambda_j(X, Z) \right) \mathbb{I}_{\{\delta = 1, c = j\}} \exp \{ -\beta_j A t - \Lambda_j(X, Z) \} \\
\times \left( \lambda_c(X|A, Z) \right)^{1-c} \exp \{ -\Lambda_c(X|A, Z) \} P(A|Z)f(Z),
\]

where \( \Lambda_j(t, z) = \int_0^t \lambda_j(u, z)du \) for \( j = 1, \ldots, J \) and \( \Lambda_c(t|a, z) = \int_0^t \lambda_c(u|a, z)du \). From the likelihood, one can derive the score for the parameter of interest. In the Supplementary Material we show that under model (1),

\[
S_{\beta} = \frac{\partial \log L}{\partial \beta} = \left\{ \int_0^T \frac{A \cdot dM_j(t)}{h_j(t|A, Z)} \right\}_{j=1}^J.
\]

In addition, if \( \eta \) were finite dimensional, the score for the nuisance parameter would be \( S_\eta = \partial \log L / \partial \eta \). The nuisance tangent space, denoted by \( \Lambda \), is the space spanned by the nuisance score. When \( \eta \) has infinite dimension, as in our case, the notion of nuisance tangent space can be extended through the definition of parametric submodels. We leave the technicality of this definition to chapter 4 of Tsiatis.19

An estimator \( \hat{\beta} \) is asymptotically linear if there exists a function of the data \( \varphi \), such that \( \sqrt{n}(\hat{\beta} - \beta_0) = \sum_{i=1}^n \varphi_i / \sqrt{n} + o_p(1) \). The function \( \varphi \), named influence function, has mean zero and finite variance, thus guarantees the asymptotic normality of the estimator \( \hat{\beta} \). Such estimators are therefore desirable and they are uniquely defined by their influence functions. Every influence function belongs to the orthogonal complement of the nuisance tangent space; see theorem 4.2 of Tsiatis.15 This space, denoted by \( \Lambda^\perp \), is therefore the starting point to define semiparametric estimators for \( \beta \) that are consistent and asymptotically normal.

The space \( \Lambda^\perp \) is also important because it allows one to find orthogonal scores. A score \( \psi(\beta, \eta) \) is orthogonal if

\[
\left. \frac{\partial}{\partial r} E \{ \psi(\beta_0; \eta_0 + r(\eta - \eta_0)) \} \right|_{r=0} = 0,
\]

where we use the subscript “0” to indicate the true value of the parameters. Orthogonal scores are invariant to small perturbations of the nuisance parameter around the truth and so the estimation of the nuisance parameter may not greatly
affect the estimation of the treatment effect. It is shown in the Supplementary Material (Lemma 3) that an estimating function belongs to $\Lambda^+$ if and only if it is an orthogonal score.

The following lemma gives the form of the orthogonal complement of the nuisance tangent space under model (1).

**Lemma 1.** Under model (1), the orthogonal complement of the nuisance tangent space takes the following form:

$$
\Lambda^+ = \left\{ \sum_{j=1}^J \int_0^T \left[ g_j(t, A, Z) - \frac{E \left\{ g_j(t, A, Z) \eta_j^{-1}(t|A, Z) S_c(t|A, Z) e^{-\sum_{l=1}^J \beta_l Z_l|Z} \right\}}{E \left\{ \eta_j^{-1}(t|A, Z) S_c(t|A, Z) e^{-\sum_{l=1}^J \beta_l Z_l|Z} \right\}} \right] dM_j(t) \right\}.
$$

for all $g_j(t, A, Z)$, $j = 1, \ldots, J$.

The proofs of all results can be found in the Supplementary Material.

Among all the semiparametric asymptotically linear estimators of $\beta$, the efficient score is $S_{\beta} = \Pi [S_{\beta}|\Lambda] = \Pi [S_{\beta}|\Lambda^+]$, where $\Pi [S_{\beta}|\Lambda]$ and $\Pi [S_{\beta}|\Lambda^+]$ are the projections of $S_{\beta}$ onto $\Lambda$ and $\Lambda^+$, respectively. From (2) it can be seen that this would be the element of $\Lambda^+$ that corresponds to $g_j(t, A, Z) = Ae_j$, where $e_j$ is a vector with 1 at the $j$th position and 0 elsewhere, $j = 1, \ldots, J$. Therefore we have the following result.

**Theorem 1.** Under model (1) the efficient score has the following form:

$$
S_{\text{eff}}(\beta) = \left\{ \int_0^T \left[ A - \frac{E \left\{ A \eta_j^{-1}(t|A, Z) S_c(t|A, Z) e^{-\sum_{l=1}^J \beta_l Z_l|Z} \right\}}{E \left\{ \eta_j^{-1}(t|A, Z) S_c(t|A, Z) e^{-\sum_{l=1}^J \beta_l Z_l|Z} \right\}} \right] dM_j(t) \right\}.
$$

**Remark 1.** If we make the stronger assumption of $C \perp (A, T)|Z$ as in Dukes et al and Hou et al., $S_c(t|A, Z) = S_c(t|Z)$, and so the efficient score simplifies to:

$$
S_{\text{eff}}(\beta) = \left\{ \int_0^T \left[ A - \frac{E \left\{ A \eta_j^{-1}(t|A, Z) e^{-\sum_{l=1}^J \beta_l Z_l|Z} \right\}}{E \left\{ \eta_j^{-1}(t|A, Z) e^{-\sum_{l=1}^J \beta_l Z_l|Z} \right\}} \right] dM_j(t) \right\}.
$$

In this case, $S_c$ is therefore no longer needed for the estimation of $\beta$. If further $J = 1$, the efficient score in Dukes et al is recovered.

**Remark 2.** Traditionally, estimation of parameters from competing risks data allows estimating the parameters for one type of failure at a time. This is the case for the widely used cause-specific Cox model and the traditional additive cause-specific hazards model. However, for both the efficient score and the doubly robust scores below, the components of $\beta$ are estimated jointly from a multi-dimensional score.

The above score is locally efficient in the sense that its asymptotic variance attains the semiparametric efficiency bound when $h_j(t, Z)$, $S_c(t|a, z)$ and $P(a|z)$ are known or correctly estimated; see theorem 4.1 of Tsiatis. Unfortunately, since $h_j(t|A, Z)$ in (5) is unknown and estimators for it are not readily available, the efficient score may not be directly used in practice. We will however exploit both (4) and (5) to derive two doubly robust scores for the estimation of $\beta$.

### 3 | DOUBLY ROBUST SCORES

#### Doubly robust Score 1

Denote $\pi(Z) = P(A = 1|Z)$. Inspired by Hou et al., we choose in (4) $g_j(t, A, Z) = e^\sum_{l=1}^J \beta_l Z_l S_c^{-1}(t|A, Z) [A - \pi(Z)] h_j(t, A, Z)$, and obtain

$$
S_1(\beta; S_c, \pi, A) = \left\{ \int_0^T e^\sum_{l=1}^J \beta_l Z_l S_c^{-1}(t|A, Z) [A - \pi(Z)] dM_j(t; \beta_j, \lambda_j) \right\}.
$$

In this case, $S_c$ is therefore no longer needed for the estimation of $\beta$. If further $J = 1$, the efficient score in Dukes et al is recovered.
where \( \Lambda = (\Lambda_1, \ldots, \Lambda_J)^T \), and \( dM_j(t; \beta, \Lambda_j) = dN_j(t) - Y(t)\beta_j A dt - Y(t)d\Lambda_j(t, Z) \). We note that \( M_j(t; \beta_0, \Lambda_0) \) is a martingale under model (1), where again the subscript “0" indicates the true value. The main difference between (6) when \( J = 1 \) and the score from Hou et al\(^{8} \) is the incorporation of the censoring distribution \( S_c \), so that we do not need the stronger assumption \( C \perp (T, A) | Z \). We note also that Hou et al\(^{8} \) directly constructed their score as a member of \( \Lambda^\perp \) using definition (3).

### Doubly robust Score 2

The second approach removes the unknown hazard weights from the efficient score (5), as done in Lin and Ying\(^{33} \) for the additive hazards regression model, and we have

\[
S_2(\beta; S_c, \pi, \Lambda) = \left\{ \int_0^T \left[ A - E(t; \beta, S_c, \pi) \right] dM_j(t; \beta, \Lambda_j) \right\}_{j=1}^J,
\]

where:

\[
E(t; \beta, S_c, \pi) = \frac{E\left[ Ae^{-\sum_{j=1}^J \beta_j A t} S_c(t|A, Z)|Z \right]}{E\left[ e^{-\sum_{j=1}^J \beta_j A t} S_c(t|A, Z)|Z \right]} = \frac{e^{-\sum_{j=1}^J \beta_j t} S_c(t|A = 1, Z)\pi(Z)}{e^{-\sum_{j=1}^J \beta_j t} S_c(t|A = 1, Z)\pi(Z) + S_c(t|A = 0, Z) \{ 1 - \pi(Z) \}}.
\]

We note that Score 2 in (7) is completely new to our best knowledge, even in the absence of competing risks.

Since both scores (6) and (7) belong to \( \Lambda^\perp \), they are orthogonal scores. In addition, they are doubly robust with respect to the estimation of both \( S_c \) and \( \pi \), and that of \( \Lambda \), as stated in the theorem below.

**Theorem 2.** \( E\{ S_1(\beta_0; S_c, \pi, \Lambda) \} = E\{ S_2(\beta_0; S_c, \pi, \Lambda) \} = 0 \) if either \( S_c = S_{c0} \) and \( \pi = \pi_0 \), or \( \Lambda = \Lambda_0 \), where subscript “0" indicates the true quantities.

Finally, if we are willing to make the stronger assumption \( C \perp (T, A) | Z \), the above two scores simplify to:

\[
\tilde{S}_1(\beta; \pi, \Lambda) = \left\{ \int_0^T e^{\sum_{j=1}^J \beta_j A t} \{ A - \pi(Z) \} dM_j(t; \beta, \Lambda) \right\}_{j=1}^J,
\]

\[
\tilde{S}_2(\beta; \pi, \Lambda) = \left\{ \int_0^T \left[ A - \frac{e^{\sum_{j=1}^J \beta_j t} \pi(Z)}{e^{\sum_{j=1}^J \beta_j t} \pi(Z) + \{ 1 - \pi(Z) \}} \right] dM_j(t; \beta, \Lambda) \right\}_{j=1}^J,
\]

where \( S_c \) is no longer involved. We will consider the implementation of these two simplified scores in the simulation below as well.

### 4 | ESTIMATION AND INFEERENCE

Given a random sample of size \( n \) we write

\[
S_{1,n}(\beta; S_c, \pi, \Lambda) := \frac{1}{n} \sum_{i=1}^n S_{1i}(\beta; S_c, \pi, \Lambda),
\]

and

\[
S_{2,n}(\beta; S_c, \pi, \Lambda) := \frac{1}{n} \sum_{i=1}^n S_{2i}(\beta; S_c, \pi, \Lambda).
\]

Both (10) and (11) depend on the quantities \( S_c, \pi \) and \( \Lambda \) that need to be estimated.
For estimation of the propensity score \( \pi(\cdot) \) and the censoring model \( S_c(\cdot, \cdot) \), we leave it to the users to choose any working model as long as some mild assumptions, given later, are satisfied. From here on we use \( \hat{S}_c, \hat{\pi}, \) and \( \hat{\Lambda} \) to denote estimators of the nuisance parameters \( S_c, \pi, \) and \( \Lambda \); note that the estimator for \( \Lambda \) may also depend on \( \beta \) as described below.

For the estimation of \( \Lambda \) we consider here the usual linear working models:

\[
\Lambda_j(t; Z; G_j, \gamma_j) = G_j(t) + \gamma_j^\top Z t.
\]

(12)

The parameters \( \gamma = (\gamma_1, \ldots, \gamma_J)^\top \) and \( G = (G_1, \ldots, G_J)^\top \) can be estimated using the approach of Shen and Cheng, which is equivalent to applying the estimating equations of Lin and Ying separately to each failure type, and the expressions of \( \hat{\gamma} \) and \( \hat{G} \) are given in the Appendix.

For the estimation of \( G_j \), following Hou et al we consider also the weighted Breslow estimator (see the Appendix). The advantage of using this estimator is that it leads to the closed-form solution \( \hat{\beta}_j^{(1)} \) to \( S_{1,n} \) for \( j = 1, \ldots, J \), the expression of which is also provided in the Appendix.

For \( S_{2,n} \), using \( \hat{G} \) above and after some algebra, we have:

\[
S_{2,n}(\beta; \hat{S}_c, \hat{\pi}, \hat{\Lambda}) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_0^T \left\{ A_i - \mathcal{E}_i(t; \beta, \hat{S}_c, \hat{\pi}) - \hat{A}(t) + \bar{\mathcal{E}}(t) \right\} dt \right\}_{j=1}^J.
\]

where

\[
\Lambda(t) = \frac{\sum_{i=1}^{n} Y_i(t) A_i}{\sum_{i=1}^{n} Y_i(t)}, \quad \bar{\mathcal{E}}(t) = \frac{\sum_{i=1}^{n} Y_i(t) \mathcal{E}_i(t; \beta, \hat{S}_c, \hat{\pi})}{\sum_{i=1}^{n} Y_i(t)}.
\]

Once the estimators \( \hat{\pi}, \hat{S}_c \) and \( \hat{\gamma} \) are available, we define \( \hat{\beta}^{(2)} \) to be the root of \( S_{2,n}(\beta; \hat{S}_c, \hat{\pi}, \hat{\Lambda}) \).

We study the asymptotic properties of these estimators below. For simplification of argument we assume \( J = 2 \); extension to \( J > 2 \) should be straightforward. We need the following main assumption concerning the convergence of the nuisance parameter estimators.

**Assumption 1.** There exist \( \mathcal{S}_c^*(\cdot, \cdot), \pi^*(\cdot), \gamma_j^*, G_j^*(\cdot) \) such that:

\[
\sup_{t \in [0, r], z \in Z, a = 0, 1} \left| \hat{S}_c(t|a, z) - \mathcal{S}_c^*(t|a, z) \right| = O_p(a_n), \quad \sup_{z \in Z} \left| \hat{\pi}(z) - \pi^*(z) \right| = O_p(b_n),
\]

\[
\left\| \hat{\gamma}_j - \gamma_j^* \right\|_1 = O_p(c_n), \quad \sup_{t \in [0, r]} \left| \hat{G}_j(t) - G_j^*(t) \right| = O_p(c_n).
\]

for some \( a_n = o(1), b_n = o(1), c_n = o(1) \) and for \( j = 1, 2 \), where \( Z \) is the sample space of \( Z \).

The proof of the asymptotic properties below makes use of the additive structure of model (12). On the other hand, if some other estimator \( \hat{\Lambda}_j \) instead of that under (12) is used, cross-fitting is generally needed in order to ensure the asymptotic normality, under a similar condition of \( \sup_{t \in [0, r], z \in Z} \left| \hat{\Lambda}_j(t, z) - \Lambda_j^*(t, z) \right| = O_p(c_n), j = 1, 2 \).

### 4.1 Asymptotic properties using Score 1

Under Assumption 1 and additional General Assumptions in the Supplementary Material, which consist of standard regularity assumptions, positivity assumptions often used in causal inference, and additional technical assumptions, under case (a), (b) or (c) listed below, \( \hat{\beta}^{(1)} - \beta_0 = o_p(1) \) and

\[
\sqrt{n} \left( \hat{\beta}^{(1)} - \beta_0 \right) \xrightarrow{D} \mathcal{N}(0, \Sigma),
\]
Asymptotic properties using Score 2

where:

(a) $\Sigma = \Sigma^{(a)}$ given in the Supplementary Material, as long as $S^*_c = S_{c0}$ and $\pi^* = \pi_0$, $a_n = b_n = n^{-1/2}$, and Assumptions A1-A2 in the Supplementary Material hold.

(b) $\Sigma = \Sigma^{(b)}$ given in the Supplementary Material, as long as $\gamma^* = \gamma^0$ and $G^* = G_0$, $c_n = n^{-1/2}$, and Assumptions B1-B2 in the Supplementary Material hold;

(c) $\Sigma = \Sigma^{(c)}$ given in the Supplementary Material, as long as $S^*_c = S_{c0}$, $\pi^* = \pi_0$, $\gamma^* = \gamma^0$ and $G^* = G_0$, $a_n c_n = o(n^{-1/2})$ and $b_n c_n = o(n^{-1/2})$, and Assumptions C1, C2 in the Supplementary Material hold. In this case a consistent estimator of $\Sigma^{(c)}$ is also given in the Supplementary Material.

In the above (a) and (b) are known as model double robustness, and (c) is known as rate double robustness. To provide some intuition for rate double robustness, it is shown in the Supplementary Material that the estimated treatment effect is asymptotically linear. In the Supplementary Material we also show that the variance estimator derived in case (c) above is somehow robust to model misspecification. Alternatively we may use nonparametric bootstrap, the validity of which is guaranteed by the fact that the estimated treatment effect is asymptotically linear. In the Supplementary Material we also show that the explicit form of the asymptotic variance can be derived for specific working models in the cases (a) and (b). In particular, we illustrate with proportional hazards modeling for $S_c$ and logistic regression for $\pi$ in case (a), and additive hazards modeling for $\Lambda_j$ in case (b). However, due to their complex forms and also because in practice one does not know which model is correct, we do not derive an estimator for the asymptotic variance.

4.2 Asymptotic properties using Score 2

We obtain similar asymptotic results using Score 2. Under Assumption 1 and additional General Assumptions in the Supplementary Material, under case (a), (b) or (c) listed below, $\hat{\beta}^{(2)} - \beta_0 = o_p(1)$. In addition,

$$\sqrt{n} \left( \hat{\beta}^{(2)} - \beta_0 \right) \overset{D}{\longrightarrow} \mathcal{N}(0, \Gamma),$$

where:

(a) $\Gamma = \Gamma^{(a)}$ given in the Supplementary Material, as long as $S^*_c = S_{c0}$, $\pi^* = \pi_0$, $a_n = b_n = n^{-1/2}$, and Assumption A’1 in the Supplementary Material hold (model double robustness 1).

(b) $\Gamma = \Gamma^{(b)}$ given in the Supplementary Material, as long as $\gamma^* = \gamma^0$ and $G^* = G_0$, $c_n = n^{-1/2}$, and Assumption B’1 in the Supplementary Material hold (model double robustness 2).

(c) $\Gamma = \Gamma^{(c)}$ given in the Supplementary Material, as long as $S^*_c = S_{c0}$, $\pi^* = \pi_0$, $\gamma^* = \gamma^0$ and $G^* = G_0$, $a_n c_n = o(n^{-1/2})$ and $b_n c_n = o(n^{-1/2})$, and Assumptions C’1, 2 in the Supplementary Material hold (rate double robustness). In this case a consistent estimator of $\Gamma^{(c)}$ is also given in the Supplementary Material.

5 SIMULATION EXPERIMENTS

In this section we investigate the performance of the proposed estimators on a series of simulated data sets. For each scenario, we simulate 500 data sets of 1000 observations. True $\beta_1 = \beta_2 = 0.1$. The percentage of treated subjects is 40% – 50% and the percentage of censored subjects is 10% – 30%, with additional results under heavy censoring of
around 50% in the Supplement Materials, which also contain the results for \( J = 3 \). For both estimators \( \hat{\beta}^{(1)} \) and \( \hat{\beta}^{(2)} \), model-based standard errors are used to construct 95% confidence intervals. As illustration, in one of the scenarios we also report the nonparametric bootstrap standard error based on the first 100 simulations due to the intensive computation demand, where we draw 100 resamples with replacement from \((X_i, \epsilon_i, \delta_i, A_i, Z_i), i = 1, \ldots, n\).

We consider separately independent and dependent censoring. For independent censoring, estimation of the censoring distribution is not required and we use the simplified scores (8) and (9).

### 5.1 Independent censoring

Here the censoring variable \( C \) is simulated independently of \( T, A, Z \). We consider five different simulation scenarios described in Table 1. For estimation of the propensity score, the working models \( A_{\text{logit}}, A_{\text{logit}}^*, \text{and } A_{\text{nu}} \) are given in the footnote of the table, which are, respectively, logistic regression, logistic regression with interaction, and the R package “twang” implementing gradient boosted models for estimation of the propensity score.\(^{35} \) For the competing risks, we fit as working models \( B \) the semiparametric additive hazards model.

Both \( \hat{\beta}^{(1)} \) and \( \hat{\beta}^{(2)} \) are consistent and asymptotically normal as long as one of the \( A \) or \( B \) working models is correct in low dimensions, that is, when one of the working models is parametric or semiparametric and correctly specified. For comparison we also report the estimate of \( \beta \) under “Regression” from fitting \( B \), which is valid when \( B \) is correctly specified.

The results of the simulations are reported in Table 2. It can be seen that when model \( B \) is misspecified as in Scenarios 3, 4, and 5, the direct regression estimator of \( \beta \) is severely biased with poor coverage of the confidence intervals (CI). The estimators from both Scores 1 and 2 have little bias in Scenarios 1, 2, and 3, with good coverage of CI’s using model-based standard errors (SE), even when one of the models is wrong. In Scenario 4 where the competing risks are generated by

| Scenario | Data-generating mechanism | Fitted models |
|----------|--------------------------|---------------|
| 1        | \( Z_1, Z_2 \sim U(0,0.5) \) | \( A_{\text{logit}}: \text{CORRECT and } A_{\text{nu}} \) |
|          | logit \( \pi(Z) = Z_1 - Z_2 \) | \( B: \text{CORRECT} \) |
|          | \( \lambda_j(t) = 0.1A + 1 + Z_1 + Z_2 \) | |
|          | \( C \sim U(0, 3) \) | |
| 2        | \( Z_1, Z_2 \sim U(0,0.5) \) | \( A_{\text{logit}}: \text{WRONG and } A_{\text{nu}} \) |
|          | logit \( \pi(Z) = 0.25(Z_1 - Z_2) - 0.5Z_1Z_2 \) | \( B: \text{CORRECT} \) |
|          | \( \lambda_j(t) = 0.1A + 0.3 + Z_1 + Z_2 \) | |
|          | \( C \sim U(0, 3) \) | |
| 3        | \( Z_1 \sim \mathcal{N}(0, 1), Z_2 \sim \mathcal{N}(Z_1, 1) \) | \( A_{\text{logit}}^*: \text{CORRECT and } A_{\text{nu}} \) |
|          | logit \( \pi(Z) = 0.25(Z_1 - Z_2) + 0.5Z_1Z_2 - 1 \) | \( B: \text{WRONG} \) |
|          | \( \lambda_j(t) = 0.1A + 0.3 + |Z_1| + \log(1 + |Z_2|) \) | |
|          | \( C \sim U(0, 3) \) | |
| 4        | \( Z_1 \sim \mathcal{N}(0, 1), Z_2 \sim \mathcal{N}(Z_1, 1) \) | \( A_{\text{logit}}^*: \text{CORRECT and } A_{\text{nu}} \) |
|          | logit \( \pi(Z) = 0.25(Z_1 - Z_2) + 0.5Z_1Z_2 - 1 \) | \( B: \text{WRONG} \) |
|          | \( \lambda_j(t) = 0.1A + \exp(Z_1 + Z_2) \) | |
|          | \( C \sim U(0, 3) \) | |
| 5        | \( Z_1 \sim \mathcal{N}(0, 1), Z_2 \sim \mathcal{N}(Z_1, 1) \) | \( A_{\text{logit}}: \text{WRONG and } A_{\text{nu}} \) |
|          | logit \( \pi(Z) = 0.25(Z_1 - Z_2) + 0.5Z_1Z_2 - 1 \) | \( B: \text{WRONG} \) |
|          | \( \lambda_j(t) = 0.1A + 0.3 + |Z_1| + \log(1 + |Z_2|) \) | |
|          | \( C \sim U(0, 3) \) | |

Note: \( A_{\text{logit}}: \pi(z; a) = \expit(a^Tz); A_{\text{nu}}: \text{twang}; A_{\text{logit}}^*: \pi(z; a) = \expit(a^Tz + a^*z_1z_2); B: \lambda_j(t, z; G_j, \gamma_j) = G_j(t) + \gamma_j^Tz, j = 1, 2. \)
### TABLE 2 Results of simulations from Scenarios 1-5, independent censoring; true $\beta_1 = \beta_2 = 0.1$

| Scenario | $\beta_1$   | Score 1 | Score 2 | Regression |
|----------|-------------|---------|---------|------------|
|          | PS          | Bias    | SD      | SE         | CP         | Bias    | SD      | SE         | CP         | Bias    | SD      | SE         | CP         |
| 1        | logistic    | -0.012  | 0.156   | 0.146     | 0.93       | -0.006  | 0.157   | 0.146     | 0.93       | -0.006  | 0.157   | 0.147     | 0.93       |
|          | twang       | -0.012  | 0.159   | 0.150     | 0.93       | -0.006  | 0.161   | 0.149     | 0.93       |          |         |           |            |
|          | logistic    | 0.0006  | 0.144   | 0.147     | 0.95       | 0.006   | 0.146   | 0.146     | 0.95       | 0.006   | 0.146   | 0.147     | 0.95       |
|          | twang       | -0.003  | 0.147   | 0.150     | 0.96       | 0.005   | 0.148   | 0.149     | 0.95       | 0.005   | 0.148   | 0.149     | 0.95       |
| 2        | logistic    | -0.010  | 0.108   | 0.120     | 0.97       | -0.007  | 0.108   | 0.120     | 0.97       | -0.007  | 0.108   | 0.121     | 0.97       |
|          | twang       | -0.011  | 0.110   | 0.124     | 0.97       | -0.008  | 0.110   | 0.124     | 0.97       |          |         |           |            |
|          | logistic    | 0.001   | 0.127   | 0.121     | 0.95       | 0.005   | 0.129   | 0.121     | 0.95       | 0.005   | 0.128   | 0.121     | 0.95       |
|          | twang       | 0.001   | 0.131   | 0.124     | 0.95       | 0.004   | 0.133   | 0.124     | 0.94       |          |         |           |            |
| 3        | logistic    | -0.009  | 0.160   | 0.163     | 0.96       | 0.001   | 0.162   | 0.163     | 0.95       | 0.336   | 0.170   | 0.163     | 0.48       |
|          | twang       | -0.006  | 0.158   | 0.162     | 0.97       | 0.006   | 0.161   | 0.162     | 0.96       |          |         |           |            |
|          | logistic    | 0.000   | 0.157   | 0.163     | 0.97       | 0.011   | 0.158   | 0.163     | 0.96       | 0.350   | 0.163   | 0.163     | 0.42       |
|          | twang       | 0.006   | 0.153   | 0.162     | 0.97       | 0.018   | 0.155   | 0.163     | 0.96       |          |         |           |            |
| 4        | logistic    | 0.004   | 0.091   | 0.077     | 0.90       | 0.006   | 0.091   | 0.080     | 0.91       | 0.570   | 0.127   | 0.095     | 0           |
|          | twang       | 0.044   | 0.089   | 0.075     | 0.84       | 0.047   | 0.092   | 0.079     | 0.86       | 0.099   | 0.089   | 0.099     | 0.98       |
|          | logistic    | 0.002   | 0.094   | 0.077     | 0.89       | 0.003   | 0.095   | 0.080     | 0.90       | 0.566   | 0.128   | 0.095     | 0           |
|          | twang       | 0.041   | 0.092   | 0.075     | 0.89       | 0.044   | 0.095   | 0.079     | 0.86       | 0.099   | 0.096   | 0.099     | 0.96       |
| 5        | logistic    | 0.302   | 0.166   | 0.162     | 0.55       | 0.337   | 0.164   | 0.163     | 0.47       | 0.340   | 0.177   | 0.164     | 0.47       |
|          | twang       | -0.006  | 0.166   | 0.163     | 0.95       | 0.007   | 0.170   | 0.162     | 0.94       |          |         |           |            |
|          | logistic    | 0.307   | 0.156   | 0.162     | 0.55       | 0.340   | 0.177   | 0.163     | 0.45       | 0.343   | 0.164   | 0.164     | 0.46       |
|          | twang       | 0.003   | 0.153   | 0.163     | 0.97       | 0.014   | 0.156   | 0.163     | 0.96       |          |         |           |            |

**Note:** Column PS indicates the working model for the propensity score. For Scenario 4, the first row of SE4 and CP are model-based, and the second row from bootstrap.

**Abbreviations:** CP, coverage of the 95% confidence interval; SD, standard deviation; SE, standard error.

the Cox-Aalen model, the model-based SE’s underestimate the SD’s, and the CI’s undercover. Bootstrap increases the SE and hence the coverage of the CI’s. In addition in Scenario 4 when “twang” is used, there is no guarantee according to our theory and the estimation bias is much more substantial, although this seems less an issue in Scenario 3 when “twang” is used. In Scenario 5, when logistic regression is used for the propensity score, both the proposed estimators and the direct regression estimator are biased, which is not unexpected since both models are wrong. On the other hand, when ’twang’ is used for the propensity score, the proposed estimators turn out to have little bias, with good coverage of the CI’s even when the model-based standard errors are used.

### 5.2 Dependent censoring

For dependent censoring we consider the five scenarios described in Table 3. For estimating the censoring distribution, we consider the proportional hazards working model for all five scenarios; in Scenario 8 we also estimate the censoring distribution using the random survival forest. For the latter we use the R package “randomForestSRC” and its default hyperparameters. We report the results using both the simplified scores, $(8)$ and $(9)$, which assume independent censoring.
TABLE 3 Data-generating mechanisms of Scenarios 6-10; $Z = [Z_1, Z_2]^\top$, $C \perp T|A, Z$

| Scenario | Data-generating mechanism | Fitted models |
|----------|--------------------------|---------------|
| 6        | $Z_1, Z_2 \sim U(0,0.5)$ | $A_{\logit}$: CORRECT and $A_{\twang}$ |
|          | $\logit \, \pi(Z) = Z_1 - Z_2$ | $\mathcal{C}$: CORRECT |
|          | $C \sim \text{Exp}(\exp(-1 + A + Z_1 + Z_2))$ | $B$: CORRECT |
|          | $\lambda_j(t) = 0.1A + 1 + Z_1 + Z_2$ | |
| 7        | $Z_1, Z_2 \sim U(0,0.5)$ | $A_{\logit}$: WRONG and $A_{\twang}$ |
|          | $\logit \, \pi(Z) = 0.25(Z_1 - Z_2) - 0.5Z_1 Z_2$ | $\mathcal{C}$: CORRECT |
|          | $C \sim \text{Exp}(\exp(-1 + A + Z_1 + Z_2))$ | |
|          | $\lambda_j(t) = 0.1A + 0.3 + Z_1 + Z_2$ | $B$: CORRECT |
| 8        | $Z_1, Z_2 \sim U(0,0.5)$ | $A_{\logit}$: WRONG and $A_{\twang}$ |
|          | $\logit \, \pi(Z) = 0.25(Z_1 - Z_2) - 0.5Z_1 Z_2$ | $\mathcal{C}$: WRONG and RSF |
|          | $\lambda_j(t|A, Z) = 2t + A - Z_1 - Z_2$ | $B$: CORRECT |
|          | $\lambda_j(t) = 0.1 + A + 0.3 + Z_1 + Z_2$ | |
| 9        | $Z_1 \sim \mathcal{N}(0,1), Z_2 \sim \mathcal{N}(Z_1, 1)$ | $A_{\logit}^*: \text{CORRECT and } A_{\twang}$ |
|          | $\logit \, \pi(Z) = 0.25(Z_1 - Z_2) + 0.5Z_1 Z_2 - 1$ | $\mathcal{C}$: CORRECT |
|          | $C \sim \text{Exp}(\exp(-A + Z_1 - Z_2))$ | |
|          | $\lambda_j(t) = 0.1A + 0.3 + |Z_1| + \log(1 + |Z_2|)$ | $B$: WRONG |
| 10       | $Z_1 \sim \mathcal{N}(0,1), Z_2 \sim \mathcal{N}(Z_1, 1)$ | $A_{\logit}$: WRONG and $A_{\twang}$ |
|          | $\logit \, \pi(Z) = 0.25(Z_1 - Z_2) + 0.5Z_1 Z_2 - 1$ | $\mathcal{C}$: CORRECT |
|          | $C \sim \text{Exp}(\exp(-A + Z_1 - Z_2))$ | |
|          | $\lambda_j(t) = 0.1A + 0.3 + |Z_1| + \log(1 + |Z_2|)$ | $B$: WRONG |

Note: $A_{\logit}$: $\pi(z; a) = \expit(a^\top z)$; $A_{\twang}$: twang; $A_{\logit}^*$: $\pi(z; a) = \expit(a^\top z + a^\top z_1 z_2)$; $B$: $\lambda_j(t|z; G_j, \gamma_j) = G_j(t) + \gamma_j^\top z, j = 1, 2$; $\mathcal{C}$: $S_j(t|a, z; \eta, \Lambda_j) = \exp \{-\Lambda_j(t)e^{\gamma}d\}$ where $d = (a, z)^\top$.

Abbreviation: RSF, random survival forest.

and Score 1 in (A4) with the estimated $S_c$. Model-based standard errors are used to construct 95% confidence intervals in all cases.

The results of the simulations are reported in Table 4. It is interesting to note that although censoring depends on $A$ and $Z$, the simplified scores using (8) and (9) in general perform better than (A4) with the estimated $S_c$, which has generally over 10% bias except when the random survival forest is used to estimate $S_c$ in Scenario 8. In Scenario 8, when the proportional hazards model, which is wrong, is used to estimate $S_c$, the model-based SE underestimates SD, leading to substantial undercoverage of the CI’s. On the other hand, when the random survival forest is used to estimate $S_c$, the bias becomes small and the coverage is relatively accurate, so that the performance of (A4) is similar to those of (8) and (9). In Scenario 10 where both sets of the models are wrong, like previously with independent censoring, when logistic regression is used for the propensity score, the proposed estimators are biased; but when ‘twang’ is used, the proposed estimators have little bias, with good coverage of the CI’s.

In the Supplementary Material we also provide the results under the heavier censoring of around 50%. We notice that in Scenario 28, the CIs based on the simplified scores seem to undercover a little, not having properly accounted for covariate dependent censoring, as compared to Scenario 8 where the censoring is around 10-30%. The performance of Score 1 using $S_c$, on the other hand, appears comparable to that in Scenario 8.

6 | APPLICATION

Here we study the effect of mid-life alcohol exposure on late life development of cognitive impairment. Cognitive impairment is assessed using the Cognitive Assessment and Screening Instrument (CASI), collected from the participants
starting in 1991 during the HAAS period. A score below 74 is considered moderate impairment, which is the event of interest. The data set consist of 1881 observations with normal cognitive functions at the start of HAAS, which is considered baseline for this competing risks analysis.

Mid-life alcohol exposure was assessed during the HHP period between 1965 and 1974, and is divided into two groups of 1390 light drinkers, and 491 heavy drinkers at some point during mid-life. We note that the setting of this data may be seen as semi-competing risks, which is under investigation separately by our group. For the purposes of this work, we consider the two competing risks: cognitive impairment and death without impairment. At the end of follow-up, among light drinkers 557 (40%) had developed cognitive impairment and 474 (34%) had died without impairment, while among heavy drinkers 216 (44%) had developed cognitive impairment and 163 (33%) had died without impairment. The cumulative incidence function curves for the two groups are presented in Figure 1.

The covariates used in this study are maximum years of education, age, systolic blood pressure and heart rate at the start of HHP, and ApoE genotype. ApoE is known to be related to Alzheimer’s disease (AD) and AD related dementia. In addition, since CASI at baseline (ie, start of HAAS in 1991) is post mid-life alcohol exposure, it might be considered as

| Scenario | PS | Score 1 - Simplified | | Score 1 with $\hat{S}_c$ | | Score 2 - Simplified | |
|----------|----|----------------------|---|----------------------|---|----------------------|---|
|          |     | Bias      SD    SE    CP |   | Bias      SD    SE    CP |   | Bias      SD    SE    CP |
| 6        | $\beta_1$ logistic | -0.021 | 0.157 | 0.181 | 0.97 | -0.032 | 0.163 | 0.180 | 0.97 | -0.006 | 0.154 | 0.180 | 0.98 |
|          | twang | -0.021 | 0.159 | 0.186 | 0.97 | -0.030 | 0.166 | 0.184 | 0.97 | -0.007 | 0.157 | 0.184 | 0.97 |
|          | $\beta_2$ logistic | -0.018 | 0.160 | 0.182 | 0.97 | -0.026 | 0.168 | 0.181 | 0.96 | -0.003 | 0.158 | 0.180 | 0.97 |
|          | twang | -0.018 | 0.163 | 0.186 | 0.97 | -0.026 | 0.171 | 0.185 | 0.95 | -0.003 | 0.162 | 0.184 | 0.97 |
| 7        | $\beta_1$ logistic | -0.009 | 0.115 | 0.113 | 0.95 | -0.012 | 0.115 | 0.119 | 0.96 | -0.006 | 0.115 | 0.113 | 0.95 |
|          | twang | -0.009 | 0.120 | 0.117 | 0.95 | -0.013 | 0.118 | 0.123 | 0.96 | -0.005 | 0.119 | 0.117 | 0.95 |
|          | $\beta_2$ logistic | -0.011 | 0.127 | 0.113 | 0.92 | -0.013 | 0.124 | 0.119 | 0.94 | -0.008 | 0.127 | 0.113 | 0.93 |
|          | twang | -0.013 | 0.131 | 0.116 | 0.91 | -0.014 | 0.128 | 0.123 | 0.93 | -0.009 | 0.131 | 0.116 | 0.91 |
| 8        | $\beta_1$ logistic | -0.007 | 0.126 | 0.136 | 0.96 | -0.012 | 0.128 | 0.120 | 0.83 | -0.000 | 0.127 | 0.135 | 0.95 |
|          | twang | -0.007 | 0.128 | 0.140 | 0.96 | -0.012 | 0.129 | 0.122 | 0.83 | 0.001 | 0.129 | 0.139 | 0.97 |
|          | $\beta_2$ logistic | -0.008 | 0.128 | 0.136 | 0.97 | -0.014 | 0.131 | 0.124 | 0.82 | -0.002 | 0.129 | 0.135 | 0.97 |
|          | twang | -0.009 | 0.135 | 0.140 | 0.96 | -0.014 | 0.138 | 0.126 | 0.81 | -0.003 | 0.135 | 0.139 | 0.96 |
| 9        | $\beta_1$ logistic | 0.001 | 0.170 | 0.160 | 0.93 | -0.030 | 0.184 | 0.197 | 0.96 | 0.012 | 0.168 | 0.161 | 0.94 |
|          | twang | 0.011 | 0.168 | 0.159 | 0.94 | -0.024 | 0.185 | 0.197 | 0.96 | 0.020 | 0.166 | 0.160 | 0.95 |
|          | $\beta_2$ logistic | 0.002 | 0.164 | 0.160 | 0.95 | -0.034 | 0.177 | 0.197 | 0.97 | 0.013 | 0.163 | 0.161 | 0.94 |
|          | twang | 0.014 | 0.160 | 0.159 | 0.95 | -0.026 | 0.176 | 0.198 | 0.97 | 0.022 | 0.160 | 0.159 | 0.94 |
| 10       | $\beta_1$ logistic | 0.348 | 0.161 | 0.167 | 0.44 | 0.253 | 0.180 | 0.193 | 0.75 | 0.358 | 0.168 | 0.161 | 0.40 |
|          | twang | 0.016 | 0.160 | 0.159 | 0.95 | -0.022 | 0.178 | 0.184 | 0.98 | 0.023 | 0.161 | 0.160 | 0.94 |
|          | $\beta_2$ logistic | 0.338 | 0.166 | 0.166 | 0.47 | 0.233 | 0.267 | 0.193 | 0.76 | 0.349 | 0.171 | 0.160 | 0.41 |
|          | twang | 0.014 | 0.171 | 0.159 | 0.95 | -0.038 | 0.273 | 0.184 | 0.97 | 0.022 | 0.170 | 0.159 | 0.93 |

Note: Column PS indicates the working model for the propensity score. For Scenario 8, the first row of “Score 1 with $\hat{S}_c$” uses the Cox model to estimate $\hat{S}_c$ while the second row uses the random survival forest.

Abbreviation: CP, coverage of the 95% confidence interval; SD, standard deviation; SE, standard error.

*Median is reported when the distribution of SE is left skewed, that is, with heavy right tail.

The covariates used in this study are maximum years of education, age, systolic blood pressure and heart rate at the start of HAAS, and ApoE genotype. ApoE is known to be related to Alzheimer’s disease (AD) and AD related dementia. In addition, since CASI at baseline (ie, start of HAAS in 1991) is post mid-life alcohol exposure, it might be considered as
a mediator for the later development of cognitive impairment. Under the additive effect model (1), similar to Lange and Hansen $^{38}$ and VanderWeele $^{39}$ if $M^a$ is the potential value of the mediator under treatment $a = 0, 1$, we have

$$h_j(t|A = 1, M^1, Z) - h_j(t|A = 0, M^0, Z) = \beta_j + \lambda_j(t|M^1, Z) - \lambda_j(t|M^0, Z) = \beta_j + h_j(t|A = 1, M^1, Z) - h_j(t|A = 1, M^0, Z),$$

for $j = 1, 2$. The above gives the usual decomposition of the total effect as the difference of the hazards in the left-hand side of the above, so that $\beta_j (j = 1, 2)$ may be seen as the direct effect of mid-life alcohol exposure on the outcome (ie, competing risk) of interest, when we include CASI at baseline in the regression model (1). For estimation of the total effect on the left-hand side above, if we make the standard consistency assumption (ie, $T^{u, m} = T$ if $A = a$, $M = m$ and $T^{u} = T$ if $A = a$) and the composition assumption $^{40}$ for mediators (ie, $T^{u, M^m} = T^u$), where we again use the superscripts to indicate the potential values, it can be shown that

$$h_j(t|A = a, M^a, Z) = \tilde{h}_j(t|A = a, Z)$$

for $j = 1, 2$ and $a = 0, 1$, where $\tilde{h}_j$ denotes the conditional cause-specific hazard when the mediator is not included in the regression model. Therefore the exposure effect from the latter, that is, when CASI at baseline is not included in model (1), may be considered as the total effect of mid-life alcohol on the outcome.

In order to estimate the above effects we use both logistic regression without interaction and “twang” to estimate the propensity score. An investigation (data not shown here) of the censoring distribution for different groups defined by the exposure and the covariates suggest that the stronger assumption of $C \perp (T, A)|Z$ may not hold here, although with the sparsity of censoring events in some of these groups we do not have good power to perform a formal hypothesis test.

We utilize the scores studied in the above simulations to estimate the effect of mid-life alcohol exposure on the development of moderate cognitive impairment and on the competing risk of death without cognitive impairment. The censoring distribution is estimated using the proportional hazards model. The results of the analysis are reported in Table 5. The results are similar quantitatively regardless of the estimation method, and seem to indicate that mid-life alcohol exposure has a significant effect on both the development of cognitive impairment and death without cognitive impairment, where the hazards are both increased (total effects). While there seems to be no obvious difference between the estimated total and direct effect on death without cognitive impairment, the estimate direct effect is visibly less than the estimated total effect of alcohol on late life cognitive impairment, once the baseline CASI score has been accounted for. In other words, mid-life alcohol exposure conceivably contributed to late life cognitive impairment both through its earlier impact on cognitive function as well as through its sustained (ie direct) impact later in life.
TABLE 5 Estimated treatment effect for the HHP-HAAS data

| Outcome | PS | Effect   | Score 1 | 95% CI     | Score 1-Cens | 95% CI     | Score 2 | 95% CI     |
|---------|----|----------|---------|------------|--------------|------------|---------|------------|
|         |    |          | $\hat{\beta}$ |            | $\hat{\beta}$ |            | $\hat{\beta}$ |            |
| Cog. imp. | logistic | Total | 0.013 | [0.004, 0.022] | 0.015 | [0.005, 0.025] | 0.012 | [0.004, 0.021] |
|         |    | Direct  | 0.010 | [0.001, 0.019] | 0.009 | [0.000, 0.019] | 0.009 | [0.001, 0.018] |
| twang   |    | Total  | 0.012 | [0.003, 0.020] | 0.013 | [0.004, 0.023] | 0.011 | [0.003, 0.020] |
|         |    | Direct  | 0.008 | [0.000, 0.017] | 0.008 | [−0.002, 0.017] | 0.008 | [0.000, 0.016] |
| Death   | logistic | Total | 0.012 | [0.005, 0.020] | 0.013 | [0.004, 0.022] | 0.012 | [0.005, 0.019] |
|         |    | Direct  | 0.012 | [0.005, 0.020] | 0.013 | [0.004, 0.022] | 0.012 | [0.005, 0.019] |
| twang   |    | Total  | 0.012 | [0.004, 0.020] | 0.012 | [0.003, 0.022] | 0.011 | [0.004, 0.019] |
|         |    | Direct  | 0.010 | [0.003, 0.018] | 0.011 | [0.002, 0.020] | 0.010 | [0.003, 0.018] |

Note: Column PS indicates the working model for the propensity score. Every first row of 95% CI is model-based, second row uses bootstrap. Abbreviation: CI, confidence interval.

7 | DISCUSSION

In this article we have proposed two doubly robust estimators for the conditional cause-specific hazard difference under competing risks. We proposed two estimators that are model doubly robust: they are consistent and asymptotically normal if both the propensity score and the censoring distribution models are correctly specified, or if the outcome models for the competing risks are correctly specified. In addition, they are rate doubly robust: they are consistent and asymptotically normal if both sets of models are correctly specified and the product of their convergence rates is $o(1/\sqrt{n})$. Our empirical investigation showed comparable performances of the two estimators; it is plausible that Score 1 has more guaranteed numerical stability as it has a closed-form solution.

Rate double robustness gives the user the possibility to use modern nonparametric methods, which are known to have rates of convergence slower than $1/\sqrt{n}$. In simulations we showed the performance of the proposed estimators when gradient boosted method is used for estimation of the propensity score, as well as random survival forest for estimation of the censoring distribution. The theoretical results in this paper require certain rate conditions, and these are not always established for a given machine learning or nonparametric estimator. Cui et al.\textsuperscript{41} established that a rate of $n^{-1/2+\tau}$ is achievable by random survival forest, where $d$ is the number of covariates. Kooperberg et al.\textsuperscript{42} derived the $L^2$ rate of convergence for splines and showed that, under some conditions, the rate can reach $n^{-p/(2p+d)}$, where $p$ is a smoothness parameter. Rates were also established for other machine learning methods; for example, a uniform rate for regression trees was shown in Wager and Walther;\textsuperscript{43} while a root-mean-square rate was derived for neural networks.\textsuperscript{44} The rates, of course, depend on the tuning parameter values. Our simulation studies are empirical investigations that complement these theoretical results.

In the absence of competing risks, McKeague and Sasieni\textsuperscript{45} proposed a partly parametric additive risk model that can be seen as a special case of (1), where the covariate effects are linear but allowed to vary over time. It is possible then, to use that in place of (12) as the working model, with the corresponding $c_n = n^{-1/2}$ just as under (12). Also in the absence of competing risks, Hou et al.\textsuperscript{5} proposed in their discussion to estimate nonparametrically the cumulative hazard function separately for the treated and the untreated, and then combine them using some weights to estimate what corresponds to
\(
A_j(t, Z) \text{ under our model (1). The procedure was not implemented or further investigated. For competing risks Ishwaran et al}^{16} \text{ proposed survival random forest for estimation of both cumulative cause-specific hazard functions and cumulative incidence functions; these might be adapted in an approach similar to Hou et al.}\text{.}^8 \text{ This would be of interest for future work.}

Recently further considerations of the competing event as a mediator to the event type of interest were described in Young et al,\text{47} Stensrud et al\text{48} and Martinussen and Stensrud,\text{49} in the potential outcomes setting. Our application here focused on the total effect of alcohol on cognitive impairment, while acknowledging “truncation by death”. The mediation type analysis would further disentangle causal exposure effects on cognitive impairment and death, along the lines of the direct effect of alcohol on cognitive impairment and the indirect effect mediated through death. The latter might sound counter-intuitive initially, and we refer interested readers to the literature referenced above. We note that as with any such decomposition of total effects into direct and indirect effects, additional assumptions are needed.

In simulations we have seen that the simple estimators using (8) and (9) appear to be somewhat robust when the stronger independent censoring assumption is violated, at least when censoring is not heavy. On the other hand, when the censoring distribution is estimated, the model-based confidence intervals can have inaccurate coverage if the censoring model is misspecified. Using random survival forest to estimate the censoring distribution improves the performance of the treatment effect estimate. In practice bootstrap variance estimate might be used to construct confidence intervals in general.

ACKNOWLEDGEMENTS

This research was partially supported by NIH/NIA grant R03 AG062432. We thank Drs. Steve Edland and Lon White for helpful discussion regarding the HHP-HAAS data, and Ms. Yiran Zhang for preparation of the data.

DATA AVAILABILITY STATEMENT

The R codes developed in this work have been implemented in the R package “HazardDiff” and are publicly available on CRAN (https://CRAN.R-project.org/package=HazardDiff).

ORCID

*Ronghui Xu* https://orcid.org/0000-0002-2822-0561

REFERENCES

1. Ding P, Geng Z, Yan W, Zhou XH. Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *J Am Stat Assoc*. 2011;106(496):1578-1591.
2. Tchetgen Tchetgen EJ. Identification and estimation of survivor average causal effects. *Stat Med*. 2014;33(21):3601-3628.
3. Estes JP, Nguyen DV, Dalrymple LS, Mu Y, Şentürk D. Time-varying effect modeling with longitudinal data truncated by death: conditional models, interpretations, and inference. *Stat Med*. 2016;35(11):1834-1847.
4. Yang P, Small DS. Using post-outcome measurement information in censoring-by-death problems. *J R Stat Soc Series B Stat Methodology*. 2016;78(1):299-318.
5. Aalen OO. A model for nonparametric regression analysis of counting processes. New York: Springer; 1980:1-25.
6. Aalen OO. A linear regression model for the analysis of life times. *Stat Med*. 1989;8(8):907-925.
7. Dukes O, Martinussen T, Tchetgen Tchetgen EJ, Vansteelandt S. On doubly robust estimation of the hazard difference. *Biometrics*. 2019;75:100-119.
8. Hou J, Bradic J, Xu R. Treatment effect estimation under additive hazards models with high-dimensional confounding. *J Am Stat Assoc*. 2021. doi:10.1080/01621459.2021.1930546.
9. Robins JM, Ritov Y. Towards a curse of dimensionality appropriate (CODA) asymptotic theory for semiparametric models. *Stat Med*. 1997;16:285-319.
10. Robins JM, Rotnitzky A. Semiparametric efficiency in multivariate regression models with missing data. *J Am Stat Assoc*. 1995;90(429):122-129.
11. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J Am Stat Assoc*. 1995;90(429):106-121.
12. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J Am Stat Assoc*. 1999;94(448):1096-1120.
13. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550-560.
14. Robins JM. *Marginal Structural Models Versus Structural Nested Models as Tools for Causal Inference*. New York: Springer; 2000:95-133.
15. Robins JM, Rotnitzky A, Laan v DM. On profile likelihood: comment. *J Am Stat Assoc*. 2000;95(450):477-482.
16. Van Der Laan MJ, Laan M, Robins JM. *Unified methods for censored longitudinal data and causality*. New York: Springer Science & Business Media; 2003.
17. Robins JM, Rotnitzky A. Comment on “Inference for semiparametric models: Some questions and an answer”. *Stat Sci*. 2001;11(4):920-936.
18. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*. 2005;61(4):962-973.
19. Tsiatis A. *Semiparametric theory and missing data*. New York: Springer; 2006.
20. Tchetgen Tchetgen EJ, Robins JM, Rotnitzky A. On doubly robust estimation in a semiparametric odds ratio model. *Biometrika*. 2010;97(1):171-180.
21. Zhang M, Schaubel DE. Contrasting treatment-specific survival using double-robust estimators. *Stat Med*. 2012;31(30):4255-4268.
22. Bae X, Tsiatis AA, Lu W, Song R. Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. *Lifitime Data Anal*. 2017;23(4):585-604.
23. Sjölander A, Vansteelandt S. Doubly robust estimation of attributable fractions in survival analysis. *Stat Methods Med Res*. 2017;26(2):948-969.
24. Bai X, Tsiatis AA, Lu W, Song R. Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. *Lifetime Data Anal*. 2017;23(4):585-604.
25. Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, Van Der Laan M. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *JCausal Inference*. 2014;2(2):147-185.
26. Zheng W, Petersen M, Van Der Laan MJ. Doubly robust and efficient estimation of marginal structural models for the hazard function. *Int J Biostat*. 2016;12(1):233-252.
27. Holt J. Competing risk analyses with special reference to matched pair experiments. *Biometrika*. 1978;65(1):159-165.
28. Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. 2nd ed. New York: John Wiley & Sons; 2011.
29. Shen Y, Cheng S. Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*. 1999;55(4):1093-1100.
30. Bickel PJ, Klaassen CAJ, Lewis TM, Ritov Y, Wellner JA. *Efficient and adaptive estimation for semiparametric models*. Baltimore MD: Johns Hopkins University Press; 1993.
31. Newey WK. Semiparametric efficiency bounds. *J Appl Econom*. 1990;5(2):99-135.
32. Newey WK. The asymptotic variance of semiparametric estimators. *Econometrica*. 1994;62:1349-1382.
33. Lin DY, Ying Z. Semiparametric analysis of the additive risk model. *Biometrika*. 1994;81:61-71.
34. Smucler E, Rotnitzky A, Robins JM. A unifying approach for doubly-robust $L_1$ regularized estimation of causal contrasts. arXiv preprint, arXiv:1904.03737, 2019.
35. Cefalu M, Riddaway G, McCaffrey D, Morral A, Griffin BA, Burgette L. *twang: Toolkit for Weighting and Analysis of Nonequivalent Groups (version 2.3)*. CRAN R package . 2021.
36. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2(3):841-860.
37. Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001;88(4):907-919.
38. Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology*. 2011;22:575-581.
39. VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology*. 2011;22(4):582.
40. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface*. 2009;2(4):457-468.
41. Cui Y, Zhu R, Zhou M, Kosorok M. Consistency of survival tree and forest models: splitting bias and correction. *Statistica Sinica*. 2022;32(3):1245-1268.
42. Kooperberg C, Stone CJ, Truong YK. The $L_2$ rate of convergence for hazard regression. *Scand J Stat*. 1995;90:78-94.
43. Wager S, Walther G. Adaptive concentration of regression trees, with application to random forests. arXiv preprint, arXiv:1503.06388, 2015.
44. Chen X, White H. Improved rates and asymptotic normality for nonparametric neural network estimators. *IEEE Trans Inf Theory*. 1999;45:682-691.
45. McKeague IW, Sasieni PD. A partly parametric additive risk model. *Biometrika*. 1994;81:501-514.
46. Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. *Biostatistics*. 2014;15(4):757-773.
47. Young JG, Stensrud MJ, Tchetgen Tchetgen E, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med*. 2020;39(8):1199-1236.
48. Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA. Separable effects for causal inference in the presence of competing events. *J Am Stat Assoc*. 2022;117:175-183. doi:10.1080/01621459.2020.1765783
49. Martinussen T, Stensrud MJ. Estimation of separable direct and indirect effects in continuous time. *Biometrics*. 2022;78. doi:10.1111/biom.13559

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.
APPENDIX

Below are expressions from Section 4.

For $\gamma$ and $G$ in the linear working models (12) we have

\[
\hat{\gamma}_j = \left[ n \sum_{i=1}^{n} \int_{0}^{\tau} Y_i(t)(Z_i - \bar{Z}(t)) \otimes^2 dt \right]^{-1} \left[ n \sum_{i=1}^{n} \int_{0}^{\tau} Y_i(t)(Z_i - \bar{Z}(t)) dN_{ji}(t) \right],
\]

(A1)

where $\bar{Z}(t) = \{ \sum_{i=1}^{n} Y_i(t) \}^{-1} \sum_{i=1}^{n} Y_i(t)Z_i$, $Z \otimes^2 = ZZ^T$, and

\[
\hat{G}_j(t; \hat{\beta}_j, \hat{\gamma}_j) = \int_{0}^{t} \sum_{i=1}^{n} \left\{ dN_{ji}(u) - Y_i(u)\hat{\beta}_jA_idu - Y_i(u)\gamma_j^TZ_idu \right\} \sum_{i=1}^{n} Y_i(u).
\]

(A2)

The weighted Breslow estimator is

\[
\tilde{G}_j(t; \hat{\beta}_j, \hat{\gamma}_j, \hat{S}_c, \hat{\pi}) = \int_{0}^{t} \sum_{i=1}^{n} w_i(S_c, \pi) \left\{ dN_{ji}(u) - Y_i(u)\hat{\beta}_jA_idu - Y_i(u)\gamma_j^TZ_idu \right\} \sum_{i=1}^{n} w_i(S_c, \pi)Y_i(u),
\]

(A3)

where $w_i(S_c, \pi) = A_i \{ 1 - \pi(Z_i) \} S_c^{-1}(u|A_i, Z_i)$. This leads to the closed-form solution to $S_{1,n}$ for $j = 1, \ldots, J$:

\[
\hat{\beta}_j^{(1)} = -\left\{ \sum_{i=1}^{n} \int_{0}^{\tau} \hat{S}_c^{-1}(t|A_i, Z_i)(1 - A_i)\hat{\pi}(Z_i)Y_i(t)dt \right\}^{-1} \times \sum_{i=1}^{n} \int_{0}^{\tau} \hat{S}_c^{-1}(t|A_i, Z_i)
\cdot (1 - A_i)\hat{\pi}(Z_i) \cdot \left\{ dN_{ji}(t) - Y_i(t) \left[ \hat{\gamma}_j^T \left\{ Z_i - \bar{Z}(t; \hat{S}_c, \hat{\pi}) \right\} dt + d\bar{N}_j(t; \hat{S}_c, \hat{\pi}) \right] \right\},
\]

(A4)

where

\[
\bar{Z}(t; \hat{S}_c, \hat{\pi}) = \sum_{i=1}^{n} Y_i(t)Z_iw_i(\hat{S}_c, \hat{\pi}) \sum_{i=1}^{n} Y_i(t)w_i(\hat{S}_c, \hat{\pi}), \quad d\bar{N}_j(t; \hat{S}_c, \hat{\pi}) = \sum_{i=1}^{n} dN_{ji}(t)w_i(\hat{S}_c, \hat{\pi}) \sum_{i=1}^{n} Y_i(t)w_i(\hat{S}_c, \hat{\pi}).
\]