



Published in final edited form as:

Stat Med. 2017 October 15; 36(23): 3683–3707. doi:10.1002/sim.7350.

Semiparametric regression on cumulative incidence function with interval-censored competing risks data

Giorgos Bakoyannis^{a,*}, Menggang Yu^b, and Constantin T. Yiannoutsos^a

^aDepartment of Biostatistics, Fairbanks School of Public Health and School of Medicine, Indiana University, Indianapolis, IN, USA

^bDepartment of Biostatistics & Medical Informatics, University of Wisconsin-Madison, WI, USA

Abstract

Many biomedical and clinical studies with time-to-event outcomes involve competing risks data. These data are frequently subject to interval censoring. This means that the failure time is not precisely observed, but is only known to lie between two observation times such as clinical visits in a cohort study. Not taking into account the interval censoring may result in biased estimation of the cause-specific cumulative incidence function, an important quantity in the competing risk framework, used for evaluating interventions in populations, for studying the prognosis of various diseases and for prediction and implementation science purposes. In this work we consider the class of semiparametric generalized odds-rate transformation models in the context of sieve maximum likelihood estimation based on B-splines. This large class of models includes both the proportional odds and the proportional subdistribution hazard models (i.e., the Fine-Gray model) as special cases. The estimator for the regression parameter is shown to be semiparametrically efficient and asymptotically normal. Simulation studies suggest that the method performs well even with small sample sizes. As an illustration we use the proposed method to analyze data from HIV-infected individuals obtained from a large cohort study in sub-Saharan Africa. We also provide the R function `ciregic` that implements the proposed method and present an illustrative example.

Keywords

Competing risks; Cumulative incidence function; Interval censoring; Semiparametric efficiency; R function

1. Introduction

Many biomedical studies with time-to-event outcomes involve competing risk data. The basic identifiable quantities from such data, are the cause-specific hazard and the cause-specific cumulative incidence function. The latter quantity plays a key role in the evaluation of interventions on populations, in studying the prognosis of various diseases, and in prediction. Prediction in turn, is especially useful in quality of life studies and for general

*Correspondence to: Giorgos Bakoyannis, Department of Biostatistics, Fairbanks School of Public Health and School of Medicine, Indiana University, 410 West 10th Street, Indianapolis, IN 46202, USA (gbakogia@iu.edu).

implementation science purposes. Frequently, competing risks data are subject to interval censoring, that is, the times to the occurrence of an event are only known to lie between two observation times. An important example of interval-censored competing risk data, and the motivation for the present paper, comes from studies in HIV care and treatment programs. Individuals living with HIV who receive care in these programs may die while in care or become lost to care. Patients not retained in care are less likely to receive antiviral treatment (ART) and, frequently, any other clinical services. They are thus more likely to experience adverse clinical outcomes including death and viral rebound. The resulting increased infectivity of these patients may in addition significantly contribute to further expansion of the HIV epidemic in their communities. Identifying factors associated with mortality while in care and loss to care is useful for optimizing interventions, especially in sub-Saharan Africa, the region most affected by the HIV epidemic and the locus of the present study.

In many cases in this context, when a death has occurred, the corresponding death date is known to be between the date the death was reported and the last clinic visit of the deceased patient. Also, since loss to care is defined as no clinic visits for a certain period (e.g., three months from the last observed clinic visit or for the maximum period that a patient may have adequate medication supplies), the actual date of disengagement from care (being alive but not receiving anti-HIV care) is unknown. Consequently, studies in this context involve interval-censored competing risk data where death and loss to care are competing events. This research was motivated from studies in this setting and the fact that the issue of interval censoring in this context has, to our knowledge, been largely ignored in the literature.

Both nonparametric and fully parametric approaches have been proposed to estimate the cause-specific cumulative incidence function with interval-censored competing risk data. For example, Hudgens et al. [1] proposed a nonparametric maximum likelihood estimator for the cause-specific cumulative incidence under interval censoring. More recently, a parametric estimation method for the cause-specific cumulative incidence under a Gompertz distribution assumption was developed by Hudgens et al. [2]. These methods however do not consider covariates. To incorporate covariate information into a regression setting, fully parametric likelihood methods for the cause-specific cumulative incidence can be modified to deal with interval-censored competing risk data. For example, Jeong and Fine [3] proposed a parametric maximum likelihood approach along with a generalized odds-rate transformation [4] to model the cause-specific cumulative incidence under the Gompertz distribution. On the other hand, Cheng [5] proposed a three-parameter modified logistic model, while Shi et al. [6] considered a parametric method for explicitly modeling the cause-specific cumulative incidence of the cause of interest based on the generalized odds-rate transformation [4] and the modified three-parameter logistic function proposed by Cheng [5].

A key limitation of the fully parametric methods is that they make strong distributional assumptions about the true cause-specific cumulative incidence functions. Additionally, using more flexible three-parameter distributions may be challenging computationally and may lead to non-convergence problems as in univariate survival analysis, as cautioned by Lambert et al. [7] for example. A general approach to overcome the limitations of the fully parametric methods is to consider a semiparametric model. This approach was used

successfully by Zhang et al. [8] for the Cox model with interval-censored (univariate) survival data. The authors used a sieve maximum likelihood estimation methodology [9] along with monotone B-splines to approximate the baseline hazard of the Cox proportional hazards model. Recently, Li [10] used the estimation approach by Zhang et al. [8] to the problem of semiparametric regression of the cumulative incidence function with interval-censored competing risk data under the proportional subdistribution hazards or the Fine-Gray model [11]. However, the assumption of proportional subdistribution hazards may not be correct and also the resulting parameter estimates are somewhat difficult to interpret as reported by Fine [12]. As we show in the illustration of our proposed methodology, using a proportional subdistribution hazard model for both competing risks did not provide a good fit to the data in our motivating HIV study compared to other models which better fit these data. Consequently, considering a more general class of semiparametric models for the cumulative incidence function, which also includes models which are easily interpretable by practitioners, such as, for example, the proportional odds model [12, 13], has important practical implications.

The main goal of this paper is to extend the B-spline sieve maximum likelihood estimator for interval-censored data used by [8] and [10], to the class of semiparametric generalized odds rate transformation models for the cause-specific cumulative incidence function and evaluate its asymptotic and finite sample properties. The large class of the generalized odds rate transformation models includes both the proportional odds model and the proportional subdistribution hazards model (i.e., the Fine-Gray model) [11] as special cases. Additionally, we explicitly incorporate the boundedness constraint of the cumulative incidence functions during the likelihood maximization. This is not the case with many existing methods for right-censored competing risk data that analyze each cause of failure separately and which can produce cumulative incidence estimates having a sum, over all possible causes of failure, which is greater than 1. For example, in the illustration presented by Choi and Huang [14], the analysis based on the Fine-Gray model [11] produced cumulative incidence estimates that added to 1.29. This may indicate a model misspecification and may adversely influence the validity of any predictions based on these estimates. The boundedness constraint was not considered by Li [10]. Finally, we develop and illustrate the R function `ciregic` that easily implements the proposed estimation methodology for general interval-censored competing risks data.

We show that the B-spline sieve maximum likelihood estimator for the regression (Euclidean) parameter is shown to be semiparametrically efficient and asymptotically normal. We also demonstrate that the estimator for the functional parameter (baseline cause-specific cumulative incidence functions) converges at the optimal rate for nonparametric regression, which is nevertheless slower than the usual parametric \sqrt{n} rate. Simulation studies suggest that the method performs well even with small sample sizes. We illustrate our methods by considering a number of candidate models like the proportional subdistribution and proportional odds model in the analysis of death and loss to care in the context of data from HIV care and treatment programs. Finally, we conclude this manuscript with a discussion, while examples on how to use the R function `ciregic` created to fit these models and asymptotic theory proofs are reserved for the Appendix.

2. Data and method

We assume the existence of a finite number k of competing risks. The cause of failure is denoted by C , the actual failure time is denoted by T , whereas (U_1, \dots, U_m) denote the $m \in (0, \infty)$ distinct observation times. Observation times could be clinical visits, laboratory tests and all other occasions where the sampling unit (e.g., a patient) is actually observed. The number of observation times m may differ from subject to subject. To facilitate our further discussion let $V \in \{0, U_1, \dots, U_m\}$ correspond to the last observation time prior to the failure and $U \in \{U_1, \dots, U_m, \infty\}$ the first observation time after the failure. Using this notation a left-censored observation will correspond to $(V, U) = (0, U_1)$, while a right-censored observation to $(V, U) = (U_m, \infty)$. For $j = 1, \dots, k$, if a subject fails from the j -th cause of failure before the first observation time U_1 (i.e. it is left-censored), we observe $\delta_j^1 = 1$. If he/she fails between $V > U_1$ and $U = U_m$, we observe $\delta_j = 1$. But if he/she is right-censored (i.e. $T > U_m$), then we observe $\delta = \sum_{j=1}^k (\delta_j + \delta_j^1) = 0$. We assume the observation interval to be $[a, b]$, that is $a = U_1 < U_m = b$. Along with the $C, V, U, \{\delta_j\}_{j=1}^k$ and $\{\delta_j^1\}_{j=1}^k$, we also observe a vector of covariates $Z \in \mathbb{R}^d$, whose effects are the primary parameters of interest. We further assume the following two fundamental conditions:

- A1** The observation times (U_1, \dots, U_m) are independent of (T, C) conditional on Z .
- A2** The distribution of (U_1, \dots, U_m) does not contain the parameters that govern the distribution of (T, C) (non-informative interval censoring).

The cause-specific cumulative incidence function is defined as

$$F_j(t; z) = \Pr(T \leq t, C = j | Z = z), \quad j = 1, \dots, k$$

and corresponds to the cumulative probability of failure from the j -th cause of failure in the presence of the remaining causes of failure, conditional on the covariates. Given the two assumptions above, the likelihood function in terms of the cause-specific cumulative incidence functions is

$$L(\theta; D) \propto \prod_{i=1}^n \left\{ \prod_{j=1}^k [F_j(u_i; z_i, \theta_j) - F_j(v_i; z_i, \theta_j)]^{\delta_{ij}} \right\} \left\{ \prod_{j=1}^k [F_j(v_i; z_i, \theta_j)]^{\delta_{ij}^1} \right\} \times \left[1 - \sum_{j=1}^k F_j(u_i; z_i, \theta_j) \right]^{1 - \delta_i}, \quad (1)$$

where $\theta = (\theta'_1, \dots, \theta'_k)'$ is a parameter that includes, in our case, regression coefficients for the effect of the covariate vector Z on each cause-specific cumulative incidence function and unspecified functions of time. A natural choice for the model for F_j is the class of semiparametric transformation models [11, 3, 15]:

$$g_j [F_j(t; z)] = \phi_j(t) + \beta_j' z, \quad j=1, \dots, k \quad (2)$$

where g_j is a known increasing cause-specific link function and ϕ_j is an unspecified, strictly increasing and invertible function of time. A special subset of the class of semiparametric transformation models is the class of the generalized odds rate transformation models [4, 16, 12, 3]:

$$g_j(F_j; \alpha_j) = \begin{cases} \log \left[\frac{(1-F_j)^{-\alpha_j} - 1}{\alpha_j} \right] & \text{if } 0 < \alpha_j < \infty \\ \log [-\log(1-F_j)] & \text{if } \alpha_j = 0 \end{cases}$$

for $j = 1, \dots, k$. The proportional odds model and the proportional subdistribution hazards model [11] are special cases of the generalized odds rate transformation with $\alpha_j = 1$ and $\alpha_j = 0$, respectively [3]. Note that the link functions for different causes of failure are allowed to be different. All authors considering general classes of models assume that the true link functions are known [16, 17, 12, 18] and we adopt this assumption in our present work as well. This is because estimating α is not always possible because not all models with unknown α are identifiable from the data as in the non-competing-risk setting [19]. Even when there is no non-identifiability issues, reliable estimation of α requires very large sample sizes [19, 18].

As in [8] and [10], we avoid imposing parametric assumptions on ϕ_j in (2) and thus the likelihood involves k infinite-dimensional or functional parameters. In general, maximization of the likelihood function with an infinite-dimensional parameter $\theta \in \Theta$ over Θ may lead to inconsistent maximum likelihood estimates [9]. One approach to overcome this problem is to use sieve maximum likelihood estimation. A sieve [9] is a sequence $\{\Theta_n\}_{n=1}$ of parameter spaces that approximate in a certain sense the original parameter space Θ , with the approximation error tending to zero as $n \rightarrow \infty$. A sieve maximum likelihood estimate is the estimate obtained by maximizing the likelihood function over Θ_n . Another practical advantage of using the sieve maximum likelihood approach is that it reduces the dimensionality of the optimization problem, and, thus, the computational burden, compared to a fully semiparametric likelihood approach [8]. This is because the dimension of Θ_n is significantly smaller (i.e., it involves fewer number of parameters to be estimated) compared to that of the full parameter space Θ in finite samples. The computational advantage of the sieve maximum likelihood approach compared to a fully semiparametric maximum likelihood approach for interval-censored survival data has been shown in the simulation study by Zhang et al. [8]. In our case, the sequence of approximating functional parameter spaces is chosen to be spaces of monotone (due to the monotonicity of the cumulative incidence function) B-spline functions as defined in [8], i.e.,

$$\mathcal{M}_n(\gamma_j, N_j, m_j) = \left\{ \phi: \phi(t; \gamma_j) = \sum_{s=1}^{N_j+m_j} \gamma_{j,s} B_{s,m_j}(t), \gamma_j \in \mathbb{R}^{N_j+m_j+1}, \gamma_{j,1} < \dots < \gamma_{j,N_j+m_j} \right\},$$

for $j = 1, \dots, k$, where N_j and m_j are the number of internal knots and the order of the B-spline for the j -th cause of failure, γ_j are the unknown control points and $t \in [a, b]$, with a and b being the minimum and maximum observation times in the study. The number of internal knots N_j increases with sample size and is selected to satisfy $N_j \approx n^\nu$ such that $\max_{1 \leq t \leq N_j+1} (c_{t,j} - c_{t-1,j}) = O(n^{-\nu})$, where $c_{t,j}$ is the place of the t -th knot for the j -th cause-specific cumulative incidence function. During the likelihood maximization we impose the monotonicity constraints

$$\gamma_{j,s} < \gamma_{j,s+1},$$

for every $s = 1, \dots, N_j + p_j$ and $j = 1, \dots, k$. Additionally, the constraint

$$\max_z \left\{ \sum_{j=1}^k F_j(b; z, \theta_j) \right\} < 1 \tag{3}$$

is needed to ensure that the sum of the estimated cumulative incidence functions at the maximum follow-up time b is bounded above by 1, where the maximum is over all the observed covariate patterns.

Let the true parameter values be denoted by $\theta_0 = (\beta'_0, \phi'_0)'$, where $\beta_0 = (\beta'_{1,0}, \dots, \beta'_{k,0})'$ and $\phi_0 = (\phi_{1,0}, \dots, \phi_{k,0})'$ and the corresponding estimator by $\hat{\theta}_n = (\hat{\beta}'_n, \hat{\phi}'_n)'$, where $\hat{\beta}_n = (\hat{\beta}'_{1,n}, \dots, \hat{\beta}'_{k,n})'$ and $\hat{\phi}_n = (\hat{\phi}'_{1,n}, \dots, \hat{\phi}'_{k,n})'$. Also, define the L_2 -metric for the distance between two parameters $\theta_1 = (\beta^{(1)'}, \phi^{(1)'})'$ and $\theta_2 = (\beta^{(2)'}, \phi^{(2)'})'$ as

$$d(\theta_1, \theta_2) = \left(\sum_{j=1}^k \|\beta_j^{(1)} - \beta_j^{(2)}\|^2 + \sum_{j=1}^k \|\phi_j^{(1)} - \phi_j^{(2)}\|_{\Phi}^2 \right)^{\frac{1}{2}},$$

where

$$\|\phi_j^{(1)} - \phi_j^{(2)}\|_{\Phi}^2 = E \left[\phi_j^{(1)}(V) - \phi_j^{(2)}(V) \right]^2 + E \left[\phi_j^{(1)}(U) - \phi_j^{(2)}(U) \right]^2, \quad j=1, \dots, k$$

and $\|\cdot\|$ denotes the Euclidean norm. Here we assume that each unspecified function $\phi_0, f(t), j = 1, \dots, k$, is smooth in the sense that it has a bounded p -th derivative in $[a, b]$ for $p \geq 1$ and that the first derivative of $\phi_0, f(t)$ is strictly positive and continuous on $[a, b]$, along with some other technical regularity conditions that are listed in the Appendix. Our regularity conditions are similar to those used in [8]. Given assumptions A1, A2 and regularity conditions C1–C6, it can be shown that

$$d(\hat{\theta}_n, \theta_0) \xrightarrow{p} 0$$

if the number of internal knots $N_j = O(n^\nu)$ for all $j = 1, \dots, k$, where ν satisfies $1/[2(1+p)] < \nu < 1/(2p)$ (more details on the selection of the number of internal knots are presented in subsection 2.1). Therefore, the B-spline-based sieve maximum likelihood estimator is consistent. Additionally,

$$d(\hat{\theta}_n, \theta_0) = O_p \left\{ n^{-\min[p\nu, (1-\nu)/2]} \right\}.$$

This implies that the convergence rate of the estimator for the functional parameters is slower than the usual \sqrt{n} rate. This estimator achieves the optimal convergence rate for nonparametric regression estimators which is $n^{-p/(1+2p)}$ when one chooses $\nu = 1/(1+2p)$ [8]. Finally,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N[0, I^{-1}(\beta_0)],$$

which implies that the convergence rate of the estimator for the Euclidean parameter is \sqrt{n} . This also points to the efficiency of this estimator as the corresponding variance matrix attains the semiparametric efficiency bound $I(\beta_0)$. The definition of the information matrix $I(\beta_0)$ is provided in the Appendix. Since finding $I(\beta_0)$ involves solving an integral equation with no explicit solution, estimation of $I(\beta_0)$ by $I(\hat{\beta}_n)$ is not straightforward [8, 10]. Consequently, one could either use the approach by Zhang et al. [8] and Li [10] for standard error estimation or rely on the computationally simpler nonparametric bootstrap method. The validity of the bootstrap for the Euclidean parameter estimates in general semiparametric M-estimation problems has been verified by Cheng and Huang [20]. The above mentioned asymptotic properties are proved in the Appendix.

2.1. Practical implementation of the method

The proposed methodology can be easily implemented using our R function `ciregic`. An example on how to use this function is provided in the Appendix. In this subsection we discuss the three steps that are crucial for applying our methodology in practice. The first and third steps are automatically implemented using our `ciregic` function.

The first step to apply the methodology is to consider the number and the placement of the knots for the B-splines. As it was mentioned previously, to achieve the optimal nonparametric convergence rate for the functional parameters, one has to select the knots based on the smoothness of the underlying unspecified parameters $\phi_0, f(t), j = 1, \dots, k$, as measured by the order p of the derivatives of these parameters that are bounded. However, this degree of smoothness will be unknown in practice. In this case, one can impose the weakest possible assumption that only the first derivative is bounded, i.e. $p = 1$, and proceed with setting $\nu = 1/(1+2p) = 1/3$. Under this choice, the number of internal knots N_j has the

order of $n^{1/3}$ and, therefore, N_j can be set equal to $\lceil n^{1/3} \rceil$ which corresponds to the largest integer that is smaller than or equal to $n^{1/3}$. The price, however, for imposing the weakest possible assumption regarding the smoothness of the functional parameters is that the convergence rate of the estimator for the baseline cause-specific cumulative incidence function is $n^{1/3}$. This is the slowest achievable rate for this estimator using the B-spline-based sieve maximum likelihood approach. Based on our simulation study presented in the next Section, the performance of this approach is quite satisfactory even with small sample sizes. It has to be noted that this minimal assumption (i.e., $p = 1$) has been imposed in our R function `ciregic`. The placement of knots for the B-splines is an easier task since, due to the monotonicity constraints, these splines have better behavior compared to unconstrained B-splines and are less sensitive to the placement of the knots. Consequently, placing the knots at the quantiles of the empirical distribution of the observation times provides a reasonable and simple approach to knot placement.

Next, selecting the link function parameters is needed in cases where the true link functions are unknown a priori (as it is usually the case in practice). In this case, we propose, following [18], performing a grid search over a plausible combination of $\alpha_1, \dots, \alpha_k$ values and selecting the combination with the largest maximized likelihood value. We must acknowledge however, that the additional uncertainty due to this type of model selection procedure is very difficult to estimate and this is still an open problem in the literature [19, 18]. For this reason we assumed in all theoretical derivations that the true link functions are a priori known.

The final step is to obtain the parameter estimates of the model by maximizing the likelihood function (1) through a constrained optimization algorithm. The monotonicity constraints are linear inequality constraints and they can be easily incorporated in the optimization by using for example the logarithmic barrier method. However, incorporating the boundedness constraint (3) is more challenging as this is a nonlinear inequality constraint. For the special case with $k = 2$ and $\alpha_1 = \alpha_2 = 1$, i.e. with two competing risks and logit link functions for both risks, the boundedness constraint is equivalent to a linear inequality constraint. In this special case the optimization can be performed using a constrained optimization package such as the R package `constrOptim`, which can handle linear inequality constraints through the logarithmic barrier method. If there are more than 2 competing risks or one needs to fit a wide variety of models under different combinations of α values, then an optimization algorithm that incorporates nonlinear inequality constraints is needed. In this case, the R package `alabama` provides a useful set of functions for optimization under both linear and nonlinear inequality constraints. Our `ciregic` function automatically imposes the monotonicity and boundedness constraints using the `alabama` package.

3. Simulation study

In order to evaluate the finite sample performance of our method, we performed a series of simulation experiments. We considered two causes of failure and two covariates, one binary (Z_1) and one continuous (Z_2), in the simulations. The covariates $Z = (Z_1, Z_2)'$ were generated from the Bernoulli distribution with probability of success 0.4 and from the

standard normal distribution, respectively. The cumulative incidence for causes 1 and 2 were considered to have a proportional odds form:

$$F_j(t) = \frac{\exp[\phi_j(t) + \beta_j'Z]}{1 + \exp[\phi_j(t) + \beta_j'Z]}, \quad j=1, 2$$

where $\exp[\phi_1(t)] = 0.4 [1 - \exp(-0.6t)]/0.6$ and $\exp[\phi_2(t)] = 0.75 [1 - \exp(-0.5t)]/0.5$ following a Gompertz distribution [3]. The true values for the regression parameters were $\beta_1 = (\beta_{11}, \beta_{12})' = (0.5, -0.3)$ and $\beta_2 = (\beta_{21}, \beta_{22})' = (-0.5, 0.3)$.

Based on these models, we simulated the failure times and causes of failure. The first observation time U_{1i} (e.g. clinic visit) was simulated based on the exponential distribution with a hazard parameter equal to 3. Each one of the next observations were placed after time V from the previous observation, with V being generated based on the exponential distribution with a hazard parameter equal to 3, for a maximum period of 3 years. This choice led to an average time of four months between two consecutive observations. The number of simulated data sets was 1,000, with the sample size per data set ranging between 200, 350 and 500. The proposed methodology was applied in each data set using cubic B-splines, and the number of knots was set equal to the larger integer up to and including $N^{1/3}$, with N being the total number of distinct time points V_i and U_i for the non-right-censored observations, plus the number of right-censored observations. The knots were placed at the corresponding percentiles of the distribution of the observed times (V_i, U_i) . The nonparametric bootstrap method was used for standard error estimation of the parameter of interest β , based on 50 bootstrap samples. More specifically, for each simulated dataset, 50 random samples of size n were drawn with replacement from the dataset, and then the model was fitted to each of these samples. The standard deviation of the corresponding 50 regression parameter estimates was the estimated standard error of $\hat{\beta}_n$.

Simulation results regarding the baseline cumulative incidence functions are presented in Figure 1. Even with a sample size of 200, the average of the estimated baseline cumulative incidence functions were in close agreement with the corresponding true quantities. Regarding the regression parameter (Table 1), the method performs well, with the bias being very small and the empirical coverage probability staying close to the nominal 95% level in all cases.

4. Analysis of the HIV data

The methodology is applied to the evaluation of risk factors for the cause-specific cumulative incidence of being lost to care and of death while in care. Loss to care is a measure of patient non-retention in care, and thus is an extremely important measure for the success of any intervention program for the HIV/AIDS epidemic, since patients who have disengaged from HIV care do not receive antiretroviral treatment (ART) and thus have increased HIV-related mortality and morbidity, are more infectious and contribute to further expanding the HIV epidemic. The data came from the East Africa IeDEA (International Epidemiologic Databases to Evaluate AIDS) Regional Consortium, a network of HIV care

and treatment programs in Kenya, Uganda and Tanzania. The working definition of loss to care was having no clinic visits for a 3-month period. This cutoff was chosen by the clinicians because typically patients on ART are expected to have run out of medication if they have not had a clinic visit within a 3-month period. A schema of typical patient temporal trajectories is presented in Figure 2. Based on this figure it is clear that neither loss to care nor death times are exactly known due to interval censoring. In the current analysis we included patients with a CD4 count of at least 100 cells/ μ l at ART initiation. Basic descriptive characteristics of the study sample are presented in Table 2.

Similarly to the simulation study, we applied the proposed methodology using cubic B-splines and with the number of knots being equal to the larger integer up to and including $N^{1/3}$, with N being the total number of distinct time points V_j and U_j for the non-right-censored patients plus the number of right-censored patients. The knots were placed at the corresponding percentiles of the distribution of the observed times (V_j, U_j). The nonparametric bootstrap method was used for standard error estimation of the parameter of interest, based on 50 bootstrap samples, as described in the simulation study Section. The analysis was performed using our R function `ciregic` that is illustrated in the Appendix.

In order to perform the analysis we needed to select the $\alpha = (\alpha_1, \alpha_2)$ parameter indexing the link functions for loss to care and death while in care. To do so we performed a grid search over all possible combinations of $\alpha_1 \in \{0, 0.2, \dots, 10\}$ and $\alpha_2 \in \{0, 0.2, \dots, 10\}$. The maximum log-likelihood value (-3144.47) was achieved at the point $\alpha = (2, 1.2)$. The maximized log-likelihood value for the popular Fine-Gray proportional subdistribution hazards model for both competing risks (FGFG), i.e., $\alpha = (0, 0)$, was -3147.05 . Consequently, the fit of the FGFG model is clearly not optimal. In order to illustrate our method for the Fine-Gray model and the proportional odds model, which are the most popular choices in practice, we also fitted a model which assumes a proportional subdistribution hazards model for loss to care and a proportional odds model for death (FGPO model). In this case α was set equal to $(0, 1)$. The maximized log-likelihood for the FGPO model was -3146.91 . The results of the analysis based on the FGPO and the optimal model are presented in Table 3.

Based on Table 3, lower age at ART initiation and male gender are associated with an increased cumulative incidence of loss to care. Specifically, the estimate associated with age implies that patients older by 10 years have 21.7% lower subdistribution hazard [11] of being lost to care (p -value <0.001). In addition, males have 17.0% higher subdistribution hazard of being lost to care compared to females (p -value $=0.031$). Also, there is no evidence for an association between CD4 count at ART initiation and the cumulative incidence of loss to care (p -value $=0.947$). Older age, male gender and lower CD4 count at ART initiation are strongly associated with an increased odds of death while in care. It is clear that the interpretability of the estimated regression parameter for death while in care as odds ratios is very desirable, especially for non-statistically inclined investigators. We also compared the results from the FGPO model with those from the optimal model based on the grid search with $\alpha = (2, 1.2)$. The estimated regression coefficients are presented in Table 3 as well. The results from the two analyses are qualitatively similar. However, the regression coefficient estimates for gender and age differ substantially between the two models for the loss to care

outcome. As a comparison, we also performed a naïve analysis based on the midpoint imputation approach, under the optimal model with $\alpha = (2, 1.2)$. Using this approach, the unknown failure times are replaced by the midpoint between the two observation times that surround the true failure time, and then the data are analyzed as usual right-censored data. The estimate for the effect of CD4 on loss to care, based on the naïve analysis, was 10.9 times larger (0.011 versus 0.001) compared to the estimate from the proposed approach, although this effect was still not statistically significant. Additionally, the estimate of the effect of age on death was attenuated by 8.3% based on the naïve analysis compared to the estimate from the proposed method. However, this effect was still statistically significant. The difference between the rest estimates from the two approaches was less than 3.4%.

It is straightforward to derive predicted cause-specific cumulative incidence functions for specific patient subgroups. Figure 3 presents results of these quantities, based on the estimates from the optimal model, for a female patient with CD4 count at ART initiation of 120 cells/ μL , by age at ART initiation. Such predictions are useful for evaluating program effectiveness and can also be used as inputs to mathematical models of cost-effectiveness analyses to identify the best strategy to optimize care in our setting. The comparison between the predictions from the optimal model and the FGPO model are presented in the bottom row of Figure 3. For the loss to care outcome, the predicted cumulative incidences for a 40-year old male patient with a CD4 count at ART initiation of 120 cells/ μL from the two models differ substantially. More specifically, the difference between the two age groups is smaller after the first year based on the optimal model compared to the FGPO model. This provides some evidence for the violation of the proportional subdistribution hazards assumption of the Fine-Gray model. However, when considering the death outcome, the predicted cumulative incidences from the two models are quite close, and this is because the difference between the link function parameters in the two models is somewhat small (i.e., $\alpha_2 = 1.2$ for the optimal model and $\alpha_2 = 1$ for the FGPO model).

5. Discussion

In this paper we extended the work by Li [10] on semiparametric analysis of the proportional subdistribution hazards or the Fine-Gray model [11] for cumulative incidence estimation with interval-censored competing risk data, to the class of the semiparametric generalized odds rate transformation models. This general class includes the proportional odds model [3, 6] and the proportional subdistribution hazards model as special cases. This extension is very important for medical and epidemiological applications when the proportional subdistribution hazards model does not provide good fit to the data, as was the case with our HIV data analysis. Unlike [10], we also explicitly incorporate the boundedness of the cause-specific cumulative incidence function constraint into the optimization, and we also provide and illustrate an R function called `ciregic`, that easily implements the proposed method.

We have showed that the B-spline sieve maximum likelihood estimator for the regression (Euclidean) parameter, under the class of the generalized odd rate transformation models is consistent, semiparametrically efficient and asymptotically normal. It was also shown that the estimator for the functional parameter (baseline cumulative incidence functions) is

consistent in an L_2 -metric and converges at the optimal rate for nonparametric regression which is nevertheless slower than the usual parametric \sqrt{n} rate. This methodology is also applicable to the case of current status data [21, 22] since these data are special cases of interval-censored data [8]. It is straightforward to show that our results also hold for the more general class of semiparametric transformation models.

Throughout this paper we assumed that the true link functions are known as in [16, 17, 12, 18]. A practical issue regarding the application of the proposed method in practice is when this assumption is not true. To this end, and similarly to Mao and Wang [18], we proposed performing a grid search over a plausible set of combinations of $(\alpha_1, \dots, \alpha_k)$ and choosing the combination with the largest maximized likelihood value. This approach for selecting the proper link function parameter values was illustrated in an analysis of data from a large network of HIV care and treatment programs in sub-Saharan Africa. The issue of the additional variability due to this type of model selection is still an open problem and requires further investigation.

Our approach simultaneously estimates the models for all causes of failure in contrast to many semiparametric methods for right-censored competing risk data. Analysis of all the competing causes of failure is essential for a more complete understanding of the underlying competing risks process. However, this comes at the cost of specifying a model for every cause of failure and can increase the risk of model misspecification, even though our method is very flexible in terms of the baseline cumulative incidence functions and the link functions.

In general, fully parametric methods can be straightforwardly extended to the case of interval censoring, and have the advantage of being efficient. However, the validity of these methods relies on strong assumptions for each cause of failure and this limits their generality. One solution is to adopt more flexible 3-parameter distributions as in [5] and [6]. However, while these distributions are clearly less flexible than the B-spline approximations used in this paper, they may also be plagued by non-convergence issues such as in univariate survival analysis [7]. By contrast, when using our proposed methods, we did not face non-convergence problems in either in the simulations or the analysis of the HIV data. A disadvantage of our method, compared to parametric approaches, is that it does not provide estimates of the long term proportions for each cause of failure. Even though this is possible with parametric approaches as well [23, 3, 6, 2], care is needed when making extrapolations beyond the observable time interval as mentioned by Jeong and Fine [3].

In this paper, we did not consider the asymptotic distribution of the estimated baseline and covariate-specific cumulative incidence functions as the slower rate of convergence of this estimator make this task considerably difficult. However, evaluating the uncertainty of model-based predictions of the cumulative incidence is very useful from a clinical and implementation science perspective. Therefore, it would be useful to study the asymptotic properties of the estimated baseline and covariate-specific cumulative incidence functions and propose a method for the calculation of confidence intervals and simultaneous confidence bands. Additionally, the issue of quantifying the uncertainty due to the link function selection is another interesting and challenging issue for future research.

Acknowledgments

The authors would like to thank the Associate Editor and the two anonymous reviewers for their constructive feedback. Research reported in this publication was supported by the National Institute Of Allergy And Infectious Diseases (NIAID), Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), National Institute On Drug Abuse (NIDA), National Cancer Institute (NCI), and the National Institute of Mental Health (NIMH), in accordance with the regulatory requirements of the National Institutes of Health under Award Number U01AI069911 East Africa IeDEA Consortium. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research has also been supported by the National Institutes of Health grant R01-AI102710 “Statistical Designs and Methods for Double-Sampling for HIV/AIDS” and by the President’s Emergency Plan for AIDS Relief (PEPFAR) through USAID under the terms of Cooperative Agreement No. AID-623-A-12-0001 it is made possible through joint support of the United States Agency for International Development (USAID). The contents of this journal article are the sole responsibility of AMPATH and do not necessarily reflect the views of USAID or the United States Government.

References

- Hudgens MG, Satten GA, Longini IM. Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics*. 2001; 57:74–80. [PubMed: 11252621]
- Hudgens MG, Li C, Fine JP. Parametric likelihood inference for interval censored competing risk data. *Biometrics*. 2014; 70:1–9. [PubMed: 24400873]
- Jeong JH, Fine JP. Parametric regression on cumulative incidence function. *Biostatistics*. 2007; 8:184–196. [PubMed: 16636138]
- Dabrowska DM, Doksum KA. Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*. 1988; 83:744–749.
- Cheng Y. Modeling cumulative incidences of dementia and dementia-free death using a novel three-parameter logistic function. *The International Journal of Biostatistics*. 2009; 5:1–17.
- Shi H, Cheng Y, Jeong JH. Constrained parametric model for simultaneous inference of two cumulative incidence functions. *Biometrical Journal*. 2013; 55:82–96. [PubMed: 23090878]
- Lambert PC, Dickman PW, Weston CL, Thompson JR. Estimating the cure fraction in population-based cancer studies by using finite mixture models. *Journal of the Royal Statistical Society: Series C Applied Statistics*. 2010; 59:35–55.
- Zhang Y, Hua L, Huang J. A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scandinavian Journal of Statistics*. 2010; 37:338–354.
- Shen X, Wong WH. Convergence rate of sieve estimates. *The Annals of Statistics*. 1994; 22:580–615.
- Li C. The Fine-Gray model under interval censored competing risk data. *Journal of Multivariate Analysis*. 2016; 143:327–344. [PubMed: 26543275]
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999; 94:496–509.
- Fine JP. Regression modeling of competing crude failure probabilities. *Biostatistics*. 2001; 2:85–97. [PubMed: 12933558]
- Eriksson F, Li J, Scheike T, Zhang M-J. The proportional odds cumulative incidence model for competing risks. *Biometrics*. 2015; 71:687–695. [PubMed: 26013050]
- Choi S, Huang X. Maximum likelihood estimation of semiparametric mixture component models for competing risk data. *Biometrics*. 2014; 70:588–598. [PubMed: 24734912]
- Mao L, Lin DY. Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks. *Journal of the Royal Statistical Society: Series B*. 2017; 79:573–587.
- Scharfstein DO, Tsiatis AA, Gilbert PB. Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis*. 1998; 4:355–391. [PubMed: 9880995]
- Fine JP. Analysing competing risk data with transformation models. *Journal of the Royal Statistical Society B: Statistical Methodology*. 1999; 61:817–830.

18. Mao MM, Wang J-L. Semiparametric efficient estimation for a class of generalized proportional odds cure models. *Journal of the American Statistical Association*. 2010; 105:302–311.
19. Zeng D, Yin G, Ibrahim JG. Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*. 2006; 101:670–684.
20. Cheng G, Huang JZ. Bootstrap consistency for general semiparametric M-estimation. *The Annals of Statistics*. 2010; 38:2884–2915.
21. Groeneboom P, Maathuis MH, Wellner JA. Current status data with competing risks: Consistency and rates of convergence of the MLE. *The Annals of Statistics* 2008. 2008a; 36:1031–1063.
22. Groeneboom P, Maathuis MH, Wellner JA. Current status data with competing risks: Limiting distribution of the MLE. *The Annals of Statistics*. 2008; 36:1064–1089. [PubMed: 19888358]
23. Jeong JH, Fine JP. Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society C: Applied Statistics*. 2006; 55:187–200.
24. van der Vaart, AW., Wellner, JA. *Weak convergence and empirical processes with applications to statistics*. Springer-Verlag; New York: 1996.
25. Wellner JA, Zhang Y. Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics*. 2007; 35:2106–2142.
26. Lu, M. PhD dissertation. Department of Biostatistics, University of Iowa; 2007. Monotone spline estimations for panel count data.
27. Kosorok, MR. *Introduction to empirical processes and semiparametric inference*. Springer; New York: 2008.
28. van der Vaart, AW. *Asymptotic statistics*. Cambridge University Press; Cambridge: 2000.
29. Huang, J., Wellner, JA. *Proceedings of the First Seattle Symposium in Biostatistics*. Springer; US: 1997. Interval censored survival data: A review of recent progress; p. 123-169.
30. Billingsley, P. *Probability and measure*. Wiley; New York: 1986.

Appendix

The R function `ciregic`

The R function `ciregic`, which can be found in the online material of this manuscript, implements the methods proposed in this work. This function depends on the package `alabama` that performs optimization with both linear and nonlinear inequality constraints. Therefore, prior installation of this package is required in order to use the `ciregic` function. The first step is to install/define the `ciregic` function, which can be done by either running the script `ciregic.R` provided in the online material of this manuscript, or by storing it in the working folder and then using the command

```
source("ciregic.R")
```

In this illustration we use the HIV dataset that was analyzed in this manuscript. We load and list the first records of these data using the commands

```
data<-read.csv("iedea.csv")
head(data)

##   id      v      u c age_art male cd4_art
## 1  1 0.26830938 0.2710472 0 35.67146  0    192
## 2  2 0.31485283 0.3175907 0 45.65366  1    191
## 3  3 0.14784394 0.1724846 2 62.52977  1    102
## 4  4 0.05475701 0.3011636 1 30.77892  0    144
## 5  5 2.44216300 2.4449008 0 43.16496  0    664
## 6  6 0.19164956 0.1943874 0 39.76454  1    122
```

The variables v and u are the lower and upper observation times V and U , respectively. These times correspond to years since ART initiation. The variable c is the cause of failure, where the loss to care event is coded as $c=1$, the death event as $c=2$, and the right-censoring as $c=0$. Note that the u values for the right-censored observations (i.e., $c=0$) are not included in any calculation. This is because in theory, $U = \infty$ for right-censored observations. Therefore, these times have been previously replaced with the lower observation times v plus 1 day as

```
data$u<-(data$v+1/365.25)*(data$c==0)+data$u*(data$c>0)
```

Performing the analysis that was presented in the HIV data example section based on the Fine-Gray model for loss to care and the proportional odds model for death, which correspond to $\alpha = (0, 1)$, requires the commands

```
set.seed(20835)
fit.01<-ciregic(formula=Surv2(v, u, c)~ male + age_art + cd4_art,
  data=data, alpha=c(0,1), nboot=50)

## [1] "Failure cause 1"
##      Estimate Std. Error z value Pr(>|z|)
## male    1.567e-01  7.256e-02   2.159  0.0308 *
## age_art -2.444e-02  4.645e-03  -5.262 1.43e-07 ***
## cd4_art  1.657e-05  2.485e-04   0.067  0.9468
## ---
## Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
## [1] "Failure cause 2"
##      Estimate Std. Error z value Pr(>|z|)
## male    0.563947  0.212187   2.658  0.00787 **
## age_art  0.030785  0.007342   4.193 2.75e-05 ***
## cd4_art -0.003576  0.001216  -2.942  0.00326 **
## ---
## Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

In order to achieve reproducibility of the standard error estimates, as they are based on the bootstrap, we need to manually set a random seed as presented above. The first argument in the `Surv2` function is the lower observation time V , the second is the upper observation time U , and the third is the cause of failure variable. The last variable should include 0 for right-censored observations, and 1 and 2 for causes of failure 1 and 2, respectively. Currently, we allow for only two competing risks, but we plan to extend this to three or more in the near future. The output presented above corresponds to the first column of Table 3. Note that in that table, the effects of age and CD4 count have been multiplied by 10 and 100, respectively, for interpretation purposes. The object `fit.01` is a list that contains an output of the analysis in the first component, the regression parameter estimates and the corresponding variance-covariance matrix in the second and third components respectively, and the value of the maximized log-likelihood in the eighth component. Indication of successful convergence of the optimization algorithm is stored in the ninth component of `fit.01`. The components 4 to 7, include the estimated baseline cumulative incidence functions and some other elements that are useful for estimating the covariate-specific cumulative incidence function using our `pred.cif` function. The estimation with 50 bootstrap replications and a dataset of 3,053 observations required 11.65 minutes. We plan to optimize the computational efficiency of our function by utilizing multiple cores, and to upload it in CRAN as a package. Estimation of the optimal model for this analysis with $\alpha = (2, 1.2)$ can be performed using the command

```
set.seed(20835)
fit.opt<-ciregic(formula=Surv2(v, u, c)~ male + age_art + cd4_art,
                data=data, alpha=c(2,1.2), nboot=50)
```

Incorporation of an interaction effect between gender and age at ART initiation could be simply incorporated as

```
fit.opt<-ciregic(formula=Surv2(v, u, c)~ male*age_art + cd4_art,
                data=data, alpha=c(2,1.2), nboot=50)
```

Estimation of the covariate-specific cumulative incidence function can be easily performed using the `pred.cif` function. The first argument in this function is the object containing the results from the model fit generated by the `ciregic` function, and the second the desired covariate pattern as a vector. For example, in order to estimate the cumulative incidence function of loss to care and death for a 20-year old female with a CD4 count of 120 cells/ μ l at ART initiation, we need the command

```
f20.01<-pred.cif(fit.01, covp=c(0,20,120))
head(f20.01)
##           t           cif1           cif2
## 1 0.00000000 4.996004e-15 0.001595043
## 2 0.05265523 7.882583e-15 0.007162271
## 3 0.10531046 5.606626e-14 0.012532213
## 4 0.15796569 1.477572e-07 0.015031502
## 5 0.21062092 4.201278e-02 0.017972113
## 6 0.26327615 1.896563e-01 0.019366670
```

The option `covp` should include the desired covariate values, in the same order as the listing order of the corresponding covariates in the right side of the formula argument in `ciregic` function. As in any other prediction problem, these covariate values should not be outside the observable range of covariate values in order to avoid extrapolation. The function `pred.cif` returns a data frame that includes times and the estimated cumulative incidence functions for both failure causes and for the desired covariate pattern. Plotting the object `f20.01` in the example above (e.g., with the `plot` function) will provide the estimated cumulative incidences that are depicted in the second row of Figure 3 as solid grey lines.

Proof of consistency

Empirical process theory techniques will be applied to derive the consistency of the B-spline sieve maximum likelihood estimator. As usual $PF = \int_{\mathcal{X}} f(x) dP(x)$ and

$\mathbb{P}_n f = n^{-1} \sum_{i=1}^n f(X_i)$, which is the empirical process indexed by the function f evaluated at $X_i = (V_i, U_i, \delta_{i1}, \dots, \delta_{ik}, \delta_{i1}^1, \dots, \delta_{ik}^1, \delta_i, Z_i)$, the sample point of individual i , with \mathcal{X}

denoting the sample space. Also, let C denote a general constant which may differ in different settings. The following regularity conditions are assumed

- C1** $E(ZZ')$ is non-singular and Z is bounded, i.e. there exists a $z_0 > 0$ such that $P(\|Z\| \geq z_0) = 1$.
- C2** $\beta_j \in \mathcal{B}_j$, where \mathcal{B}_j is a compact subset of \mathbb{R}^d for every $j = 1, \dots, k$.
- C3** There exists an $\eta > 0$ such that $P(U - V \leq \eta) = 1$ and the union of the supports of V and U are contained in $[a, b]$, where $0 < a < b < \infty$ and $0 < \min_{j \in \{1, \dots, k\}} F_j(a) < \sum_{j=1}^k F_j(b) < 1$.
- C4** $\phi_{0,j} \in \Phi$, where Φ is a class of functions with bounded p -th derivative in $[a, b]$ for $p \geq 1$ and the first derivative of $\phi_{0,j}$ is strictly positive and continuous on $[a, b]$. This holds for all $j = 1, \dots, k$.
- C5** The conditional density of (V, U) given Z has bounded partial derivatives with respect to (v, u) and the bounds of these derivatives do not depend on (v, u, z) .
- C6** For some $\kappa \in (0, 1)$, $a \text{Var}(Z|V)a \leq \kappa a \text{E}(ZZ'|V)a$ a.s. and $a \text{Var}(Z|U)a \leq \kappa a \text{E}(ZZ'|U)a$ a.s. for all $a \in \mathbb{R}^d$.

Let $\mathcal{A}_1 = \{l(\theta; X): \theta \in \Theta_n\}$ denote the class of log-likelihood functions indexed by the parameter space $\Theta_n = \times_{j=1}^k (\mathcal{B}_j \times \mathcal{M}_{n,j})$. The Euclidean parameters are $\beta_j \in \mathcal{B}_j$ and the functional parameters are $\phi_j \in \mathcal{M}_{n,j}$ for $j = 1, \dots, k$. The functional parameter space $\mathcal{M}_{n,j} \equiv \mathcal{M}_n(\gamma_j, N_{n,j}, p_j)$ is the space of monotone B-spline functions defined on the observation time interval $[a, b]$. Convergence will be proved in the L_2 -metric

$$d(\theta_1, \theta_2) = \left(\sum_{j=1}^k \|\beta_j^{(1)} - \beta_j^{(2)}\|^2 + \sum_{j=1}^k \|\phi_j^{(1)} - \phi_j^{(2)}\|_{\Phi}^2 \right)^{\frac{1}{2}},$$

for $\theta_1 = (\beta^{(1)'}, \phi^{(1)'})'$ and $\theta_2 = (\beta^{(2)'}, \phi^{(2)'})'$, where

$$\|\phi_j^{(1)' } - \phi_j^{(2)' } \|_{\Phi}^2 = E \left[\phi_j^{(1)' } (V) - \phi_j^{(2)' } (V) \right]^2 + E \left[\phi_j^{(1)' } (U) - \phi_j^{(2)' } (U) \right]^2, \quad j=1, \dots, k.$$

Let $\mathbb{M}(\theta) = P l(\theta; X)$ and $\mathbb{M}_n(\theta) = P_n l(\theta; X)$, which leads to $\mathbb{M}_n(\theta) - \mathbb{M}(\theta) = (P_n - P) l(\theta; X)$ for each $\theta \in \Theta_n$. In order to prove that $d(\hat{\theta}_n, \theta_0) \xrightarrow{P} 0$ we need to verify the following conditions:

1. $\sup_{\theta \in \Theta_n} |\mathbb{M}_n(\theta) - \mathbb{M}(\theta)| \xrightarrow{P} 0$
2. $\sup_{\theta: d(\theta, \theta_0) \leq \varepsilon} \mathbb{M}(\theta) < \mathbb{M}(\theta_0)$
3. The sequence of estimators $\hat{\theta}_n$ satisfy $\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) = o_p(1)$

It has been shown that there exists a set of brackets $\{[\phi_{i,j}^L, \phi_{i,j}^U]: i = 1, \dots, (1/\epsilon)^{Cq_n,j}\}$, where C is a constant and $q_{n,j} = N_{n,j} + p_j + 1$, with $N_{n,j}$ and p_j being the number of internal knots and the degree of the B-spline for the j -th cause of failure, such that $\mathbb{P}_n|\phi_{i,j}^U - \phi_{i,j}^L| \leq \epsilon$ and $\phi_{i,j}^L(t) \leq \phi_j(t) \leq \phi_{i,j}^U(t)$ for any $t \in [a, b]$ and $i = 1, \dots, (1/\epsilon)^{Cq_n,j}$ [9]. By condition C2, $\mathbb{B}_j \subset \mathbb{R}^d$ is compact and thus it can be covered by $C(1/\epsilon)^d$ ϵ -balls, i.e. for any $\beta \in \mathbb{B}_j$ there exists an $s = 1, \dots, C(1/\epsilon)^d$ such that $\beta \in B(\beta_{s,j}) \equiv \{b \in \mathbb{B}_j: |b - \beta_{s,j}| \leq \epsilon\}$. By condition C1 it follows that for any $\beta \in \mathbb{B}_j$ and any $z, \beta'z \in [\beta'_{s,j}z - C\epsilon, \beta'_{s,j}z + C\epsilon]$. Let

$\eta_j(t, z) \equiv \phi_j(t) + \beta'z$. Then, for any $j = 1, \dots, k$ we can construct a set of brackets $\{[\eta_{j,s,i}^L, \eta_{j,s,i}^U]: i = 1, \dots, (1/\epsilon)^{Cq_n,j}; s = 1, \dots, C(1/\epsilon)^d\}$, with

$\eta_{j,s,i}^L(t, z) = \phi_{i,j}^L(t) + \beta'_{s,j}z - C\epsilon$ and $\eta_{j,s,i}^U(t, z) = \phi_{i,j}^U(t) + \beta'_{s,j}z + C\epsilon$, such that

$\mathbb{P}_n|\eta_{j,s,i}^U - \eta_{j,s,i}^L| \leq C\epsilon$ and $\eta_j(t, z) \in [\eta_{j,s,i}^L(t, z), \eta_{j,s,i}^U(t, z)]$ for any $t \in [a, b]$ and $z \in Z$. So, for any cause of failure $j = 1, \dots, k$ it is straightforward to construct a set of brackets

$\{[F_{j,s,i}^L, F_{j,s,i}^U]: i = 1, \dots, (1/\epsilon)^{Cq_n,j}; s = 1, \dots, C(1/\epsilon)^d\}$ with:

$$F_{j,s,i}^L = g_j^{-1}(\eta_{j,s,i}^L; \alpha_j)$$

and

$$F_{j,s,i}^U = g_j^{-1}(\eta_{j,s,i}^U; \alpha_j)$$

such that there exists an $s = 1, \dots, C(1/\epsilon)^d$ and an $i = 1, \dots, (1/\epsilon)^{Cq_n,j}$ such that

$F_j(t, z) \in [F_{j,s,i}^L(t, z), F_{j,s,i}^U(t, z)]$ for any $t \in [a, b]$ and $z \in Z$. Using Taylor approximation

along with conditions C1, C2 the constraint $\sum_{j=1}^k F_j(b; z, \theta_j) < 1$ and the fact that $a_j < \infty$ for all $j = 1, \dots, k$, we can show that $\mathbb{P}_n|F_{j,s,i}^U - F_{j,s,i}^L| \leq C\epsilon$ and the bracketing number with the $L_1(\mathbb{P}_n)$ -norm is bounded by $C(1/\epsilon)^{d+Cq_n,j}$, for any $j = 1, \dots, k, s = 1, \dots, C(1/\epsilon)^d$ and $i = 1, \dots, (1/\epsilon)^{Cq_n,j}$. Consequently, we can construct a set of brackets

$\{[l_{i_1, \dots, i_k, s_1, \dots, s_k}^L, l_{i_1, \dots, i_k, s_1, \dots, s_k}^U]: i_j = 1, \dots, (1/\epsilon)^{Cq_n,j}, s_j = 1, \dots, C_j(1/\epsilon)^{d_j}\}$, where k is the fixed and finite number of competing risks, with

$$\begin{aligned} l_{i_1, \dots, i_k, s_1, \dots, s_k}^L(X) &= \sum_{j=1}^k \delta_j^L \log [F_{j,s_j,i_j}^L(V, Z)] + \sum_{j=1}^k \delta_j \log [F_{j,s_j,i_j}^L(U, Z) - F_{j,s_j,i_j}^U(V, Z)] \\ &+ (1 - \delta) \log \left[1 - \sum_{j=1}^k F_{j,s_j,i_j}^U(U, Z) \right], \end{aligned}$$

where $\delta_j^L = I_{\{X \leq V, \delta=j\}}$, $\delta_j = I_{\{V < X \leq U, \delta=j\}}$, $\delta = \sum_{j=1}^k (\delta_j + \delta_j^1)$ and

$$l_{i_1, \dots, i_k, s_1, \dots, s_k}^U(X) = \sum_{j=1}^k \delta_j^U \log [F_{j, s_j, i_j}^U(V, Z)] + \sum_{j=1}^k \delta_j \log [F_{j, s_j, i_j}^U(U, Z) - F_{j, s_j, i_j}^L(V, Z)] + (1 - \delta) \log \left[1 - \sum_{j=1}^k F_{j, s_j, i_j}^L(U, Z) \right]$$

such that for any $j = 1, \dots, k$ there exists an $i_j = 1, \dots, (1/\varepsilon)^{Cq_{n,j}}$, $s_j = 1, \dots, C_j(1/\varepsilon)^\nu$ so that $l(\theta; X) \in [l_{i_1, \dots, i_k, s_1, \dots, s_k}^L(X), l_{i_1, \dots, i_k, s_1, \dots, s_k}^U(X)]$ for any sample point $X \in \mathcal{X}$. Using Taylor approximation and regularity conditions C1–C3 it can be shown that

$\mathbb{P}_n[l_{i_1, \dots, i_k, s_1, \dots, s_k}^U - l_{i_1, \dots, i_k, s_1, \dots, s_k}^L] \leq C\varepsilon$, for any j, i_j and s_j . Additionally, the bracketing number $N_{[]}(\mathcal{C}\varepsilon, \mathcal{F}_1, L_1(\mathbb{P}_n))$ is bounded by $\prod_{j=1}^k C_j(1/\varepsilon)^{\nu + Cq_{n,j}}$. Since $q_{n,j} = n^r + m$, with $r_j \in (0, 1)$, it follows that $\log N_{[]}(\mathcal{C}\varepsilon, \mathcal{F}_1, L_1(\mathbb{P}_n)) = o(n)$, for any $\varepsilon > 0$. Also, due to condition C3 there exists an envelope F for the class of likelihood functions \mathcal{F}_1 such that $P^*F < \infty$ and $\log N(\varepsilon, \mathcal{F}_1, L_1(\mathbb{P}_n)) - \log N_{[]}(\varepsilon, \mathcal{F}_1, L_1(\mathbb{P}_n)) \rightarrow 0$ [8]. Consequently, \mathcal{F}_1 is Glivenko-Cantelli by Theorem 2.4.3 of [24]. Consequently

$$\sup_{\theta \in \Theta_n} |\mathbb{M}_n(\theta) - (\theta)| \xrightarrow{P} 0$$

Next we are going to show that condition 2., i.e. $\sup_{\theta: d(\theta, \theta_0) \leq \varepsilon} \mathbb{M}(\theta) < \mathbb{M}(\theta_0)$, holds. By definition we have:

$$\begin{aligned} (\theta_0) - (\theta) &= E \left\{ \sum_{j=1}^k F_j^0(V, Z) \log \frac{F_j^0(V, Z)}{F_j(V, Z)} + \sum_{j=1}^k [F_j^0(U, Z) - F_j^0(V, Z)] \log \frac{F_j^0(U, Z) - F_j^0(V, Z)}{F_j(U, Z) - F_j(V, Z)} + \left[1 - \sum_{j=1}^k F_j^0(U, Z) \right] \log \frac{1 - \sum_{j=1}^k F_j^0(U, Z)}{1 - \sum_{j=1}^k F_j(U, Z)} \right\} \\ &= E \left\{ \sum_{j=1}^k F_j(V, Z) m \left[\frac{F_j^0(V, Z)}{F_j(V, Z)} \right] + \sum_{j=1}^k [F_j(U, Z) - F_j(V, Z)] m \left[\frac{F_j^0(U, Z) - F_j^0(V, Z)}{F_j(U, Z) - F_j(V, Z)} \right] + \left[1 - \sum_{j=1}^k F_j(U, Z) \right] m \left[\frac{1 - \sum_{j=1}^k F_j^0(U, Z)}{1 - \sum_{j=1}^k F_j(U, Z)} \right] \right\}, \end{aligned}$$

where, as in [8], $m(x) = x \log x - x + 1 - (1/4)(x - 1)^2$ for $x \in [0, 5]$, and thus for any θ in a sufficiently small neighbourhood of θ_0 we have

$$\begin{aligned}
 (\theta_0) - (\theta) &\geq \frac{1}{4} E \left\{ \sum_{j=1}^k F_j(V, Z) \left[\frac{F_j^0(V, Z)}{F_j(V, Z)} - 1 \right]^2 \right. \\
 &\quad + \sum_{j=1}^k [F_j(U, Z) - F_j(V, Z)] \left[\frac{F_j^0(U, Z) - F_j^0(V, Z)}{F_j(U, Z) - F_j(V, Z)} - 1 \right]^2 \\
 &\quad \left. + \left[1 - \sum_{j=1}^k F_j(U, Z) \right] \left[\frac{1 - \sum_{j=1}^k F_j^0(U, Z)}{1 - \sum_{j=1}^k F_j(U, Z)} - 1 \right]^2 \right\} \\
 &= \frac{1}{4} E \left\{ \sum_{j=1}^k \frac{1}{F_j(V, Z)} [F_j^0(V, Z) - F_j(V, Z)]^2 \right. \\
 &\quad + \sum_{j=1}^k \frac{1}{F_j(U, Z) - F_j(V, Z)} [F_j^0(U, Z) - F_j^0(V, Z) - F_j(U, Z) + F_j(V, Z)]^2 \\
 &\quad \left. + \frac{1}{F_j(U, Z)} \left[\sum_{j=1}^k F_j^0(U, Z) - \sum_{j=1}^k F_j(U, Z) \right]^2 \right\} \\
 &\geq \frac{1}{4} E \left\{ \sum_{j=1}^k [F_j^0(V, Z) - F_j(V, Z)]^2 \right. \\
 &\quad + \sum_{j=1}^k [F_j^0(U, Z) - F_j^0(V, Z) - F_j(U, Z) + F_j(V, Z)]^2 \\
 &\quad \left. + \left[\sum_{j=1}^k F_j^0(U, Z) - \sum_{j=1}^k F_j(U, Z) \right]^2 \right\} \\
 &\geq \frac{1}{4} E \left(\sum_{j=1}^k [F_j^0(V, Z) - F_j(V, Z)]^2 + \left\{ \sum_{j=1}^k [F_j^0(U, Z) - F_j(U, Z)] \right\}^2 \right) \\
 &\geq \frac{1}{4} E \left(\left\{ \sum_{j=1}^k [F_j^0(V, Z) - F_j(V, Z)] \right\}^2 + \left\{ \sum_{j=1}^k [F_j^0(U, Z) - F_j(U, Z)] \right\}^2 \right) \\
 &= CE \left(\left\{ \sum_{j=1}^k [(\beta_{0,j} - \beta_j)' Z + (\phi_{0,j} - \phi_j)(V)] \right\}^2 \right. \\
 &\quad \left. + \left\{ \sum_{j=1}^k [(\beta_{0,j} - \beta_j)' Z + (\phi_{0,j} - \phi_j)(U)] \right\}^2 \right)
 \end{aligned}$$

where the last equality follows from a Taylor expansion around the value of the linear predictor $\eta_j(X)$. Using the same arguments as in [25], pp. 2126–2127, along with regularity conditions C1–C6 it follows that

$$(\theta_0) - (\theta) \geq C \left(\sum_{j=1}^k \|\beta_j - \beta_{j,0}\|^2 + \sum_{j=1}^k \|\phi_j - \phi_{j,0}\|_{\Phi}^2 \right) = Cd^2(\theta_0, \theta).$$

Consequently

$$\begin{aligned} \sup_{\theta:d(\theta,\theta_0)\geq\varepsilon} (\theta) &\leq \sup_{\theta:d(\theta,\theta_0)\geq\varepsilon} [(\theta_0)-Cd^2(\theta_0,\theta)] \\ &= (\theta_0)-\inf_{\theta:d(\theta,\theta_0)\geq\varepsilon} Cd^2(\theta_0,\theta) \\ &= (\theta_0)-C\varepsilon^2 \\ &< (\theta_0) \end{aligned}$$

and we have proven condition 2.

Finally, we prove condition 3, i.e. that the sequence of estimators $\hat{\theta}_n$ satisfy $\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) - o_p(1)$. It has been shown that for $\phi_{0,j} \in \Phi$ there exists a $\theta_{0,n,j} \in \mathcal{M}_{n,j}$ of order $m - p + 2$ such that $\|\phi_{0,n,j} - \phi_{0,j}\|_\infty \leq Cq_n^{-p} = O(n^{-p\nu})$ [26]. This also implies that $\|\phi_{0,n,j} - \phi_{0,j}\|_\Phi \leq Cq_n^{-p} = O(n^{-p\nu})$ [8]. Of course the previous result holds for any $j = 1, \dots, k$. Letting $\theta_{0,n} = (\beta_0, \phi_{0,n})$ it follows that

$$\begin{aligned} \mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) &= \mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_{0,n}) + \mathbb{M}_n(\theta_{0,n}) - \mathbb{M}_n(\theta_0) \\ &\geq \mathbb{M}_n(\theta_{0,n}) - \mathbb{M}_n(\theta_0) \\ &= \mathbb{M}_n(\theta_{0,n}) - \mathbb{M}_n(\theta_0) + (\theta_{0,n}) - (\theta_0) \\ &= (\mathbb{P}_n - P)[l(\theta_{0,n}; X) - l(\theta_0; X)] + (\theta_{0,n}) - (\theta_0) \end{aligned} \tag{4}$$

Similar to [8], we consider the class of functions $\mathcal{F}_2 = \{l(\beta_0, \phi; X) - l(\beta_0, \phi_0; X) : \phi_j \in \mathcal{M}_{j,n}, \|\phi_j - \phi_{0,j}\|_\Phi \leq Cn^{-p\nu} \text{ for any } j = 1, \dots, k\}$ and construct a set of ε -brackets with $L_2(P)$ -norm bounded by $(1/\varepsilon) \sum_{j=1}^k C_j q_{n,j}$. Consequently, the corresponding bracketing integral

$$\begin{aligned} J_{[]} (1, \mathcal{F}_2, L_2(P)) &\equiv \int_0^1 \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_2, L_2(P))} d\varepsilon \\ &\leq \sqrt{\sum_{j=1}^k C_j q_{n,j} \int_0^1 \sqrt{\log(1/\varepsilon)} d\varepsilon} \\ &= \sqrt{\sum_{j=1}^k C_j q_{n,j} \int_0^\infty u^{1/2} e^{-u} du} \\ &< \infty \end{aligned}$$

because the integral above is bounded by 1 [27] and thus by theorem 19.5 of [28] \mathcal{F}_2 is P -Donsker. Using the fact that $\|\phi_j - \phi_{0,j}\|_\Phi \leq Cn^{-p\nu}$ in \mathcal{F}_2 and the boundedness of $l(\cdot; X)$, along with the dominated convergence theorem we can show that $\lim_{n \rightarrow \infty} P[l(\beta_0, \phi; X) - l(\beta_0, \phi_0; X)]^2 = 0$. Using the fact that for a P -Donsker class of functions there exists a semimetric ρ_p that makes that class totally bounded and that $\mathbb{G}_n f = \sqrt{n} (\mathbb{P}_n - P) f$ is asymptotically equicontinuous by corollary 2.3.12 of [24] it follows that:

$$(\mathbb{P}_n - P)[l(\theta_{0,n}; X) - l(\theta_0; X)] = o_p(n^{-1/2})$$

Next, for the remaining term in (4) we can again use the boundedness of the loglikelihood function for any sample point $X \in \mathcal{X}$ and that $\|\phi_{0,n,j} - \phi_{0,j}\|_{\Phi} = O(n^{-p\nu})$, along with the dominated convergence theorem it follows that

$$(\theta_{0,n}) - (\theta_0) > -o(1)$$

as $n \rightarrow \infty$. Thus

$$\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq o_p(n^{-1/2}) - o(1) = -o_p(1)$$

and this proves the final consistency condition. Consequently

$$\left(\sum_{j=1}^k \|\hat{\beta}_{j,n} - \beta_{j,0}\|^2 + \sum_{j=1}^k \|\hat{\phi}_{j,n} - \phi_{j,0}\|_{\Phi}^2 \right)^{\frac{1}{2}} \xrightarrow{P} 0$$

Proof of rate of convergence

Derivation of the rate of convergence will be based on theorem 3.2.5 of [24]. We showed in the proof of consistency that

$$(\theta_0) - (\theta) \geq Cd^2(\theta_0, \theta).$$

and that

$$\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq I_{1,n} + I_{2,n},$$

where $I_{1,n} = (\mathbb{P}_n - P)[l(\beta_0, \phi_{0,n}; X) - l(\beta_0, \phi_0; X)]$ and $I_{2,n} = P[l(\beta_0, \phi_{0,n}; X) - l(\beta_0, \phi_0; X)]$. Applying a Taylor expansion leads to

$$I_{1,n} = (\mathbb{P}_n - P) \left[\dot{l}_2(\beta_0, \tilde{\phi}; X)(\phi_{0,n} - \phi_0) \right] = n^{-p\nu + \varepsilon} (\mathbb{P}_n - P) \left[\dot{l}_2(\beta_0, \tilde{\phi}; X) \frac{\phi_{0,n} - \phi_0}{n^{-p\nu + \varepsilon}} \right]$$

for any $0 < \varepsilon < 1/2 - p\nu$. The uniform boundedness of $\dot{l}_2(\beta_0, \tilde{\phi}; X)$ due to the conditions C1–C4 and the fact that $\|\phi_{0,n} - \phi_0\|_{\infty} = O(n^{-p\nu})$ [26] lead to

$$P \left[\dot{l}_2(\beta_0, \tilde{\phi}; X) \frac{\phi_{0,n} - \phi_0}{n^{-p\nu + \varepsilon}} \right]^2 \rightarrow 0.$$

Since \mathcal{F}_2 is P -Donsker and by corollary 2.3.12 of [24] we have that

$$(\mathbb{P}_n - P) \left[i_2(\beta_0, \tilde{\phi}; X) \frac{\phi_{0,n} - \phi_0}{n^{-p\nu + \varepsilon}} \right] = o_p(n^{-1/2}).$$

Consequently,

$$I_{1,n} = o_p(n^{-p\nu + \varepsilon} n^{-1/2}) = o_p(n^{-2p\nu}).$$

Since the function $m(x) = x \log x - x + 1$ is bounded by $(x - 1)^2$ in a neighbourhood of $x = 1$, it follows that

$$(\theta_0) - (\theta_{0,n}) \leq C \|\phi_0 - \phi_{0,n}\|_{\Phi}^2 = O(n^{-2p\nu})$$

and thus

$$I_{2,n} = (\theta_{0,n}) - (\theta_0) \geq -O(n^{-2p\nu}).$$

Consequently,

$$\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq -O_p(n^{-2p\nu}) = -O_p(n - 2 \min(p\nu, (1-\nu)/2)).$$

Defining the class of functions $\mathcal{F}_3(\eta) = \{I(\theta; X) - I(\theta_0; X): \phi_j \in \mathcal{M}_{n,j} \text{ and } d(\theta, \theta_0) \leq \eta\}$ and using similar arguments as in the proofs of consistency, we have that

$$\log N_{[]}(\varepsilon, \mathcal{F}_3(\eta), L_2(P)) \leq Ck(N_j + p_j) \log(\eta/\varepsilon)$$

and thus

$$J_{[]}(\eta, \mathcal{F}_3(\eta), L_2(P)) = \int_0^\eta \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{F}_3(\eta), L_2(P))} d\varepsilon \leq Ck(N_j + p_j)^{1/2} \eta.$$

Now, using the uniform boundedness of $I(\theta; X)$ as a result of conditions C1 and C3, and theorem 3.4.1 of [24], the function $\phi_n(\eta)$ of theorem 3.2.5 [24] is

$$\phi_n(\eta) = (N_j + p_j)^{1/2} \eta + \frac{(N_j + p_j)}{n^{1/2}}.$$

Also, we have that

$$n^{2p\nu} \phi_n(1/n^{p\nu}) = n^{p\nu} n^{\nu/2} + \frac{n^{2p\nu} n^{\nu}}{n^{1/2}} = n^{1/2} \left[n^{p\nu - (1-\nu)/2} + n^{2p\nu - (1-\nu)} \right].$$

Consequently, if $p\nu > (1 - \nu)/2$ then $n^{2p\nu} \phi_n(1/n^{p\nu}) = o_p(n^{1/2})$. Based on this and in accordance with [8] we conclude that if we choose $r_n = \min(p\nu, (1 - \nu)/2)$, then $r_n^2 \phi_n(1/r_n) \leq n^{1/2}$ and $\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq -O_p(r_n^{-2})$. Therefore

$$d(\hat{\theta}_n, \theta_0) = O_p(n^{-r_n})$$

Proof of asymptotic normality

In order to prove the asymptotic normality of the estimator for Euclidean parameter $\hat{\beta}_n$ we need to verify that the information matrix $I(\beta_0)$ is nonsingular and also the conditions of theorem 3 of [8]. These conditions are

- B1** $\mathbb{P}_n \dot{l}_1(\hat{\beta}_n, \hat{\phi}_n; X) = o_p(n^{-1/2})$ and $\mathbb{P}_n \dot{l}_2(\hat{\beta}_n, \hat{\phi}_n; X)(\xi_0) = o_p(n^{-1/2})$
- B2** $(\mathbb{P}_n - P)[\dot{I}^*(\hat{\beta}_n, \hat{\phi}_n; X) - \dot{I}^*(\beta_0, \phi_0; X)] = o_p(n^{-1/2})$
- B3** $P[\dot{I}^*(\hat{\beta}_n, \hat{\phi}_n; X) - \dot{I}^*(\beta_0, \phi_0; X)] = -I(\beta_0)(\hat{\beta}_n - \beta_0) + o_p(\|\hat{\beta}_n - \beta_0\|) + o_p(n^{-1/2})$

Information for β

At first we will show that the information matrix $I(\beta_0)$ is nonsingular. Generically, the likelihood for an individual observation is

$$l(\beta, \phi; X) = \sum_{j=1}^k \delta_j^1 \log [F_j(V; Z, \alpha_j)] + \sum_{j=1}^k \delta_j \log [F_j(U; Z, \alpha_j) - F_j(V; Z, \alpha_j)] \\ + (1 - \delta) \log \left[1 - \sum_{j=1}^k F_j(U; Z, \alpha_j) \right],$$

where

$$F_j(t; Z, \alpha_j) = \begin{cases} 1 - \{1 + \alpha_j \exp[\phi_j(t) + \beta_j' Z]\}^{-\frac{1}{\alpha_j}} & \text{if } 0 < \alpha_j < \infty \\ 1 - \exp\{-\exp[\phi_j(t) + \beta_j' Z]\} & \text{if } \alpha_j = 0. \end{cases}$$

We will carry out the calculations for the proportional subdistribution hazards model (i.e. $\alpha_j = 0$). The proof for any other combination of link functions with $0 < \alpha_j < \infty$, follows from similar arguments and calculations. The likelihood for an individual observation, under $\alpha_j = 0$ for all $j = 1, \dots, k$ is:

$$\begin{aligned}
 l(\beta, \phi; X) &= \sum_{j=1}^k \delta_j \log \left[e^{-\Lambda_j(V)e^{\beta'_j Z}} - e^{-\Lambda_j(U)e^{\beta'_j Z}} \right] \\
 &+ \sum_{j=1}^k \delta_j^1 \log \left[1 - e^{-\Lambda_j(V)e^{\beta'_j Z}} \right] \\
 &+ (1-\delta) \log \left\{ 1 - \sum_{j=1}^k \left[1 - e^{-\Lambda_j(U)e^{\beta'_j Z}} \right] \right\},
 \end{aligned}$$

where $\Lambda_j = \exp(\phi_j)$ is the cumulative subdistribution hazard for the j -th cause of failure [11]. The score function for $\beta_j (j = 1, \dots, k)$ is

$$\begin{aligned}
 \dot{l}_{\beta_j} &= \delta_j \frac{e^{-\Lambda_j(V)e^{\beta'_j Z}} [-\Lambda_j(V)Ze^{\beta'_j Z}] - e^{-\Lambda_j(U)e^{\beta'_j Z}} [-\Lambda_j(U)Ze^{\beta'_j Z}]}{e^{-\Lambda_j(V)e^{\beta'_j Z}} - e^{-\Lambda_j(U)e^{\beta'_j Z}}} \\
 &+ \delta_j^1 \frac{e^{-\Lambda_j(V)e^{\beta'_j Z}} \Lambda_j(V)Ze^{\beta'_j Z}}{1 - e^{-\Lambda_j(V)e^{\beta'_j Z}}} + (1-\delta) \frac{e^{-\Lambda_j(U)e^{\beta'_j Z}} [-\Lambda_j(U)Ze^{\beta'_j Z}]}{1 - \sum_{j=1}^k \left[1 - e^{-\Lambda_j(U)e^{\beta'_j Z}} \right]} \\
 &= -\delta_j Z e^{\beta'_j Z} \frac{\Lambda_j(V)e^{-\Lambda_j(V)e^{\beta'_j Z}} - \Lambda_j(U)e^{-\Lambda_j(U)e^{\beta'_j Z}}}{e^{-\Lambda_j(V)e^{\beta'_j Z}} - e^{-\Lambda_j(U)e^{\beta'_j Z}}} \\
 &+ \delta_j^1 Z e^{\beta'_j Z} \frac{\Lambda_j(V)e^{-\Lambda_j(V)e^{\beta'_j Z}}}{1 - e^{-\Lambda_j(V)e^{\beta'_j Z}}} + (1-\delta) Z e^{\beta'_j Z} \frac{\Lambda_j(U)e^{-\Lambda_j(U)e^{\beta'_j Z}}}{1 - \sum_{j=1}^k \left[1 - e^{-\Lambda_j(U)e^{\beta'_j Z}} \right]}.
 \end{aligned}$$

Let

$$h_j = \frac{\partial \Lambda_{j,s}}{\partial s} \Big|_{s=0}.$$

The score operator for $\Lambda_j (j = 1, \dots, k)$ is

$$\begin{aligned}
 \dot{l}_s^j &= -\delta_j e^{\beta'_j Z} \frac{h_j(V)e^{-\Lambda_j(V)e^{\beta'_j Z}} - h_j(U)e^{-\Lambda_j(U)e^{\beta'_j Z}}}{e^{-\Lambda_j(V)e^{\beta'_j Z}} - e^{-\Lambda_j(U)e^{\beta'_j Z}}} \\
 &+ \delta_j^1 e^{\beta'_j Z} \frac{h_j(V)e^{-\Lambda_j(V)e^{\beta'_j Z}}}{1 - e^{-\Lambda_j(V)e^{\beta'_j Z}}} \\
 &- (1-\delta) e^{\beta'_j Z} \frac{h_j(U)e^{-\Lambda_j(U)e^{\beta'_j Z}}}{1 - \sum_{j=1}^k \left[1 - e^{-\Lambda_j(U)e^{\beta'_j Z}} \right]}.
 \end{aligned}$$

Let

$$\alpha_1^j = \frac{e^{-\Lambda_j(V)e^{\beta_j'Z}}}{1-e^{-\Lambda_j(V)e^{\beta_j'Z}}}, \alpha_2^j = \frac{e^{-\Lambda_j(V)e^{\beta_j'Z}}}{e^{-\Lambda_j(V)e^{\beta_j'Z}} - e^{-\Lambda_j(U)e^{\beta_j'Z}}} = \alpha_3^j + 1,$$

$$\alpha_3^j = \frac{e^{-\Lambda_j(U)e^{\beta_j'Z}}}{e^{-\Lambda_j(V)e^{\beta_j'Z}} - e^{-\Lambda_j(U)e^{\beta_j'Z}}}, \alpha_4^j = \frac{e^{-\Lambda_j(U)e^{\beta_j'Z}}}{1 - \sum_{j=1}^k \left[1 - e^{-\Lambda_j(U)e^{\beta_j'Z}} \right]},$$

$$\frac{\alpha_1^j}{1-\alpha_1^j} = e^{-\Lambda_j(V)e^{\beta_j'Z}}.$$

Then,

$$\begin{aligned} \dot{l}_{\beta_j} &= Ze^{\beta_j'Z} \left[-\delta_j \alpha_3^j \Lambda_j(U) + \delta_j \alpha_2^j \Lambda_j(V) + \delta_j^1 \alpha_1^j \Lambda_j(V) - (1-\delta) \Lambda_j(U) \alpha_4^j \right] \\ &= Ze^{\beta_j'Z} \left\{ \delta_j^1 \Lambda_j(V) \alpha_1^j - \delta_j \left[\Lambda_j(V) \alpha_3^j - \Lambda_j(U) \alpha_2^j \right] - (1-\delta) \Lambda_j(U) \alpha_4^j \right\} \end{aligned}$$

and

$$\dot{l}_s^j = e^{\beta_j'Z} \left\{ \delta_j^1 h_j(V) \alpha_1^j - \delta_j \left[h_j(V) \alpha_3^j - h_j(U) \alpha_2^j \right] - (1-\delta) h_j(U) \alpha_4^j \right\}.$$

Next, for any $j = 1, \dots, k$, we need to find

$$E(\dot{l}_{\beta_j} | T_1=t_1, \dots, T_k=t_k, Z)$$

and

$$E(\dot{l}_s^j | T_1=t_1, \dots, T_k=t_k, Z).$$

The random variables here are $(\delta_j^1, \delta_j, \delta, U, V)$. Assume U, V the same for all $j = 1, \dots, k$ and that condition C3 holds. Since the event $\delta_j^1 = 1$ is equivalent to $T_j < V$ and $T_j > T_l$ for any $l = 1, \dots, k$ we have that

$$E(\delta_j^1 = 1 | T_1=t_1, \dots, T_k=t_k, Z) = I(t_j \leq t_l, l=1, \dots, k) \int_{t_j}^b g(v, u|z) \int_{v+\eta}^b dv du,$$

where $g(v, u|z)$ if the density of (V, U) conditional on $Z = z$. Similarly, since the event $\delta_j = 1$ is equivalent to $T_j \in (V, U)$ and $T_j > T_l$ for any $l = 1, \dots, k$ we have that

$$E(\delta_j=1|T_1=t_1, \dots, T_k=t_k, Z) = I(t_j \leq t_l, l=1, \dots, k) \int_a^{t_j} \int_{t_j}^b g(v, u|z) I(u-v \geq \eta) dvdu.$$

Finally, we have that

$$\begin{aligned} E(\delta=0|T_1=t_1, \dots, T_k=t_k, Z) &= P(U \leq t_j, j=1, \dots, k, U-V \geq \eta) \\ &= \int_a^{t-\eta} \int_{v+\eta}^t g(v, u|z) dvdu, \end{aligned}$$

where $t = \min(t_j; j = 1, \dots, k)$. We also have that,

$$(i_s^j)^T i_{\theta_j} = E_Z E(i_{\theta_j} | T_1=t_1, \dots, T_k=t_k, Z).$$

Now

$$\begin{aligned} E(i_{\beta_j} | T_1=t_1, \dots, T_k=t_k, Z) &= Z e^{\beta_j' Z} \left\{ I(t_j \leq t_l, l=1, \dots, k) \int_{t_j}^b \int_{v+\eta}^b \Lambda_j(v) \alpha_1^j g(v, u|z) dvdu \right. \\ &\quad - I(t_j \leq t_l, l=1, \dots, k) \times \\ &\quad \left. \int_a^{t_j} \int_{t_j}^b [\Lambda_j(v) \alpha_3^j - \Lambda_j(u) \alpha_2^j] I(u-v \geq \eta) g(v, u|z) \times \right. \\ &\quad \left. I(u-v \geq \eta) dvdu \right. \\ &\quad \left. - \int_a^{t-\eta} \int_{v+\eta}^t \Lambda_j(u) \alpha_4^j g(v, u|z) dvdu \right\}. \end{aligned}$$

Similarly,

$$(i_s^j)^T i_s^j = E_Z E(i_s^j | T_1=t_1, \dots, T_k=t_k, Z).$$

Now,

$$\begin{aligned} E(i_s^j | T_1=t_1, \dots, T_k=t_k, Z) &= e^{\beta_j' Z} \left\{ I(t_j \leq t_l, l=1, \dots, k) \int_{t_j}^b \int_{v+\eta}^b h_j(v) \alpha_1^j g(v, u|z) dvdu \right. \\ &\quad - I(t_j \leq t_l, l=1, \dots, k) \times \\ &\quad \left. \int_a^{t_j} \int_{t_j}^b [h_j(v) \alpha_3^j - h_j(u) \alpha_2^j] I(u-v \geq \eta) g(v, u|z) \times \right. \\ &\quad \left. I(u-v \geq \eta) dvdu \right. \\ &\quad \left. - \int_a^{t-\eta} \int_{v+\eta}^t h_j(u) \alpha_4^j g(v, u|z) dvdu \right\}. \end{aligned}$$

Let,

$$\begin{aligned}
 b_1^j(v, u) &= E_z[e^{\beta_j' Z} \alpha_1^j g(u, v|z)] \\
 b_2^j(v, u) &= E_z[e^{\beta_j' Z} \alpha_2^j g(u, v|z)] \\
 b_3^j(v, u) &= E_z[e^{\beta_j' Z} \alpha_3^j g(u, v|z)] \\
 b_4^j(v, u) &= E_z[e^{\beta_j' Z} \alpha_4^j g(u, v|z)].
 \end{aligned}$$

Then,

$$\begin{aligned}
 (l_s^j)^T l_s^j &= I(t_j \leq t_l, l=1, \dots, k) \int_{t_j}^b \int_{v+\eta}^b h_j(v) b_1^j(v, u) g(v, u|z) dv du \\
 &- I(t_j \leq t_l, l=1, \dots, k) \times \\
 &\quad \int_a^{t_j} \int_{t_j}^b [h_j(v) b_3^j(v, u) - h_j(u) b_2^j(v, u)] I(u-v \geq \eta) g(v, u|z) I(u-v \geq \eta) dv du \\
 &- \int_a^{t-\eta} \int_{v+\eta}^t h_j(u) b_4^j(v, u) g(v, u|z) dv du.
 \end{aligned}$$

Let,

$$\begin{aligned}
 c_1^j(v, u) &= E_z[Z e^{\beta_j' Z} \alpha_1^j g(u, v|z)] \\
 c_2^j(v, u) &= E_z[Z e^{\beta_j' Z} \alpha_2^j g(u, v|z)] \\
 c_3^j(v, u) &= E_z[Z e^{\beta_j' Z} \alpha_3^j g(u, v|z)] \\
 c_4^j(v, u) &= E_z[Z e^{\beta_j' Z} \alpha_4^j g(u, v|z)].
 \end{aligned}$$

Then,

$$\begin{aligned}
 (l_s^j)^T l_{\beta_j} &= I(t_j \leq t_l, l=1, \dots, k) \int_{t_j}^b \int_{v+\eta}^b \Lambda_j(v) c_1^j(u, v) g(v, u|z) dv du \\
 &- I(t_j \leq t_l, l=1, \dots, k) \times \\
 &\quad \int_a^{t_j} \int_{t_j}^b [\Lambda_j(v) c_3^j(v, u) - \Lambda_j(u) c_2^j(v, u)] I(u-v \geq \eta) g(v, u|z) I(u-v \geq \eta) dv du \\
 &- \int_a^{t-\eta} \int_{v+\eta}^t \Lambda_j(u) c_4^j(v, u) g(v, u|z) dv du.
 \end{aligned}$$

Denote,

$$(l_s^j)^T l_s^j = \begin{pmatrix} L_1(t_1) \\ \vdots \\ \vdots \\ \vdots \\ L_k(t_k) \end{pmatrix}$$

and

$$\begin{pmatrix} l_s^j \end{pmatrix}^T l_{\beta_j} = \begin{pmatrix} R_1(t_1) \\ \cdot \\ \cdot \\ \cdot \\ R_k(t_k) \end{pmatrix}$$

We need to find h_1, \dots, h_k s.t.

$$L_j(t_j) = R_j(t_j), \quad j=1, \dots, k.$$

or equivalently

$$\frac{\partial L_j(t_j)}{\partial t_j} = \frac{\partial R_j(t_j)}{\partial t_j}, \quad j=1, \dots, k.$$

We have that

$$\begin{aligned}
 \frac{\partial L_j(t_j)}{\partial t_j} &= \left\{ -\int_{t_j+\eta}^b h_j(t_j) b_1^j(u, t_j) du \right. \\
 &- \int_{t_j}^b [h_j(t_j) b_3^j(u, t_j) - h_j(u) b_2^j(u, t_j)] I(u \geq t_j + \eta) du \\
 &+ \int_a^{t_j} [h_j(v) b_3^j(t_j, v) - h_j(t_j) b_2^j(t_j, v)] I(u \geq t_j + \eta) du \\
 &\left. - \int_a^{t_j-\eta} h_j(t_j) b_4^j(t_j, v) dv \right\} I(t_j \leq t_l, l=1, \dots, k)
 \end{aligned}$$

Let,

$$\begin{aligned}
 b_j(t_j) &= \int_{t_j+\eta}^b b_1^j(u, t_j) du + \int_{t_j+\eta}^b b_2^j(u, t_j) du \\
 &+ \int_a^{t_j-\eta} b_3^j(t_j, v) dv + \int_a^{t_j-\eta} b_4^j(t_j, v) dv.
 \end{aligned}$$

Then,

$$\begin{aligned}
 \frac{\partial L_j(t_j)}{\partial t_j} &= -b_j(t_j) h_j(t_j) + \int_{t_j+\eta}^b h_j(x) b_3^j(x, t_j) dx \\
 &+ \int_a^{t_j-\eta} h_j(x) b_2^j(t_j, x) dx
 \end{aligned}$$

for $t_j \leq t_l$ for any $l=1, \dots, k$. Similarly let

$$\begin{aligned}
 c_j(t_j) &= \int_{t_j+\eta}^b c_1^j(u, t_j) du + \int_{t_j+\eta}^b c_2^j(u, t_j) du \\
 &+ \int_a^{t_j-\eta} c_3^j(t_j, v) dv + \int_a^{t_j-\eta} c_4^j(t_j, v) dv.
 \end{aligned}$$

Then,

$$\begin{aligned} \frac{\partial R_j(t_j)}{\partial t_j} &= -c_j(t_j)\Lambda_j(t_j) + \int_{t_j+\eta}^b \Lambda_j(x)c_3^j(x, t_j)dx \\ &+ \int_a^{t_j-\eta} \Lambda_j(x)c_2^j(t_j, x)dx \end{aligned}$$

for $t_j = t_b$ for any $l = 1, \dots, k$. Therefore

$$h_j^*(t_j) - \int K(t_j, x)h_j^*(x)dx = m(t_j),$$

where

$$\begin{aligned} h_j^*(t_j) &= b_j(t_j)h_j(t_j), \\ K(t_j, x) &= \frac{b_3^j(x, t_j)I(t_j+\eta < x < b) + b_2^j I(a < x < t_j-\eta)}{b_j(t_j)} \end{aligned}$$

and

$$m(t_j) = -\frac{\partial R_j(t_j)}{b_j(t_j)}.$$

Now, as in [29] by lemma 6.1 [29], the fact that $K(t_j, x)$ is a bounded kernel and Fredholm integral equations results, there exists a $O(t_j, x)$ such that

$$h_j^*(t_j) = m(t_j) - \int \Gamma(t_j, x)m(x)dx.$$

Thus, similarly to [29] (pp. 159), h_j^* is differentiable with a bounded derivative.

Verification of conditions B1–B3

For condition B1 we need to show that

$$\mathbb{P}_n \dot{l}_2(\hat{\beta}_n, \hat{\phi}_n; X)(\xi_0) = o_p(n^{-1/2})$$

since $\mathbb{P}_n \dot{l}_1(\hat{\beta}_n, \hat{\phi}_n; X) = 0$. ξ_0 has a bounded derivative and it is also a function with bounded variation [8]. Define $S_n(D_n, K_n, m)$ to be the space of polynomial spline functions s of order $m - 1$, with the restriction of s to any of the K_n subinterval being a polynomial of order $m - 1$ and for $m \geq 2$ and $0 \leq m' \leq m - 2$, s is m' times continuously differentiable on $[a, b]$. Then it can be shown using the argument in [30] pp. 435–436 that there exists a $\xi_{0,n} \in S_n(D_n, K_n, m)$ such that

$$\|\xi_{0,n} - \xi_0\|_{\Phi} = O(q_n^{-1}) = O(n^{-\nu})$$

and $\mathbb{P}_n \dot{l}_2(\hat{\theta}_n; X)(\xi_{0,n}) = 0$. Consequently,

$$\mathbb{P}_n \dot{l}_2(\hat{\theta}_n; X)(\xi_0) = I_{3,n} + I_{4,n},$$

where

$$I_{3,n} = (\mathbb{P}_n - P) \dot{l}_2(\hat{\theta}_n; X)(\xi_0 - \xi_{n,0})$$

and

$$I_{4,n} = P \left[\dot{l}_2(\hat{\theta}_n; X)(\xi_0 - \xi_{n,0}) - \dot{l}_2(\theta_0; X)(\xi_0 - \xi_{n,0}) \right].$$

Let $\mathcal{F}_4 = \{ \dot{l}_2(\theta; x)(\xi_0 - \xi) : \theta \in \mathcal{M}_n, \xi \in S_n(D_n, K_n, m) \text{ and } \|\xi_0 - \xi\|_{\Phi} \leq n^{-\nu} \}$. Similarly, it can be shown that

$$N_{[]}(\varepsilon, \mathcal{F}_4, L_2(P)) \leq C(1/\varepsilon)^d (1/\varepsilon)^{Ck(N_j+p_j)} (1/\varepsilon)^{Ck(N_j+p_j)},$$

and this leads to \mathcal{F}_4 being Donsker. Also, for any $r \in \mathcal{F}_4$, $P^2 r \rightarrow 0$ as $n \rightarrow \infty$ and thus $I_{3,n} = o_p(n^{-1/2})$ by corollary 2.3.12 of [24]. Finally, based on the Cauchy-Schwartz inequality and regularity conditions C1–C4 we have that

$$\begin{aligned} I_{4,n} &\leq C d(\hat{\theta}_n, \theta_0) \|\xi_0 - \xi_{0,n}\|_{\Phi} \\ &= O_p(n - \min(p\nu, (1-\nu)/2) n^{-\nu}) \\ &= O_p(n - \min(\nu(p+1), (1+\nu)/2)) \\ &= o_p(n^{-1/2}), \end{aligned}$$

and thus we proved condition B1.

For condition B2 define the class of functions

$$\mathcal{F}_5(\eta) = \{ l^*(\theta; x) - l^*(\theta_0; x) : \theta \in \mathcal{M}_{j,n} \text{ and } d(\theta, \theta_0) \leq \eta \}.$$

Using similar arguments as before, we can show that $\mathcal{F}_5(\eta)$ is P-Donsker and that for any $r \in \mathcal{F}_5$, $P^2 r \rightarrow 0$ as $\eta \rightarrow \infty$ and therefore condition B2 holds. Finally, it is a straightforward

task to show condition B3 based on Taylor expansion and the convergence rate of the estimator. Thus

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N[0, I^{-1}(\beta_0)]$$

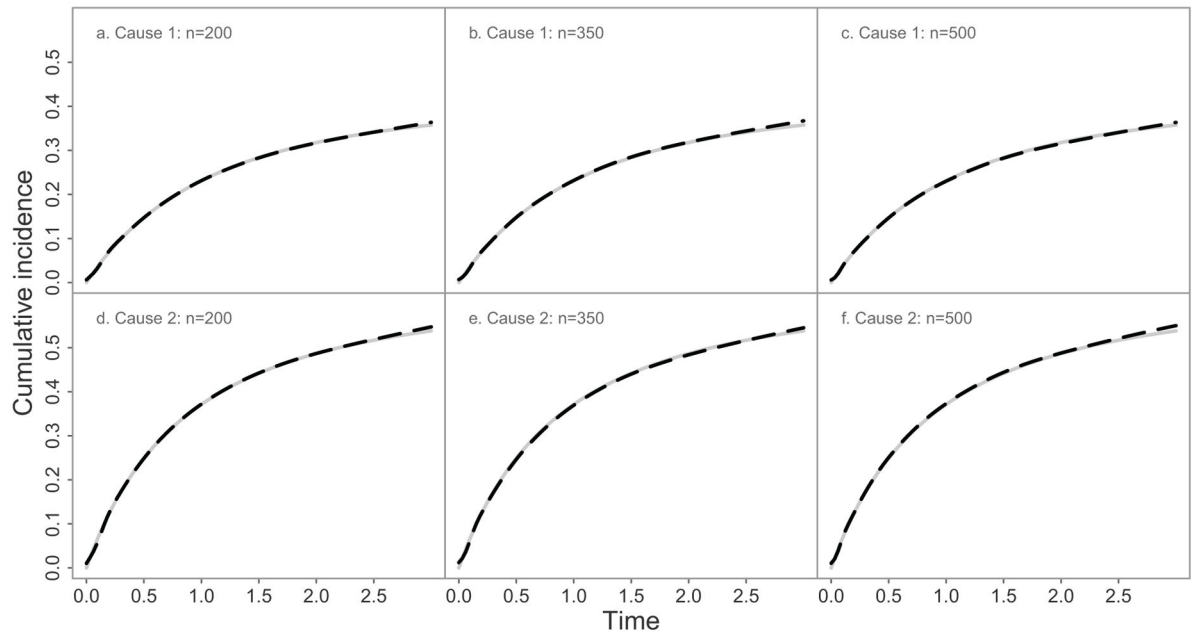


Figure 1. True (grey lines) and average estimated (black dashed lines) baseline cumulative incidence functions based on 1,000 simulation replications according to cause of failure and sample size n .

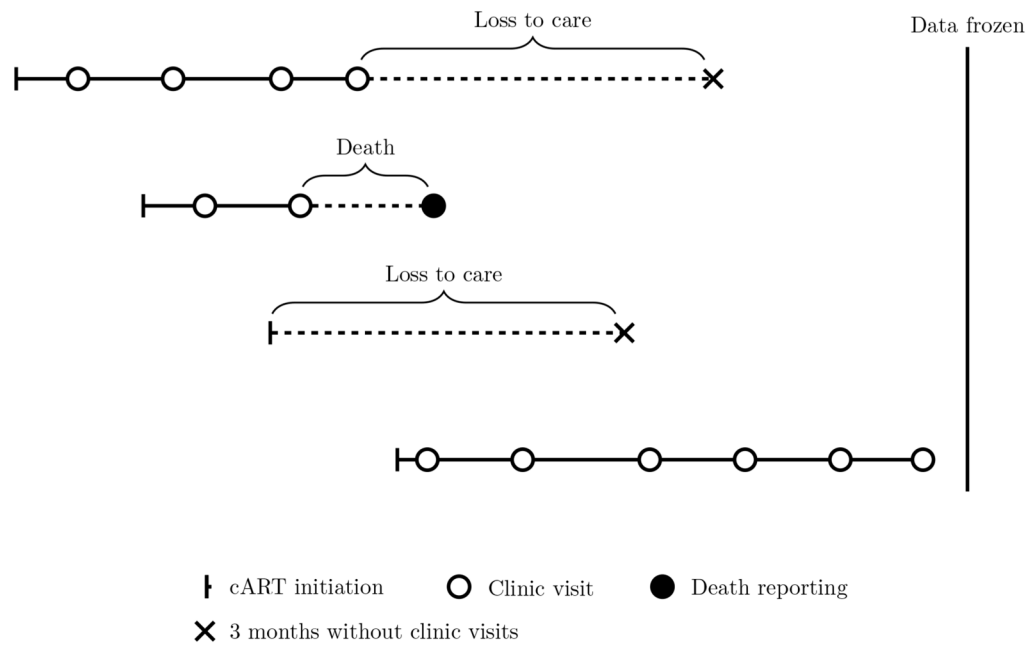


Figure 2.
Example of four patients in a typical HIV care and treatment program.

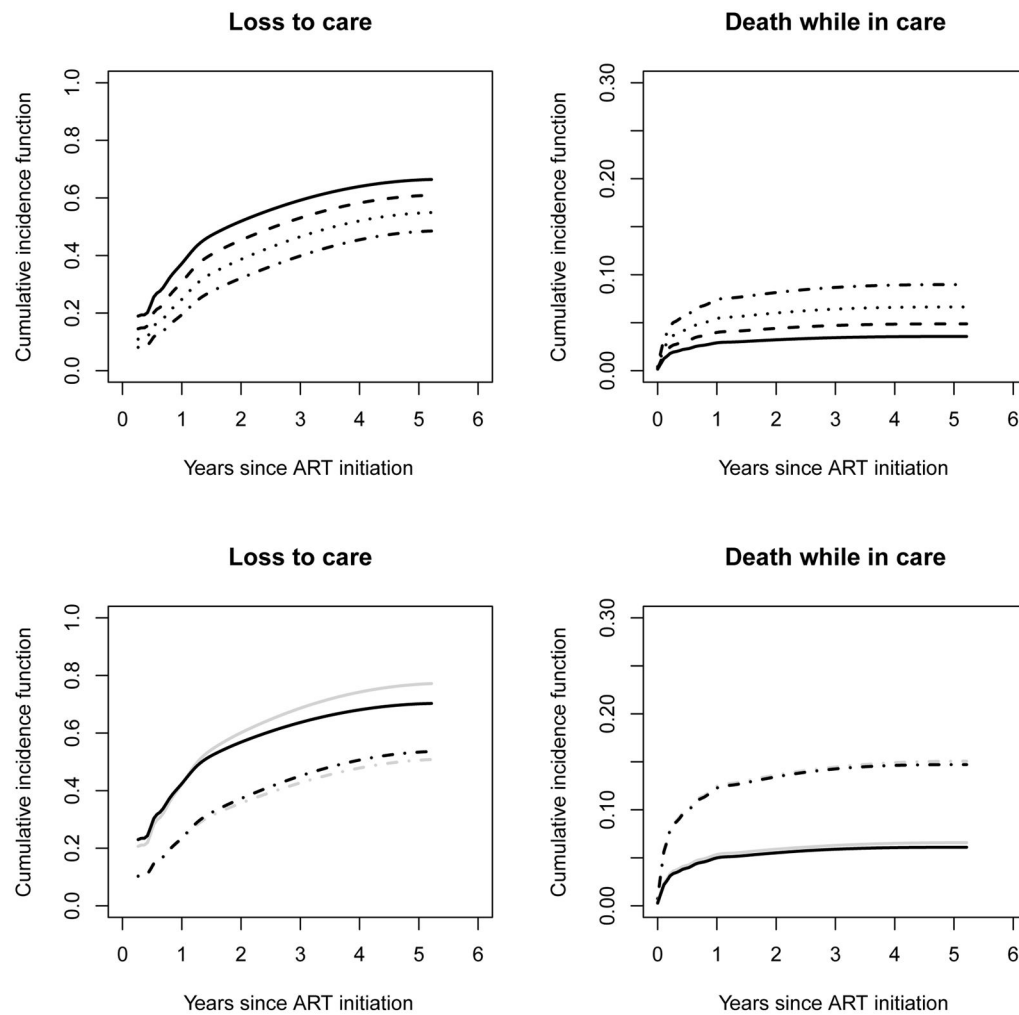


Figure 3. Estimated cumulative incidence of loss to care (top left panel) and of death (top right panel) for a female patient with a CD4 count of 120 cells/ μ l at ART initiation, according to age at ART initiation (solid lines = 20 years; dashed lines = 30 years; dotted lines = 40 years; dashed-dotted lines = 50 years) based on the optimal model according to the grid search, with $\alpha = (2, 1.2)$, and estimated cumulative incidence of loss to care (bottom left panel) and of death (bottom right panel) for a male patient with a CD4 count of 120 cells/ μ l at ART initiation, according to age (solid lines = 20 years; dashed-dotted lines = 50 years), based on the optimal model (black lines) and the Fine-Gray – proportional odds model (gray lines) with $\alpha = (0, 1)$.

Simulation results regarding the Euclidean parameter $\beta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})'$ based on 1,000 replications. Bootstrap based on 50 bootstrap samples was used for standard error estimation.

Table 1

n		β_{11}	β_{12}	β_{21}	β_{22}
200	% bias	-0.311	3.239	2.631	2.950
	MCS D	0.281	0.134	0.279	0.134
	ASE	0.275	0.133	0.273	0.133
	ECP	0.943	0.945	0.938	0.947
350	% bias	0.464	-0.037	2.626	0.085
	MCS D	0.196	0.095	0.191	0.095
	ASE	0.203	0.098	0.204	0.098
	ECP	0.952	0.941	0.955	0.941
500	% bias	-1.403	0.640	-0.137	0.622
	MCS D	0.169	0.082	0.171	0.082
	ASE	0.169	0.081	0.168	0.081
	ECP	0.945	0.940	0.946	0.938

MCS D: Monte Carlo standard deviation of the estimates;

ASE: Average standard error of the estimates;

ECP: Empirical coverage probability

Table 2

Descriptive characteristics of the study sample.

	In HIV care (<i>n</i> =2,232) <i>n</i> (%)	LTC ¹ (<i>n</i> =690) <i>n</i> (%)	Death ² (<i>n</i> =131) <i>n</i> (%)
Gender			
<i>Male</i>	681 (30.5)	207 (30.0)	62 (47.3)
<i>Female</i>	1,551 (69.5)	483 (70.0)	69 (52.7)
HIV status disclosed			
<i>No</i>	763 (34.2)	279 (40.4)	57 (43.5)
<i>Yes</i>	1,469 (65.8)	411 (59.6)	74 (56.5)
	Median (IQR)	Median (IQR)	40.9 (35.0, 50.8)
Age³ (years)	37.1 (30.4, 44.9)	35.2 (28.6, 41.7)	
CD4³ (cells/μA)	199 (152, 276)	188 (140, 262)	163 (131, 213)

¹ Loss to care² While in HIV care³ At ART initiation

Table 3

Comparison between the results from the Fine-Gray – proportional odds model (FGPO) with $\alpha = (0, 1)$ and the optimal based on the grid search model with $\alpha = (2, 1.2)$

Outcome	Covariate	$\alpha = (0, 1)$ mRR ¹ / $\hat{\beta}$ p-value)	$\alpha = (2, 1.2)$ $\hat{\beta}$ (p-value)
A. Loss to care	Age at ART initiation		
	<i>per 10 years</i>	0.783 ² / -0.244 (<0.001)	-0.348 (<0.001)
	Gender		
	<i>Male vs Female</i>	1.170 ² / 0.157 (0.031)	0.273 (0.005)
	CD4 at ART initiation		
	<i>per 100 cells/μl</i>	1.002 ² / 0.002 (0.947)	0.001 (0.976)
B. Death	Age at ART initiation		
	<i>per 10 years</i>	1.360 ³ / 0.308 (<0.001)	0.329 (<0.001)
	Gender		
	<i>Male vs Female</i>	1.758 ³ / 0.564 (0.008)	0.567 (0.008)
	CD4 at ART initiation		
	<i>per 100 cells/μl</i>	0.699 ³ / -0.358 (0.003)	-0.346 (0.006)

¹Measure of relative risk

²Subdistribution hazard ratio

³Odds ratio