Revista Colombiana de Estadística - Applied Statistics
July 2024, volume 47, issue 2, pp. 385–405
http://doi.org/10.15446/rce.v47n2.110557

Joint Occurrences of Competing Risks and Multivariate Longitudinal Data: A Prediction Investigation for the HIV.long Data

Ocurrencias conjuntas de riesgos competitivos y datos longitudinales multivariados: una investigacion de prediccion para los datos de HIV.long

Jaber Kazempoor[1,a], Arezou Habibirad[2,b], Sanjoy Sinha[3,c]

[1]Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

[2]Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

[3]School of Mathematics and Statistics, 4302 Herzberg Laboratories, Carleton University, Ottawa, Canada

Abstract

In this article, some prediction strategies are introduced for event times, where multivariate data with competing or semi-competing risks are simultaneously collected. Without loss of generality, the proposed methods can be used to analyze multivariate longitudinal data with competing or semi-competing risks, often encountered in social sciences and sports activities. Regarding the situations mentioned earlier, we can provide the prediction values of: I. Time of occurrences of any cause for specific individuals II. Time of subsequent events for some cause in other individuals III. The covariate values on predicted time of I and II. Accordingly, doctor assistants or nurses can schedule good visiting times based on I and II. Item III can provide the missing values of all covariates that are utilized for better modeling. The corresponding statistical background is extensively discussed. Finally, an actual data set has been analyzed, the prediction values are provided, and their performances are assessed.

Key words: Competing or semi-competing risks data; Covariate prognosis; Cumulative incidence functions; Multivariate longitudinal data; Order statistics.

[a]Ph.D(c). E-mail: kazempoor.jaber@mail.um.ac.ir
[b]Ph.D. E-mail: ahabibi@um.ac.ir
[c]Ph.D. E-mail: sinha@math.carleton.ca

Resumen

En este articulo, se presentan algunas estrategias de prediccion para los tiempos de eventos, donde se recopilan datos multivariados con riesgos competitivos o semi-competitivos de manera simultanea. Sin perdida de generalidad, los metodos propuestos se pueden utilizar para analizar datos longitudinales multivariados con riesgos competitivos o semi-competitivos, que a menudo se encuentran en las ciencias sociales y actividades deportivas. En relacion con las situaciones mencionadas anteriormente, podemos proporcionar los valores de prediccion de: I. Tiempo de ocurrencia de cualquier causa para individuos especificos. II. Tiempo de eventos subsiguientes para alguna causa en otros individuos. III. Los valores de covariables en el tiempo predicho de I y II. En consecuencia, los asistentes medicos o enfermeras pueden programar buenos momentos de visita en funcion de I y II. El item III puede proporcionar los valores faltantes de todas las covariables que se utilizan para un mejor modelado. El fondo estadistico correspondiente se discute ampliamente. Finalmente, se ha analizado un conjunto de datos reales, se proporcionan los valores de prediccion y se evaluan sus rendimientos.

***Palabras clave*:** Riesgos competitivos o semi-competitivos; Pronostico de covariables; Funciones de incidencia acumulativa; Datos longitudinales multivariados; Estadisticas de orden.

1. Introduction

In the vast majority of survival analysis research, the studied cases deal with an exciting event. However, this event may not be one fatal event and maybe come from a certain mutual event or one of the different risks (Putter et al., 2007; Berry et al., 2010). These various risks called "competing risk", emerge in a variety of disciplines such as biomedical methods, reliability engineering, quality life, and so forth. Besides, in a number of clinical trials, an individual can experience both non-fatal and fatal events, where the fatal censor other occasions but not vice versa, which is called "semi competing risk" (Fine et al., 2001; Haneuse & Lee, 2016; Kim et al., 2019). Nevertheless, it often occurs that both types of these data have been subject to interval censoring (Kim et al., 2019; Bakoyannis et al., 2017; Barrett et al., 2011). Interval censoring frequently appears in the biomedical and clinical studies regarding the exact event time is not precisely observed yielding the event time lies between two examination times such as clinical visits. Furthermore, in the follow-up study with competing or semi competing events, interval censoring may be needed where the last observation time prior to the interesting time and the first observation time after this time is considered lower and upper bound (Bakoyannis et al., 2017; Park et al., 2019).

Both "competing" and "semi competing" risks data frequently arise in survival analysis (Putter et al., 2007; Berry et al., 2010; Fine et al., 2001; Haneuse & Lee, 2016; Kim et al., 2019) and joint modeling of longitudinal and time to event data (Bakoyannis et al., 2017; Park et al., 2019; Williamson et al., 2008; Bhattacharjee, 2019; Rouanet et al., 2015; Hickey et al., 2018). Many different interesting topics have been studied in both of the foregoing cases, but one of the crucial issues is

the estimation of cumulative incidence function (CIF) and cause-specific CIF. The CIF is defined as

$$F_j(t) = Pr(T \leq t, C = j), \quad j = 1, 2, \ldots, m,$$

where T stands for time to event (fatal or non fatal) and j can be one of m different causes that results in an interesting event. However, in many clinical trials, biomedical research and follow up studies, in addition to the time of incidence, some covariates affecting this time are also measured noticed by the vector $\underline{Z} = (Z_1, Z_2, \ldots, Z_k)$. It is logical to assess the behavior of the cumulative probability of failure due to each cause in the presence of the remaining causes of failure, given the covariates, $\underline{z} = (z_1, z_2, \ldots, z_k)$. The corresponding function is called cause-specific CIF and is defined in the form of

$$F_j(t; z) = Pr(T \leq t, C = j \mid \underline{Z} = \underline{z}), \quad j = 1, 2, \ldots, m.$$

$\underline{z}$ is a covariate collected during the experiment. The vector can consist of empty or any dimension that completely depends on the operator's desire and design situation. Most of the papers related to the competing/semi-competing risks data condensed on these two functions. The cause-specific CIF represents the instantaneous risk of failure from a specific cause while the CIF quantifies the cumulative risk of incidence from a competing/semi-competing risk.

In a simply word, we may faced with missing or non-reporting t, and j values in practice. The reasons behind this, maybe return to operator or cost of observing and so on. The missing t, and j respectively refers to missing values and masked cause of failuress in literature (Moharib Alsarray et al., 2023; Kazempoor et al., 2022). The main innoviation of this study is using order statistics for prediction missing values and masked cause of failure with respect to covariate vector z. Previously, missing values and masked cause of failures have predicted separately. However, regarding both jointly, utilizing the order statistics are in the present investigation. The order statistics point of view and predicting two cases, in this paper have not investigate up to now. It is also worth mentioning that, another motivation is regretting the common considering time to event data, especially interval censorship sampling. Instead, we extend the presented strategy using order statistics to these data types appearing in joint modeling of longitudinal data and competing risks. The accuracy of prediction issues are shown via Tables and Figure.

Lots of scholars have investigated inference on CIF/cause-specific CIF in all three estimation areas including parametric, non-parametric, and semi-parametric. In addition to the estimator type (parametric, non-parametric, or semi-parametric), various methods based on the frequentist occurred type of competing risks data (right censored, left truncated, or interval-censored) and the kind of observed causes (entirely or partially observed) are also presented. In the point of non-parametric estimations, the cause-specific CIF can be estimated utilizing the increment of the Nelson-Aalen method by censoring individuals who have experienced competing causes of failure, and similarly, the CIF may be estimated directly using the Nelson-Aalen estimator (Aalen & Johansen, 1978; Li, 2016). Nonetheless, in

the presence of covariates, there is a need for semi-parametric estimation. Accordingly, the cause-specific CIF can be estimated using the standard Cox proportional hazards model whereas considering failures from competing causes as independent right censoring events (Li, 2016; Prentice et al., 1978). In a similar way for estimating the CIF, using the Fine-Gray model[(Fine & Gray, 1999) generalized odds transformation model (Park et al., 2019) and other semi-parametric models (Fine, 2001; Meng, 1994; Mao & Lin, 2017). Contrary to the significant advantage of non-parametric and semi-parametric estimation including no underlying distributional assumption for the cause-specific CIF/CIF, such flexibility appears at the cost of efficiency less relative to the parametric method, mainly dealing with small sample sizes (Jeong & Oakes, 2003; Jeong & Fine, 2006*b*). In addition, in this case, formal maximum likelihood and Bayes estimation can be applied, unlike the semi-parametric analysis (Fine & Gray, 1999; Fine, 2001).

The fundamental identifiable quantities and substantial estimand in the case of competing/semi-competing risks data are the cause-specific CIF and CIF. The CIF is utilized in studying various disease prognoses (Barrett et al., 2011; Haneuse & Lee, 2016), Assessment intervention in populations (Bhattacharjee, 2019) and implementing science purposes (Mao & Lin, 2017). Briefly, in accordance with the cause-specific CIF/CIF, monitoring the prognosis of interested covariates can be implemented appertaining to the event time. However, predicting the exact time of occurrence of other causes given the covariate vector has not been investigated to date. This task may be matters from two points of perspective. Firstly, in some competing/semi-competing risks studies, it is important to predict the time of occurrence of a special event when the first cause has happened, doing prevention work. For instance, when a kidney has failures, it is helpful to predict the failure time of another kidney to estimate the time we have to replace it. Of course, the term “kidney failure" includes a lot of problems. Moreover in the same situation, when an individual has left the experiment (right censoring case), time prediction of the first failure in conformity with their measured features might be vital. Secondly, In the joint modeling of longitudinal and competing/semi-competing risks data with covariate information, the scholars may be interested in the assessment of the affection of specific covariates on the events. For instance, in the HIV/Acquired immunodeficiency syndrome (AIDS) study, the evaluation of covariates CD4, sex, age ... might be of interest. There was no doubt that the preliminary task of this end was to predict the exact time occurrence time. Accordingly, applying the polynomial regression on the favorite covariates is proposed.

A significant relevant example of the interval-censored competing risks data and our motivation in the present study emanates from HIV care and treatment programs. The HIV-infected people who receive care in such a program might be died while in care or loss to care. Individuals that are not maintained in care are less presumably to receive antiretroviral treatment (ART) and often, any other clinical services. Thereby, they have more risks to experience adverse clinical results involving death and viral rebound. The increasing outcomes of these affected patients may also considerably contribute to more extension of the HIV epidemic in their society. Hence, recognizing influential factors dealing with the mortality rate of these communities while in care and loss to care is so utilizable for opti-

mizing intercession, particularly in sub-Saharan Africa, the territory most affected by the HIV epidemic and an interesting region of the present study. In many cases in this sense, the exact time of death is known between the reported death date and the last clinical visit of a deceased person (Park et al., 2019). It is also correct that the loss to care has happened as no clinical visits for a certain period, the real date time of disengagement from care is unknown (Park et al., 2019). It is also worth mentioning that death censor loss to care but not vice versa. Accordingly, such studies involved the interval-censored semi-competing risks data context. The research deal with this setting considering the fact that the interval censoring phenomenon has been largely ignored in the literature dealing with competing or semi-competing risks data.

In medical research, HIV/AIDS study is one of the frequently interesting arising issues. In such follow-up studies, viral load and CD4 cell counts are major biomarkers of the severity of viral infection, disease progression, and treatment evaluation. The important role of CD4 covariates in longitudinal analysis arises from the fact that the number of CD4 cells of an individual gives a general degree of the well-being of his/her immune system and is a great measurement feature of immunosuppression. An ordinary CD4 cell is more than 500 cells per cubic millimeter (mm 3) of blood. On the off chance that one incorporates a CD4 check of less than 300, one will be analyzed as having AIDS, therefore, CD4 check estimation is a critical list that gives a way of gaging the progression from HIV to obtained resistant insufficiency disorder (Helps) for prognostic purposes (Kim et al., 2019, 2020). Many studies have been conducted from different aspects containing quantile regression investigation from the Bayesian strategy (Zhang & Huang, 2020) and semiparametric point of view (Bakoyannis et al., 2017; Park et al., 2021) on the joint modeling of such data and time to event data. However, in real-world applications, patients in addition to HIV are also exposed to other influential factors that may affect their lifetime or any favorite time to the event. At the same time, it is obvious that the observed value of important covariates such as CD4 is only available for the first occurring cause. Thereby, it seems necessary to model these important covariates given each cause. To this end, the exact time of occurring cause should be predicted and consequently, a polynomial regression has been carried out on the favorites covariates with knowledge of other causes.

Despite the fact that most competing/semi-competing risks data emerges as interval censorship, there exist a few papers considering the estimation problem of cause-specific CIF/CIF in such a situation. For such data in incorporation with covariate information, in accordance with nonparametric approaches, Hudgens et al. (2001), Jewell et al. (2003) and Groeneboom et al. (2008) proposed non-parametric maximum likelihood estimators of cause-specific CIF. In the semiparametric point of view, Bakoyannis et al. (2017), Barrett et al. (2011)] and Li (2016) provide different estimations of cause-specific CIF. There exists a parametric estimation of this function, especially in recent years (Jeong & Fine, 2006*a*; Hudgens et al., 2014). For a further comprehensive discussion see (Collett, 2015, Chapter 12) and reference therein.

In the face of existing a majority of articles dealing with these concepts, there still can find a lack of predicting occurrence times. The novelty of presenting some prediction strategies on occurrence times will provide using order statistics and copula approaches. Considering the dependence structure between these times are completely new and the motivation is so strong due to the real-world data set. Predicting cause failures are also given utilizing the aforementioned methods and according to a real data set, the performances are examined. In what follows, not only the copula function for the occurrence times have been generally assumed but also the covariates vector are also supposed with any dimension. As a matter of fact, we didn't focus on covariate affection, and just the CIF and cause-specific CIF are calculated in accordance with the available sources. Thereafter, using these functions we provide our prediction issues for both occurrence times, cause of failures, and even the missing covariates.

The main goal of this study is to predict the exact time of occurrence of other causes and make a proper multiple regression on affective covariate information for prognosis and its value whenever each cause has happened. The dependency of failure time given each cause was generally assumed, meaning that these variables can be independently considered or even following any dependency model. It is further assumed that the statistical distribution of these variables has not necessarily been identical. The copula function has been employed for constructing such a model, which will be discussed later. On the other hand, clinical visits are done in an ordered time form, and consequently, the concept of order statistics is also applied to the introduced prediction procedures.

In what follows from this study, the required relations for prediction problems are provided in section 2. Time of causes is predicted in section 3 for kinds of situations dealing with joint occurrences of multiple longitudinal and competing or semi competing risks data. In section 4, the values of covariate are predicted given the time of happening event is previously predicted. The pseudo.HIV.long data that are accessible in the R package `intccr` is deeply analyzed and the prediction problems of this data are investigated in section 5. Finally, the conclusion of our study is presented in the last section.

2. Prerequisites of Prediction

The present section not only reviews the required concepts but also provides some useful lemmas, that can be very helpful during the next section.

2.1. Required Relations

As mentioned previously, the present study intends to investigate the predicted missing values of interval censored multistate competing risks data according to the order of occurred upper bounds in each individual. Meaning that immediately after occurred upper bound for each individual the predicted value of missing competing risks data must be provided. It is logical to expect that some theoretical formulas

related to the order statistics can be needed. These formulas are presented in this subsection.

The next Lemma gives Cumulative Distribution Function (CDF) and Probability Distribution Function (PDF) of first order statistics arising from Independent and not Necessarily Identical Distributed (INID) random variables. The first order of statistics plays an important role in the competing risks area. In fact, the first occurring competing risks is the observation value of first order statistics among all involved risks.

Lemma 1. *Assume that $X_1, X_2, \ldots, X_n$ be an INID random variable with corresponding CDFs $F_1, F_2, \ldots, F_n$ respectively and $X_{1:n}$ denote the first order statistics among the aforementioned samples, thus*

$$F_{X_{1:n}}(t) = 1 - \prod_{j=1}^{n}(1 - F_j(t))$$

$$f_{X_{1:n}}(t) = \sum_{j=1}^{n} f_j(t)[\prod_{\substack{i=1,\\ i\neq j}}^{n}(1 - F_i(t))]$$

Proof. See David & Nagaraja (2004, page 106). □

Since competing risks are one type of time to an event data, they are observed in an ordered form. Therefore, the joint PDF of the ordered form of these variables is needed aiming to predict problems that are extensively discussed in future subsections.

The Lemma 2, provides the joint pdf of order statistics arising from not Necessarily Independent and not Necessarily Identical Distributed (NINID) random variables.

Lemma 2. *If $\underline{X} \sim f_{\underline{X}}(\underline{x})$ then*

$$f_{X_{1:n},X_{2:n},\ldots,X_{n:n}}(x_1, x_2, \ldots, x_n) = \sum_{\pi \in S_n} f_{\pi(1),\pi(2),\ldots,\pi(n)}(x_1, x_2, \ldots, x_n)$$

where

$$S_n = \{(i_1, i_2, \ldots, i_n) \in \{1, 2, \ldots, n\}^n | i_j \neq i_k, j \neq k, 1 \leq j, k \leq n\}$$

denotes the set of all permutations π of $(1, 2, \ldots, n)$.

Proof. See Hogg & Craig (1978, section 4.6), David & Nagaraja (2004, section 5.3) and Balakrishnan & Cramer (2008). □

Because the value of this extra variable is observed during the experiment, it can be considered independently of the rest of the involved risks and moreover with a degenerated PDF.

In the Remark 1, the joint PDF of an ordered form of a random vector containing degenerated and a vector of continuous random variables are presented.

***Remark* 1.** If Y is a degenerated random variable at real point c and $\underline{X} \sim f_{\underline{X}}(\underline{x})$. Moreover assume that $Z = (Y, \underline{X})$, then

$$g_{Z_{1:n}, Z_{2:n}, \ldots, Z_{n+1:n+1}}(z_1, z_2, \ldots, z_{n+1}) = \sum_{i=1}^{n+1} I(Z_i = c) f(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_{n+1})$$

where f be the same function as in Lemma 2 and $I(.)$ stands for the indicator function.

***Proof*.** Use the independence features of Y and $\underline{X}$ and apply Lemma 2. □

Since in the present study, the multistate competing risk is considered, the joint CDF of these variables plays an important role in statistical inference based on the observed value of these risks. Here, the role of the copula function is highlighted and without loss of generality, one copula function has been considered to connect the CIF of competing risks.

Theorem 1. *For a d-variate random vector $\underline{Y} = (Y_1, Y_2, \ldots, Y_d)$ associated with joint CDF F, there exist a copula function $C : [0,1]^d \to [0,1]$ such that*

$$\begin{aligned} &P(Y_1 \leq y_1, Y_2 \leq y_2, \ldots, Y_d \leq y_d) \\ &= F_{Y_1, Y_2, \ldots, Y_d}(y_1, y_2, \ldots, y_d) \\ &= C(P(Y_1 \leq y_1), P(Y_1 \leq y_1), \ldots, P(Y_d \leq y_d)) \\ &= C(F_{Y_1}(y_1), F_{Y_2}(y_2), \ldots, F_{Y_d}(y_d)) \end{aligned}$$

where copula function C satisfies in the following analytical terms

I: $C(u_1, u_2, \ldots, u_d) = 0$ *if at least one* $u_j = 0$.

II: $C(1, 1, \ldots, 1, u_j, 1, \ldots, 1, 1) = u_j$ *if at most one* $u_j \neq 1$.

III:

$$\int_B dC(u) = \sum_{\mathbf{z} \in \times_{i=1}^{d} \{x_i, y_i\}} (-1)^{N(\mathbf{z})} C(\mathbf{z}) \geq 0,$$

***Proof*.** See Joe (2014, pages 7, 8 and 27). □

As can be seen from the Theorem 1, the copula function can connect the joint CDF of multistate competing risks with the marginal CIF of every risk.

3. Time of Causes Predictions

Suppose to n individuals be in a study which some of their indexes repeatedly are measured until an event has occurred. This event has one of k risks which every individual deal with. On the other hand, the exact time of occurring each of these risks can not be observed by an operator or tester and just the interval of in which each of the risks has been caused is available. Mark the observed interval of the i-th individual with $(l_i, u_i), 0 \leq l_i \leq u_i \leq \infty$. The situation $u_j = \infty$ indicate that j-th individual is right censored and consequently the related risk are not observed that results in $c_j = 0$ In addition, consider $c_0 = 0, c_1 = 1, c_2 = 2, \ldots, c_k = k$ as the $k + 1$ causes of each individual. Since we are going to predict the values of unobserved risks based on the timing of caused one risks for each individual, so set $l_i^* = u_i I(u_i \neq \infty) + l_i I(u_i = \infty)$. Consider $F_j(t, \underline{Z}), j = 1, 2, \ldots, k$ as the CIF of each risks at the time t subject to covariate vector $\underline{Z}$ that represent repeatedly measured affection covariate on each risks. Theorem 1, guarantee that there exists a function C that

$$F_{T_{c_1}, T_{c_2}, \ldots, T_{c_k}|\underline{Z}}(t_1, t_2, \ldots, t_k) = C(F_{T_{c_1}}(t_1, \underline{Z}), F_{T_{c_2}}(t_2, \underline{Z}), \ldots, F_{T_{c_k}}(t_k, \underline{Z}))$$

which represent the joint CIF of k multistate competing risks and where T_{c_j}'s stand for all possible competitive events time. Thence, if l_i^* considered as a degenerated random variable, then by Remark 1, the ordered form of joint pdf of l_i^* and random vector $(T_{c_1}, T_{c_2}, \ldots, T_{c_k})$ is given by

$$g_{T^*_{1:n}, T^*_{2:n}, \ldots, T^*_{n+1:n+1}}(t_1^*, t_2^*, \ldots, t_{n+1}^*) = \sum_{i=1}^{n+1} I(Z_i = c) f(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_{n+1}),$$

where c is a constant. In fact, in this case, we are interested in modeling the joint PDF of a vector and a constant c that frequently occurs in both left and right censoring. This is the main function that can be useful in predicting competing risks data which will be discussed in details in the next section.

According to two previous subsections, we provide strategies of predicting missing values of interval censored multistate competing risks data. These strategies consider the following assumptions.

- A general dependency between continuous competing risks has been considered. i.e., the prediction problem has not been constructed based on an assumption for joint CDFs of multistate competing risks. It is clear to understand that this assumption is delivered to the operator's desires.
- The right censoring case has been considered as an special case of interval censoring scheme where infinite value is assumed for its upper bound. The simplest competing risks mode when a cause has occurred in a specific point is also a particular case of interval censored competing risks such that lower and upper bounds for any case are equally assumed.
- In addition to the measured and known covariate that affects the observing competing risk, the time which operator can predict for an individual has

been considered as an unmeasured and unknown covariate that may affect the observed value of competing risks data.

Here, suppose the interval censored multistate competing risks data from n individuals are collected. For the i-th individual consider $l_i, u_i, c_i, \underline{Z}_i$ as its lower and upper bounds, observed competing risks, and affection covariate on observing values of competing risk. In order to apply the affection of time on the observed values of competing risks for any individuals, the prediction strategy should be done based on the order of observing upper bounds. In other words, for two individuals the prediction problems must consider the order of their upper bounds due to the point that it is logical to expect that the prediction for individuals with less upper bounds should be done earlier rather than who has greater upper bounds. For the individuals who have been right censored, the decision point for prediction is their lower bounds.

To predict the missing values in interval censored multistate competing risks data, one of three situations may occur. These terms are as follows.

- Right Censoring Occurred

  Right censoring in competing risks data happens when before any risk occurs, the individual withdraws from the test. In fact, these type of censoring is an especial case of interval censoring scheme where its upper bounds are infinite. Assume that $T_{c_1}, T_{c_2}, \ldots, T_{c_k}$ be corresponding random variables due to time events of k multistate competing risk for an individual whom to the right censoring scheme at time l_i. It is clear to understand $T_{c_j} \geq l_i, j = 1, 2, \ldots, k$. However, the right censoring may also happen in other individuals. Considering the same dependent structure of multistate competing risk, if for every right censored individuals at time l_i, we can suppose to the lower bounds L_i is itself a degenerated random variables at the point l_i and moreover is independent of the random vector $T_{c_1}, T_{c_2}, \ldots, T_{c_k}$. Therefore assume that the new random vector $U = (T_{c_1}, T_{c_2}, \ldots, T_{c_k}, L_i)$ as the combined of censoring time and multistate competing risks. Hereafter setting $U_{1:k+1}, U_{2:k+1}, \ldots, U_{k+1:k+1}$ which represent the ordered form of random vector $\underline{U}$, and calculating the conditional densities, the predicted values of competing risks data can be suggested as expectation of these conditional PDFs, i.e., $\hat{T}_{c_j} = E[U_j | U_{1:k+1} = l_i], j = 1, 2, \ldots, k$. In the case, that competing risks data is naturally observed in an ordered form, such as prognosis of a disease, the predicted values of these missing data can be provided as $\hat{T}_{c_j} = E[U_{j+1:k+1} | U_{1:k+1} = l_i], j = 1, 2, \ldots, k$. Another prediction based on Markov chain idea that can be used is

  $$\hat{T}_{c_j} = E[U_{j+1:k+1} | U_{1:k+1} = l_i, U_{2:k+1} = \hat{T}_{c_2}, \\ U_{3:k+1} = \hat{T}_{c_3}, \ldots, U_{j-1:k+1} = \hat{T}_{c_{j-1}}],$$

  for any individuals indexed by $j = 1, 2, \ldots, k$.

- Right Censoring Does Not Occurred

The prediction method for the case that right censoring happens, are discussed in the previous item. Here, the problem of prediction missing values in interval censored competing risks data for the cases that right censored does not happens are investigated. In such cases, two different situations may happen. These two modes are the competing risks happens and don't happen. The next two items deal with the prediction problems in these modes.

- ◇ Cause Occurred

  This item will pay attention to the prediction problem when right censoring does not happen and one cause has occurred in the interval. Assume that $\underline{T} = T_{c_1}, T_{c_2}, \ldots, T_{c_k}$ be corresponding random variables due to time events of k multistate competing risk for an individual that minimum of them has occurred in the interval l_i, u_i. Since the minimum value of these variables are observed, thence set a new random vector $T_{1:k}, T_{2:k}, \ldots, T_{k:k}$ as the ordered form of random vector T. Therefore the prediction of event time for the j-th ($j = 1, 2, \ldots, k$), can be considered as $\hat{T}_{c_j} = E[T_j | l_i \leq T_{1:k} \leq u_i], j = 1, 2, \ldots, k$. In the case, that competing risks data is naturally observed in an ordered form, such as prognosis of a disease, the predicted values of these missing data can be provided as $\hat{T}_{c_j} = E[T_{j:k} | l_i \leq T_{1:k} \leq u_i], j = 1, 2, \ldots, k$.

- ◇ None Cause Has Been Occurred

  The last item that remains shall be investigated in this subsection. The procedures behind the prediction strategy in this situation is similar to the first item when the right censoring has occurred but here, the random which we sure less than all of competing risks seems to be a continuous random variable supported in the given interval. Regarding the aforementioned phrase, assume that random vector $\underline{T} = T_{c_1}, T_{c_2}, \ldots, T_{c_k}$ be representations of event time of competing risks that do not happen yet. Furthermore, it looks like that additional random variable happening in the interval follows a uniform distribution on the corresponding given interval. Mark this variable with notation T', and consider the random vector $\underline{U} = (T_{c_1}, T_{c_2}, \ldots, T_{c_k}, T')$ as a combination of random vector $\underline{T}$ and random variable T'. Finally, as we had in the previous two items if $U_{1:k+1}, U_{2:k+1}, \ldots, U_{k+1:k+1}$ be an ordered form of $\underline{U}$, the prediction of missing competing risks data in such cases can be given as $\hat{T}_{c_j} = E[U_j | l_i \leq U_{1:k} \leq u_i], j = 1, 2, \ldots, k$.

When faced with missing values or masked causes of failure, we should first predict these values. Then, the problem of predicting corresponding covariates will arise. For instance, after forecasting the time of an event for an individual, it is necessary to provide the corresponding covariate values for that predicted time. This can also be done for unobserved causes or with the assumption of another cause occurring. What can we say about the covariate values? To this end, we provide some methods and investigate their features.

4. Pseudo.HIV.long Data Analysis: Independent Case

Table 1: Prediction causes (CA) for randomly selected patients (PA)

PA	30	79	93	110	118	122	161	165	167
CA-TRUE	1	0	0	0	0	0	0	1	1
CA-PREDICTED	1	1	0	0	1	0	0	0	1

Table 2: Prediction causes (CA) for randomly selected patients (PA)

PA	24	31	48	57	68	99	102	149	172
CA-TRUE	0	0	0	0	0	0	0	1	0
CA-PREDICTED	0	0	0	0	0	0	1	1	1

Table 3: Prediction causes (CA) for randomly selected patients (PA)

PA	27	35	70	76	78	101	168	171	200
CA-TRUE	2	0	0	1	0	0	1	0	0
CA-PREDICTED	2	2	0	1	0	0	1	0	0

Table 4: Prediction causes (CA) for randomly selected patients (PA)

PA	32	40	47	75	137	146	159	194	200
CA-TRUE	0	1	0	0	0	0	0	0	0
CA-PREDICTED	0	1	0	0	0	0	0	0	0

Table 5: Prediction causes (CA) for randomly selected patients (PA)

PA	14	29	55	121	154	166	179	184	195
CA-TRUE	0	1	2	0	0	0	0	0	2
CA-PREDICTED	1	0	2	2	0	0	0	0	2

The methodologies are going to apply in a real data set. Among the majority of suitable studies, investigations around HIV have their own attention for many scholars. There are some relating projects that are famous in applicational aspects. Focusing on HIV topics, Bakoyannis et al. (2017) present a real data set collected in a region of Africa, and afterward, this data set will be extensively discussed by many researchers (Park et al., 2019; Mitra et al., 2020; Curnow et al., 2021; Eleuteri et al., 2021; Emura et al., 2020). The data was gathered by the East Africa International Epidemiologic Databases to Evaluate AIDS Regional Consortium, an HIV care network, and treatment programs working in Kenya, Uganda, and Tanzania countries. They also provide a corresponding simulational data set called `pseudo.HIV.long` or equivalently `Artificial HIV dataset` in package `intccr` of R statistical software. The constructing method was discussed in Park et al. (2019) in detail and we are going to utilize this data set to provide our prediction results.

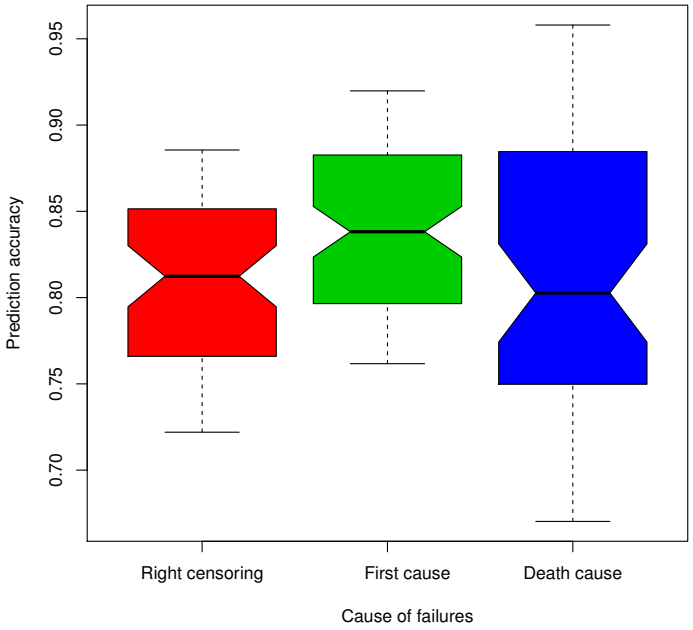

FIGURE 1: Boxplot for the accuracy of cause prediction in the case of right censoring, first cause and death time

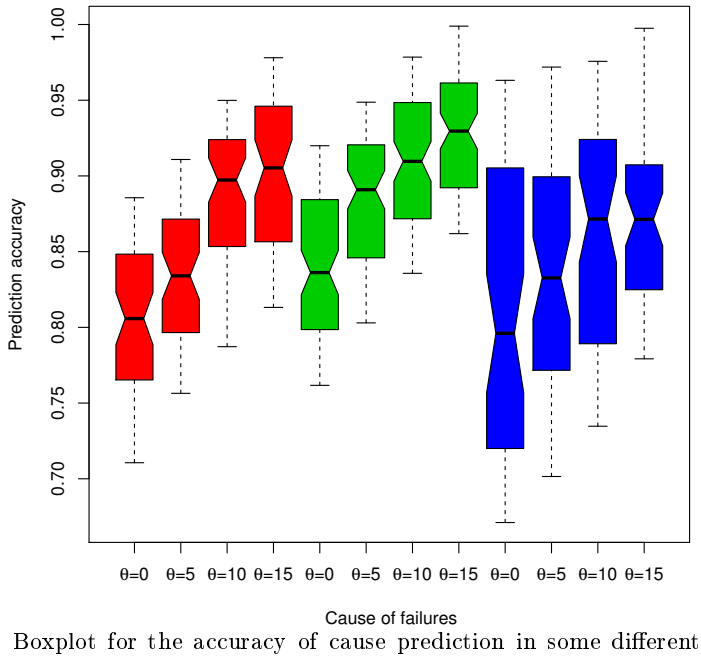

FIGURE 2: Boxplot for the accuracy of cause prediction in some different copula assumption

The artificial dataset that was simulated to resemble the HIV study on loss to HIV care and death in subSaharan Africa (Park et al., 2019). It contains subjects individual identification number (id), observation times (t), competitive events (c), and important covariates including male, age, and the number of cd4 cells for 3000 patients in 22710 rows showing all visiting processes. The `simdata` contains id, the last time point prior to the event (v), the first time point after the event (u), competitive events (c), and related covariates.

This data set is also an excellent showcase in real-world applications since it covers three competitive events and both right and left censored data that repeatedly occur in clinical trials. Left censoring occurs when competitive event 2 happens in the first clinical visit like as patient 7, and consequently, there is no useful data for those patients. Right censoring occurs when competitive event 0 is observed, which means that before occurring any failures 1 or 2, the patient leaves the process like as patients 1, 3, 4, and so on. The competitive event 1 also can be observed for the patients who lost their care.

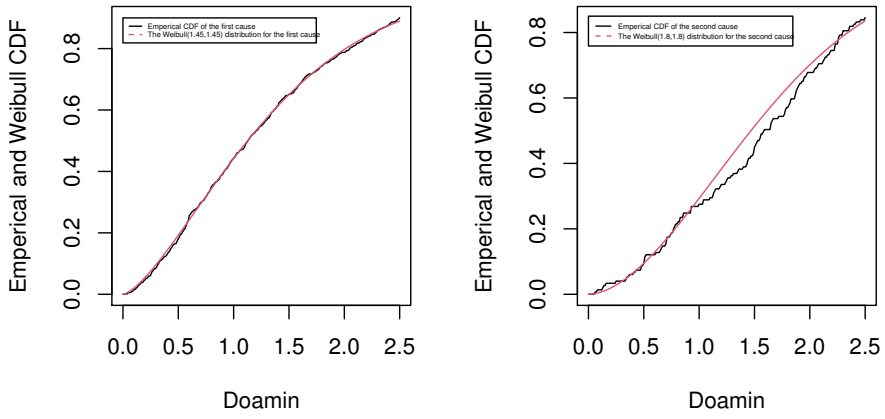

FIGURE 3: The performances of the Weibull distribution for the fitting cause of failures (first cause (left) and second cause (right))

In this case, we deal with joint occurrences of multiple longitudinal data of dimension 3 with semi-competing risks data, since competitive event 2 can not happen after competitive event 1 and it is the fatal competitive events or fatal shock. Hereafter it goes without saying that the prediction issues are so important for the competitive event 2. Accordingly, our main problem here is to predict the time of occurrences for competitive event 2 when other competitive events have happened. However, it may be interesting to provide the prediction of covariate values at that predicted time too. For instance, patient 1 was right-censored at the time 2.9, and we are looking for the time of his death in preparing clinical service. We also find the number of cd4 cells of that patient for that time.

Table 6: Predicted death time (PDT) for all 200 Patients (PA)

PA	PDT	PA	PDT	PA	PDT	PA	PDT	PA	PDT
1	3.313	41	2.912	81	0.751	121	3.216	161	2.915
2	-	42	3.012	82	3.256	122	3.714	162	2.975
3	2.612	43	2.991	83	3.012	123	2.925	163	2.524
4	3.214	44	2.975	84	2.201	124	2.912	164	2.925
5	0.975	45	3.724	85	2.012	125	2.013	165	0.726
6	3.003	46	2.974	86	3.111	126	2.802	166	3.012
7	-	47	3.207	87	-	127	2.725	167	0.973
8	2.925	48	3.314	88	3.524	128	1.013	168	1.914
9	3.127	49	3.428	89	3.317	129	3.526	169	2.914
10	2.817	50	2.991	90	2.903	130	2.983	170	2.524
11	2.873	51	2.504	91	-	131	0.985	171	2.714
12	3.483	52	3.673	92	3.712	132	2.604	172	3.519
13	2.778	53	2.997	93	0.813	133	3.011	173	2.714
14	2.871	54	2.915	94	0.712	134	2.912	174	-
15	1.017	55	-	95	2.877	135	3.017	175	2.914
16	2.973	56	2.896	96	2.614	136	2.712	176	-
17	2.901	57	3.212	97	-	137	3.114	177	3.017
18	2.313	58	2.987	98	1.812	138	2.814	178	0.974
19	2.221	59	2.813	99	1.993	139	3.011	179	3.019
20	0.712	60	2.718	100	3.015	140	2.914	180	3.214
21	1.725	61	2.903	101	2.923	141	3.214	181	1.188
22	2.011	62	2.903	102	2.904	142	3.314	182	0.905
23	0.724	63	2.814	103	0.846	143	2.826	183	3.014
24	3.114	64	2.914	104	2.913	144	2.817	184	2.546
25	3.336	65	3.114	105	3.524	145	3.624	185	3.011
26	2.719	66	3.001	106	2.224	146	2.924	186	1.715
27	-	67	0.998	107	1.381	147	1.607	187	-
28	-	68	2.918	108	3.013	148	2.514	188	2.718
29	0.725	69	3.214	109	3.105	149	1.907	189	3.007
30	2.214	70	3.234	110	2.518	150	0.991	190	3.303
31	2.718	71	2.819	111	0.381	151	2.912	191	2.917
32	2.708	72	1.724	112	0.792	152	2.524	192	3.012
33	2.997	73	2.405	113	2.308	153	2.876	193	3.117
34	1.803	74	2.703	114	3.117	154	2.905	194	2.908
35	3.214	75	3.307	115	2.907	155	2.613	195	-
36	-	76	0.718	116	2.207	156	2.998	196	3.212
37	3.110	77	3.111	117	0.656	157	3.607	197	2.714
38	3.507	78	3.302	118	2.907	158	0.514	198	3.129
39	-	79	1.817	119	2.873	159	3.012	199	-
40	1.714	80	2.974	120	3.213	160	2.687	200	3.119

In this regard, the independence assumption for the competitive events 1 and 2 is reasonable. Moreover, the most-famous distribution that can be used for modeling time to event data is the Weibull density. This model is widely utilized in modeling time to event data (Fine, 2001), mortality data (Li, 2016), cancer

data, competing and semi-copmeting risks data (Bakoyannis et al., 2017; Park et al., 2019), and also interval censoring data (Jeong & Fine, 2006*b*; Hudgens et al., 2014). Here, we have semi-competing interval censoring data, and the aim is to find maximum likelihood estimators for corresponding Weibull distributions. The Weibull density with shape parameter α, and scale parameter β, has the form:

$$F_{T_{c_j}}(t) = 1 - e^{-(\frac{t}{\beta_j})^{\alpha_j}}, j = 1, 2,$$

and we can set the likelihood function as:

$$L(\alpha_1, \beta_1, \alpha_2, \beta_2) = \prod_{i \in S} \prod_{j=1}^{2} [F_{T_{c_j}}(t) - F_{T_{c_j}}(t)]^{\delta_j}, \tag{1}$$

where S stands for the subset of "newpseudo" that only competitive events 1 and 2 have occurred, and

$$\delta_j = \begin{cases} 1 & c = j \\ 0 & c = 3 - j \end{cases}$$

Now, utilizing the following codes, it can go without saying that our estimators are $\hat{\alpha_1} = 3.135, \hat{\beta_1} = 1.786$, and $\hat{\alpha_2} = 2.798, \hat{\beta_2} = 2.013$.

Fitted Weibull distributions for both competitive events have been carried out by Figure 3. Accordingly, it is straightforward that the aforementioned Weibull densities are good choices for the data set, but there exist some alternative choices including cause-specific CIF provided by other scholars like as Bakoyannis et al. (2017); Emura et al. (2020); Mitra et al. (2020); Jeong & Fine (2006*a*); Fine et al. (2001); Li (2016). In the case of the dependent competitive events, we can use 3 famous copulas covering three kinds of dependencies including negative, positive, and weak correlations. It is easy to check that Gumbel, FGM, and Clayton copulas can include all dependency situations. In addition, we also consider the normal copula as an important model. Accordingly, we should assess the performances of these copulas based on a given data set, and then a copula model is determined as a suitable model for the corresponding variables. Taking a good approach for our selection, we can use the helpful command `gof` in the useful package `gofCopula` from statistical software R, computing for a given dataset and according to the choices of the user different tests for different copulae. For more information see Joe (2014).

It is also worth mentioning that, the likelihood function (1) should be re-calculated in the case of the dependent competitive events. In the following we have:

$$
\begin{aligned}
L(\alpha_1,\beta_1,\alpha_2,\beta_2) &= \prod_{i\in S}\prod_{j=1}^{2}\Big[P(l_{c_1}\le T_{c_1}\le u_{c_1}, l_{c_2}\le T_{c_2}\le u_{c_2})\Big]\\
&= \prod_{i\in S}\prod_{j=1}^{2}\Big[P(T_{c_1}\le u_{c_1}, T_{c_2}\le u_{c_2})\\
&\quad + P(T_{c_1}\le l_{c_1}, T_{c_2}\le l_{c_2}) - P(T_{c_1}\le u_{c_1}, T_{c_2}\le l_{c_2})\\
&\quad - P(T_{c_1}\le l_{c_1}, T_{c_2}\le u_{c_2})\Big]\\
&= \prod_{i\in S}\prod_{j=1}^{2}\Big[C(F_{T_{c_1}}(u_{c_1}), F_{T_{c_2}}(u_{c_2}))\\
&\quad + C(F_{T_{c_1}}(l_{c_1}), F_{T_{c_2}}(l_{c_2})) - C(F_{T_{c_1}}(u_{c_1}), F_{T_{c_2}}(l_{c_2}))\\
&\quad - C(F_{T_{c_1}}(l_{c_1}), F_{T_{c_2}}(u_{c_2}))\Big],
\end{aligned}
$$

where S stands as the same as relation (1) and $C(.)$ is our selected copula function.

In this regard, to our best knowledge about the performances and accuracies of provided methods, we present three simulation strategies in continue. First of all, for 200 patients, their death times have been predicted in Table 6. Clearly, it didn't require the patients who have died or experienced the second competitive events. In fact, in all similar cases that we are dealing with semi-competing risks data, there exists a fatal competitive events (in this case it is patient death time or competitive event 2) and we are interested to predict that competitive event. For other patients that are right-censored (competitive event 0) or lose their care (competitive event 1), knowing their predicted death time can help us to provide some extra acquirements like increasing visits number for the patients with small predicted time or precisely optimizing visiting number for those individuals. For instance, both patients 5 and 8 experienced first competitive events and their prediction times are $(0.975, 2.925)$. Obviously patient 5 need more help, and acquirements.

The data set also can have another perspective including missing the competitive events. It frequently happens during a study, some information has been missed such as the competitive events. The prediction way can initially be given by:

$$
\begin{cases}
0 & |p_0(t)-t| \le \min(|p_1(t)-t|, |p_2(t)-t|)\\
1 & |p_1(t)-t| \le \min(|p_0(t)-t|, |p_2(t)-t|)\\
2 & |p_2(t)-t| \le \min(|p_0(t)-t|, |p_1(t)-t|)
\end{cases}
$$

where $p_i(t)$ and t respectively stand for the prediction of occurrence time based on assumptions of happening competitive event i and observed value during the experiment. The relation was clear and it is easy to understand that its basis is related to time prediction. The predicted competitive events referred to the closest predicted time. Hereafter, we randomly choose 10 patients among the data set in five stages. For all patients, the competitive events is assumed to be missed and then, the prediction method was applied. True and predicted competitive

events are tabulated for every stage respectively in Tables 1-5. The prediction accuracy is around 82 percent on average and it shows that the method has suitable performance.

The last issue is dependence discussion. If the competitive events is set to be dependent, it can be directly seen that for this case study, we have a positive dependency, since increasing behavior of non-fatal competitive events (0 and 1) results in the same appearance of fatal competitive events. Accordingly, we assume that the competitive events following the Clayton copula provides a proper covering for such dependency (for more studies see Joe, 2014). The Clayton copula has the form:

$$C_C(u_1, u_2, \ldots, u_n) = [1 + \sum_{i=1}^{n} (u_i^{-\theta} - 1)]^{-\frac{1}{\theta}}.$$

where $\theta \geq 1, \theta \neq 0$, and $\theta \to 0$, indicated for independent copula. Using the package `copula` and corresponding command `claytonCopula` we simulate 1000 items and replicate them 100 times for similar competitive events considering the same Weibull distribution and $\theta = 5$ as a strong dependency. The accuracy percentages are available in Figure 2 and again the suitable performances of the prediction issues are clear.

5. Conclusion

The study considers joint occurrences of multiple longitudinal and competing or semi-competing risks data with some covariate values observed during the experiment. Time to event data is considered to be observed in the interval censoring because this assumption not only includes the right censoring but also covers the common case of competing or semi-competing risks data. The related cases are described in detail. In these situations, there are many values that are don't observe like competitive events and longitudinal measures. In addition, it is obvious that the corresponding covariate values for the masked competitive events have still been unknown. Considering the general dependency among the lifetime distributions, the prediction strategies for the aforementioned missed values and masked competitive events are provided. Some of the prediction advantageous features are investigated and for a real data set, the prediction methods are applied. The real data set is well analyzed and corresponding tables and figures are also presented.

[Received: August 2023 — Accepted: March 2024]

References

Aalen, O. O. & Johansen, S. (1978), 'An empirical transition matrix for non-homogeneous markov chains based on censored observations', *Scandinavian Journal of Statistics* pp. 141–150.

Bakoyannis, G., Yu, M. & Yiannoutsos, C. T. (2017), 'Semiparametric regression on cumulative incidence function with interval-censored competing risks data', *Statistics in medicine* **36**(23), 3683–3707.

Balakrishnan, N. & Cramer, E. (2008), 'Progressive censoring from heterogeneous distributions with applications to robustness', *Annals of the Institute of Statistical Mathematics* **60**(1), 151–171.

Barrett, J. K., Siannis, F. & Farewell, V. T. (2011), 'A semi-competing risks model for data with interval-censoring and informative observation: An application to the mrc cognitive function and ageing study', *Statistics in medicine* **30**(1), 1–10.

Berry, S. D., Ngo, L., Samelson, E. J. & Kiel, D. P. (2010), 'Competing risk of death: an important consideration in studies of older adults', *Journal of the American Geriatrics Society* **58**(4), 783–787.

Bhattacharjee, A. (2019), 'A joint longitudinal and survival model for dynamic treatment regimes in presence of competing risk analysis', *Clinical Epidemiology and Global Health* **7**(3), 337–341.

Collett, D. (2015), *Modelling survival data in medical research*, Chapman and Hall/CRC.

Curnow, E., Hughes, R. A., Birnie, K., Crowther, M. J., May, M. T. & Tilling, K. (2021), 'Multiple imputation strategies for a bounded outcome variable in a competing risks analysis', *Statistics in Medicine* **40**(8), 1917–1929.

David, H. A. & Nagaraja, H. N. (2004), 'Order statistics', *Encyclopedia of Statistical Sciences* .

Eleuteri, A., Rola, A. C., Kalirai, H., Hussain, R., Sacco, J., Damato, B. E., Heimann, H., Coupland, S. E. & Taktak, A. F. (2021), 'Cost-utility analysis of a decade of liver screening for metastases using the liverpool uveal melanoma prognosticator online (lumpo)', *Computers in Biology and Medicine* **130**, 104221.

Emura, T., Shih, J.-H., Ha, I. D. & Wilke, R. A. (2020), 'Comparison of the marginal hazard model and the sub-distribution hazard model for competing risks under an assumed copula', *Statistical methods in medical research* **29**(8), 2307–2327.

Fine, J. P. (2001), 'Regression modeling of competing crude failure probabilities', *Biostatistics* **2**(1), 85–97.

Fine, J. P. & Gray, R. J. (1999), 'A proportional hazards model for the subdistribution of a competing risk', *Journal of the American Statistical Association* **94**(446), 496–509.

Fine, J. P., Jiang, H. & Chappell, R. (2001), 'On semi-competing risks data', *Biometrika* **88**(4), 907–919.

Groeneboom, P., Maathuis, M. H. & Wellner, J. A. (2008), 'Current status data with competing risks: Limiting distribution of the mle', *Annals of Statistics* **36**(3), 1064.

Haneuse, S. & Lee, K. H. (2016), 'Semi-competing risks data analysis: accounting for death as a competing risk when the outcome of interest is nonterminal', *Circulation: Cardiovascular Quality and Outcomes* **9**(3), 322–331.

Hickey, G. L., Philipson, P., Jorgensen, A. & Kolamunnage-Dona, R. (2018), 'A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **181**(4), 1105–1123.

Hogg, R. V. & Craig, A. T. (1978), 'Introduction to mathematical statistics'.

Hudgens, M. G., Li, C. & Fine, J. P. (2014), 'Parametric likelihood inference for interval censored competing risks data', *Biometrics* **70**(1), 1–9.

Hudgens, M. G., Satten, G. A. & Longini Jr, I. M. (2001), 'Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation', *Biometrics* **57**(1), 74–80.

Jeong, J.-H. & Fine, J. (2006*a*), 'Direct parametric inference for the cumulative incidence function', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **55**(2), 187–200.

Jeong, J.-H. & Fine, J. P. (2006*b*), 'Parametric regression on cumulative incidence function', *Biostatistics* **8**(2), 184–196.

Jeong, J.-H. & Oakes, D. (2003), 'On the asymptotic relative efficiency of estimates from cox's model', *Sankhyā: The Indian Journal of Statistics* pp. 422–439.

Jewell, N. P., van der Laan, M. & Henneman, T. (2003), 'Nonparametric estimation from current status data with competing risks', *Biometrika* **90**(1), 183–197.

Joe, H. (2014), *Dependence modeling with copulas*, Chapman and Hall/CRC.

Kazempoor, J., Habibirad, A. & Okhli, K. (2022), 'Bounds for cdfs of order statistics arising from inid random variables', *Journal of the Iranian Statistical Society* **19**(1), 39–57.

Kim, J., Kim, J. & Kim, S. W. (2019), 'Additive-multiplicative hazards regression models for interval-censored semi-competing risks data with missing intermediate events', *BMC medical research methodology* **19**(1), 49.

Kim, J., Kim, J. & Kim, S. W. (2020), 'Regression analysis of the illness-death model with a shared frailty when all transition times are interval censored', *Communications in Statistics-Simulation and Computation* **52**(1), 1–16.

Li, C. (2016), 'The fine-gray model under interval censored competing risks data', *Journal of Multivariate Analysis* **143**, 327–344.

Mao, L. & Lin, D. (2017), 'Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**(2), 573–587.

Meng, X.-L. (1994), 'Multiple-imputation inferences with uncongenial sources of input', *Statistical Science* pp. 538–558.

Mitra, D., Das, U. & Das, K. (2020), 'Analysis of interval-censored competing risks data under missing causes', *Journal of Applied Statistics* **47**(3), 439–459.

Moharib Alsarray, R. M., Kazempoor, J. & Ahmadi Nadi, A. (2023), 'Monitoring the weibull shape parameter under progressive censoring in presence of independent competing risks', *Journal of Applied Statistics* **50**(4), 945–962.

Park, J., Bakoyannis, G. & Yiannoutsos, C. T. (2019), 'Semiparametric competing risks regression under interval censoring using the r package intccr', *Computer Methods and Programs in Biomedicine* **173**, 167–176.

Park, J., Bakoyannis, G., Zhang, Y. & Yiannoutsos, C. T. (2021), 'Semiparametric regression on cumulative incidence function with interval-censored competing risks data and missing event types', *Biostatistics* .

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T. & Breslow, N. E. (1978), 'The analysis of failure times in the presence of competing risks', *Biometrics* pp. 541–554.

Putter, H., Fiocco, M. & Geskus, R. B. (2007), 'Tutorial in biostatistics: competing risks and multi-state models', *Statistics in medicine* **26**(11), 2389–2430.

Rouanet, A., Joly, P., Dartigues, J.-F., Proust-Lima, C. & Jacqmin-Gadda, H. (2015), 'Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia', *arXiv preprint arXiv:1506.07415* .

Williamson, P., Kolamunnage-Dona, R., Philipson, P. & Marson, A. (2008), 'Joint modelling of longitudinal and competing risks data', *Statistics in Medicine* **27**(30), 6426–6438.

Zhang, H. & Huang, Y. (2020), 'Quantile regression-based bayesian joint modeling analysis of longitudinal-survival data, with application to an aids cohort study', *Lifetime Data Analysis* **26**(2), 339–368.