

Modeling Data With Semicompeting Risks: An Application to Chronic Kidney Disease in Colombia

Modelo para datos con riesgos semi competitivos: una aplicación a insuficiencia renal crónica en Colombia

ELIZABETH GONZÁLEZ PATIÑO^{1,a}, GISELA TUNES^{1,b},
MARIA ISABEL MUNERA^{2,c}

¹DEPARTAMENTO DE ESTADÍSTICA, INSTITUTO DE MATEMÁTICA E ESTATÍSTICA,
UNIVERSIDADE DE SÃO PAULO, SÃO PAULO, BRAZIL

²LABORATÓRIO CLÍNICO, HOSPITAL PABLO TOBÓN URIBE, MEDELLÍN, COLOMBIA

Abstract

In this paper, the structure of semicompeting risks data, defined by Fine, Jiang & Chappell (2001), is studied. Two events are of interest: a non-terminal and a terminal event, the last one, can censor the non-terminal event, but not vice versa. Due to the possible dependence between the times until the occurrence of such events, two approaches are evaluated: modelling the bivariate survival function through Archimedean copulas and a shared frailty model. A simulation is conducted to examine its performance and both approaches are applied to a real data set of patients with chronic kidney disease (CKD).

Key words: Archimedean Copula; Frailty model; Semicompeting risks; Survival.

Resumen

En este trabajo se estudia la estructura de datos con riesgos semicompetitivos definida por Fine et al. (2001). En esta estructura existen dos eventos de interés; uno intermedio y otro terminal, este último puede censurar el evento intermedio, pero no viceversa. Dada la posible dependencia, entre los tiempos hasta la ocurrencia de tales eventos, dos tipos de enfoques son evaluados: uno, modelando la función de supervivencia bivariada a través de cópulas Arquimedianas y el otro por medio de un modelo con fragilidad

^aPhD student. E-mail: lizapat@ime.usp.br

^bPhD. E-mail: tunes@ime.usp.br

^cMd Microbiologist. E-mail: mmunera@hptu.org.co

compartida. El desempeño se observa a través de simulación. Ambos enfoques son aplicados a un conjunto de datos reales de pacientes con insuficiencia renal crónica, que pueden o no presentar una recaída de la enfermedad luego de ser tratados con diálisis.

Palabras clave: Cópulas arquimedianas; Modelo de fragilidad; Supervivencia; Riesgos semicompetitivos.

1. Introduction

Currently, chronic kidney disease (CKD) is considered a worldwide problem and a great increase of deaths caused by CKD has been observed from 1990 to 2010 (Martín-Cleary & Ortiz 2014). In Colombia, its prevalence has increased (Gómez 2006) and, according to Fondo Colombiano de Enfermedades de Alto Costo (Cuenta del Alto Costo 2014), 975.479 people had CKD in 2013. In December 1993, the social security system was reformulated (the law 100 was enacted) and this change increased considerably the coverage of patients with renal disease. CKD is a high cost disease and the excessive cost is a major concern for healthcare systems. Therefore, there is great interest in studying CKD and its progression, mortality and response to treatments, such as hemodialysis and peritoneal dialysis. In this work, we are concerned with the analysis of a dataset with end-stage renal disease (ESRD), stages IV and V, of CKD patients, followed from 2009 to 2011, from five cities in Colombia: Manizales, Montería, Rionegro, Sincelejo and Tunja. The data contains information about 1253 patients who had started dialysis treatment. The patients were periodically examined and tests were performed to obtain the levels of calcium, inorganic phosphorus, serum albumin, serum creatinine, among other blood tests. There is a major interest in studying survival times of those patients, but it is also of interest to study the disease's progression. Researchers involved in this study were particularly interested in studying the time until a *progression* of the disease occurs. This progression can be considered as a worsening in the patient's health, determined by the levels of some substances in the blood and by the clinical evaluation of the patient. Specifically, in this study progression was defined by high levels of inorganic phosphorus (level over 5,6 mg/dl). Therefore, in this particular study, there are two main events of interest: time until progression and time until death.

Common survival analysis techniques are suitable when the interest is to study only the time until the occurrence of an event. When observations are subject to more than one type of event (e.g, different causes of death) and the occurrence of an event prevents the occurrence of the other, there is a competing risks structure. Our CKD data, nevertheless, cannot be viewed as competing risks data since some patients may experience the progression and also death. In our example, the main interest is to study two events, one of which (terminal event) prevents the occurrence of the other (non-terminal event), but not vice versa. This structure is known as Semicompeting Risks, defined by Fine et al. (2001). In cancer studies, typically the non-terminal event is the relapse of cancer and sometimes the non-terminal event is simply called progression.

Recently, increased attention has been given to semicompeting risks data, and there are two main difficulties in addition to the existence of censored data: (i) the hypothesis that we have a pair of random variables with the restriction that one of them is always smaller than the other one (progression necessarily occurs before death) or one of them is not observed and (ii) the time until the non-terminal event and time until the terminal event cannot be considered as independent random variables. This possible dependence must be taken into account into statistical analysis. According to Hsieh, Wang & Ding (2008), if the relationship between the two events is completely unspecified, the marginal distribution of the time to a non-terminal event is not identifiable due to possible dependent censoring. In this context, while the univariate Kaplan-Meier (K-M) estimator can be used to estimate the survival function of the terminal event, it cannot be used for the non-terminal event and even a descriptive analysis for semicompeting risks data must be done using appropriate methodologies. It can be found in the literature different generalizations to the Kaplan-Meier estimator for bivariate situations (see, for example, Campbell & Foides 1980, Campbell 1981, Hanley & Parnes 1983, Tsai, Leurgans & Crowley 1986, Dabrowska 1988, Prentice & Cai 1992). Although we are interested in a bivariate survival function, semicompeting risks structure is different and it is not straightforward to adapt those estimators for our context. Therefore, the approach of Lakhal & Abdous (2008) using Archimedean copulas is a natural choice since it was proposed for semicompeting risks data. It also allows one to study explicitly the dependence between the time until the terminal and non-terminal events. Moreover, Archimedean copulas are quite referenced in the literature for bivariate survival models and they are associated with frailty multiplicative models (Romeo 2005).

In this work, we are focused on the statistical analysis of CKD data set using appropriate methods for semicompeting risks data. In order to carry a complete data analysis, we combined different new methodologies already available in the literature. Initially, we are concerned with methodologies for an appropriate descriptive analysis and we consider the estimation of the bivariate survival function and the non-terminal event survival function through Archimedean copulas, as suggested by Lakhal & Abdous (2008). Estimators for the survival functions are obtained for some Archimedean copulas (see Nelsen 2006, p.116) and we discuss those estimators in Section 2. In the sequence, we are interested in a regression model for both the time until progression and time until death. We focus on a three-state process, known as Illness-Death process, initially studied by Fix & Neyman (1951) and proposed by Xu, Kalbfleisch & Tai (2010) for semicompeting risks data. In this case, the inclusion of covariates and a possible dependence between the two times (terminal and non-terminal events) is taken into account by a shared frailty and the model is presented in Section 3. We also present a short simulation study comparing the performance of this model and a naive regression model where the dependence is not included. Finally, in Section 4 we present the results of the analysis of CKD data and in Section 5 we discuss the methods and results.

2. Bivariate Survival Function Estimation Through Archimedean Copulas

In this section, we describe the methodology used for a proper descriptive analysis of semicompeting risks data. We briefly introduce the concept of copulas and discuss the four copulas considered in this work. We then discuss estimation procedures for plotting graphics of the marginal survival functions of both events of interest.

Let X be the time to the non-terminal event and let Y be the time to the terminal event. For semicompeting risks data we have the restriction $X < Y$, thus, the bivariate survival function is defined in the upper wedge. Furthermore, the assumption of independence of X and Y is not reasonable and, therefore, we cannot use K-M estimator for the survival function of the non-terminal event X . It is then necessary to incorporate the dependence between X and Y in the estimator. In the bivariate case, let (X, Y) be continuous random variables with marginal survival functions $(S_X(x), S_Y(y))$. The bivariate survival function based on a copula $C_\alpha(\cdot)$ for some $\alpha \in \mathcal{A}$ is given by

$$S_{XY}(x, y) = P(X > x, Y > y) = C_\alpha(S_X(x), S_Y(y)), \quad 0 \leq x \leq y. \quad (1)$$

The copula is a natural function and an efficient way to construct multivariate distributions for random variables, since it couples the marginal distribution and captures their dependence. Archimedean copulas are an important family of copulas associated with bivariate survival models (Romeo 2005) and are defined by

$$C_\alpha(u, v) = \phi_\alpha^{-1}\{\phi_\alpha(u) + \phi_\alpha(v)\} \quad 0 \leq u \leq 1, 0 \leq v \leq 1, \quad (2)$$

where $\phi_\alpha(\cdot)$, known as the generator of $C_\alpha(\cdot, \cdot)$, is a continuous strictly decreasing convex function from $[0, 1]$ to $[0, \infty)$ such that $\phi_\alpha(1) = 0$ and $\phi_\alpha^{-1}(\cdot)$ denotes the inverse of $\phi_\alpha(\cdot)$ (Nelsen 2006). The parameter α is associated with the intensity of the dependence between the variables. Usually, this dependence is quantified by Kendall's tau measure of association, discussed in more details in Section 2.2.

Although the approach of Lakhal & Abdous (2008) for estimating survival and association in the semicompeting risks works for any Archimedean copula, we have implemented the methodology for the most referenced families. The Clayton family (which we call Family A) is associated to the multiplicative frailty gamma model and the dependence is (strongly) concentrated in the extreme lower left. The Frank family (family C) admits both positive and negative dependence and the Gumbel family (family D) has positive dependence strongly concentrated in the extremes (tail dependence). We have also implemented the family B corresponding to the family 4.2.2 defined on page 116 of Nelsen (2006) due to the simple expression of the copula. Table 1 shows the function $C_\alpha(u, v)$ that defines each family and the corresponding generating function.

TABLE 1: Archimedean copulas families considered.

Family	Copula, $C_\alpha(u, v)$	Generating function $\phi_\alpha(k)$
A	$[\max(u^{-\alpha} + v^{-\alpha} - 1, 0)]^{-1/\alpha}, \quad \alpha \in [-1, \infty) \setminus \{0\}$	$\frac{1}{\alpha}(k^{-\alpha} + 1)$
B	$1 - [(1 - u)^\alpha + (1 - v)^\alpha]^{1/\alpha}, \quad \alpha \in [1, \infty)$	$(1 - k)^\alpha$
C	$-\frac{1}{\alpha} \ln \left(1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1} \right), \quad \alpha \in (-\infty, \infty) \setminus \{0\}$	$-\ln \left(\frac{e^{-\alpha k} - 1}{e^{-\alpha} - 1} \right)$
D	$\exp\{-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{1/\alpha}\}, \quad \alpha \in [1, \infty)$	$(-\ln k)^\alpha$

2.1. Censoring and Notation

We assume that our data is subject to right-censoring and therefore we may not observe X and Y for all subjects. The observable quantities are described as follow. Let $Z = \min(X, Y)$ and $\delta_x = I_{\{X < Y\}}$, where $I_{\{X < Y\}}$ is given by $I_{\{X < Y\}} = 1$ if $X < Y$ and $I_{\{X < Y\}} = 0$ otherwise. Let C be a right-censoring variable independent of (X, Y) and define the random variables $R = \min(Y, C)$, $S = \min(Z, C)$, $\delta_y = I_{\{Y < C\}}$, $\delta_z = I_{\{Z < C\}}$ and $\delta_{xz} = \delta_x \delta_z$. Thus, the data set consists of n independent replications of $D_i = \{(R_i, S_i, \delta_{y_i}, \delta_{z_i}, \delta_{xz_i}), i = 1, \dots, n\}$ of observable variables $(R, S, \delta_y, \delta_z, \delta_{xz})$.

It is important to notice that, due to the right-censoring, for semicompeting risks four situations are possible:

Case 1. Subject was observed to have progression at x_i and death at y_i , then, $D_i = (R_i = y_i, S_i = x_i, \delta_{y_i} = 1, \delta_{z_i} = 1, \delta_{xz_i} = 1)$;

Case 2. Subject died at time y_i without progression, $D_i = (R_i = y_i, S_i = y_i, \delta_{y_i} = 1, \delta_{z_i} = 1, \delta_{xz_i} = 0)$;

Case 3. Subject was observed to have progression at x_i and right censoring at y_i , so $D_i = (R_i = c_i, S_i = x_i, \delta_{y_i} = 0, \delta_{z_i} = 1, \delta_{xz_i} = 1)$;

Case 4. Both events progression and death were not observed, and hence, $D_i = (R_i = c_i, S_i = c_i, \delta_{y_i} = 0, \delta_{z_i} = 0, \delta_{xz_i} = 0)$.

2.2. Measures of Association

As mentioned before, the parameter α in (1) and (2) is associated with the intensity of the dependence and its interpretation depends on the specific copula considered. A common association measure used in this context is the Kendall's tau coefficient, defined by

$$\tau = E\{\text{sgn}((X_1 - X_2)(Y_1 - Y_2))\},$$

$$\text{where } \text{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases},$$

(X_1, Y_1) and (X_2, Y_2) are two independent replications of (X, Y) of any random variables. For Archimedean copulas, Kendall's tau coefficient can be easily computed, see Nelsen (2006) for more details.

In semicompeting risks, when right censoring is observed, (X_i, Y_i) and (X_j, Y_j) can be compared only when the event $A_{ij} = \tilde{X}_{ij} \leq \tilde{Y}_{ij} \leq \tilde{C}_{ij}$ occurs, with $\tilde{X}_{ij} = \min(X_i, X_j)$, $\tilde{Y}_{ij} = \min(Y_i, Y_j)$ and $\tilde{C}_{ij} = \min(C_i, C_j)$. This leads to some difficulties when working with Kendall's coefficient and it is then useful to define the conditional Kendall's tau (Oakes 1989):

$$\tau_\alpha = E\{\text{sgn}[(X_i - X_j)(Y_i - Y_j)] | A_{ij}\}.$$

Another important association measure is a local counterpart of the Kendall's tau, which is a local measure of dependence, defined by Clayton (1978) as the cross-ratio function, given by

$$\begin{aligned} \theta^*(x, y) &= \frac{P\{(X_i - X_j)(Y_i - Y_j) > 0 | \tilde{X}_{ij} = x; \tilde{Y}_{ij} = y\}}{P\{(X_i - X_j)(Y_i - Y_j) < 0 | \tilde{X}_{ij} = x; \tilde{Y}_{ij} = y\}} \\ &= \frac{S_{XY}(x, y) D_1 D_2 S(x, y)}{D_1 S_{XY}(x, y) D_2 S_{XY}(x, y)}, \end{aligned}$$

where $D_1 = -\partial/(\partial x)$ and $D_2 = -\partial/(\partial y)$. Oakes (1989) showed that $\theta_\alpha^*(x, y) = \theta(S_{XY}(x, y))$, then for the Archimedean copulas we have the result

$$\theta_\alpha^*(x, y) = -\frac{k\phi_\alpha''(k)}{\phi_\alpha'(k)} = \theta_\alpha(k),$$

where $k = S_{XY}(x, y)$.

Table 2 shows the expressions for Kendall's tau coefficient and the corresponding cross-ratio function for each of the four copulas considered in this work.

TABLE 2: Association measures of four Archimedean copulas families.

Family	Generating function $\phi_\alpha(k)$	Kendall's tau τ	cross-ratio function $\theta_\alpha^*(k)$
A	$\frac{1}{\alpha}(k^{-\alpha} + 1)$	$\frac{\alpha}{\alpha+2}$	$\alpha + 1$
B	$(1 - k)^\alpha$	$\tau = \frac{\alpha-2}{\alpha}$	$\frac{k(\alpha-1)}{1-k}$
C	$-\ln\left(\frac{e^{-\alpha k} - 1}{e^{-\alpha} - 1}\right)$	$1 - \frac{4}{\alpha^2} \int_0^\alpha \left(\frac{t}{e^{-t}-1}\right) dt - \frac{4}{\alpha}$	$\frac{\alpha k}{1-e^{-\alpha k}}$
D	$(-\ln k)^\alpha$	$-\frac{4}{\alpha} \left(\frac{1}{4}\right) + 1 = \frac{\alpha-1}{\alpha}$	$1 - \frac{\alpha-1}{\log k}$

The cross-ratio function and the conditional tau coefficient are related and it can be shown (Patiño 2012) that

$$\tau_\alpha = E\left\{\frac{\theta_\alpha\{S_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} - 1}{\theta_\alpha\{S_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1} \middle| A_{ij}\right\}. \quad (3)$$

In the next section, expression (3) will be used as motivation for an estimating equation for α .

2.3. Estimation

We are interested in estimating the marginal survival functions $S_X(x)$ and $S_Y(y)$ and also the bivariate survival $S_{XY}(x, y)$ based on our available data. Some conditional functions in which the copula parameter α is always involved may also be of interest and will be discussed later. To construct (1), we need the non-terminal and terminal survival function $S_X(x)$ and $S_Y(y)$, respectively, as well as the parameter α of the assumed copula. As mentioned before, the traditional K-M method (Kaplan & Meier 1958) can be used to estimate $S_Y(y)$. However, for the non-terminal survival function $S_X(x)$ it is not appropriate due to the possible dependence between X and Y . In this context, various estimators have been proposed: non parametric estimator by Fine et al. (2001), Pseudo Self-Consistent estimator by Jiang, Fine, Kosorok & Chappell (2005) and the copula-graphic estimator by Zheng & Klein (1982), which was adapted by Rivest & Wells (2001) for dependent censoring and by Lakhali & Abdous (2008) for semicompeting risks data.

We describe here the method proposed by Lakhali & Abdous (2008), know it as copula-graphic estimator for semicompeting risks. The authors first consider the estimation of the parameter α . Motivated by expression (3), it is possible to show that

$$E \left\{ \frac{\theta_\alpha \{S_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha \{S_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1} \middle| A_{ij} \right\} = E [I_{\{P\{(X_i - X_j)(Y_i - Y_j) > 0\}} | A_{ij}] .$$

This result motivates the following estimating equation:

$$\frac{1}{\binom{n}{2}} \sum_{i < j} w(\tilde{S}_{ij}, \tilde{R}_{ij}) 1_{\{A_{ij}\}} \times \left\{ I_{\{(X_i - X_j)(Y_i - Y_j) > 0\}} - \frac{\theta_\alpha \{S_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha \{S_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1} \right\} = 0, \quad (4)$$

where $w(\cdot, \cdot)$ is a random weight function associated to the pair (i, j) . Fine et al. (2001) suggest

$$w_{a,b}^{-1}(u, v) = \frac{1}{n} \sum_{i=1}^n 1_{\{S_i \geq \min(u,a), R_i \geq \min(b,v)\}},$$

for $a > 0$ and $b > 0$ constants. Note that equation (4) depends on the bivariate survival function, which it is what we want to estimate in (1). Nevertheless, a non parametric estimator may be used as a first approximation and Lin & Ying (1993) suggested the following

$$\tilde{S}_{XY}(x, y) = \frac{\sum_{i=1}^n I_{\{S_i > x, R_i > y\}}}{\hat{G}_C(y)}, \quad 0 \leq x \leq y,$$

where $\hat{G}_C(\cdot)$ is the K-M estimator of censoring survival function C .

For each copula considered in this work, a different estimating equation is obtained and numerical procedures may be needed. For the Clayton copula, a solution for the equation may be obtained analytically. The estimation equations or solution for the four copulas considered are shown in Table 3.

TABLE 3: The estimation equations four Archimedean copulas families.

Family	Estimation equation
A	$\hat{\alpha} = \frac{\sum_{i < j} w(\tilde{S}_{ij}, \tilde{R}_{ij}) D_{ij} (2\Delta_{ij} - 1)}{\sum_{i < j} w(\tilde{S}_{ij}, \tilde{R}_{ij}) D_{ij} (1 - \Delta_{ij})}$
B	$\frac{1}{\binom{n}{2}} \sum_{i < j} w(\tilde{S}_{ij}, \tilde{R}_{ij}) D_{ij} \times \left\{ \Delta_{ij} - \frac{\tilde{S}_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})^{(\alpha-1)}}{\tilde{S}_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})^{(\alpha-2)+1}} \right\} = 0$
C	$\frac{1}{\binom{n}{2}} \sum_{i < j} w(\tilde{S}_{ij}, \tilde{R}_{ij}) D_{ij} \times \left\{ \Delta_{ij} - \frac{\alpha \tilde{S}_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})}{\alpha \tilde{S}(\tilde{X}_{ij}, \tilde{Y}_{ij}) + 1 - \exp(-\alpha \tilde{S}_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij}))} \right\} = 0$
D	$\frac{1}{\binom{n}{2}} \sum_{i < j} w(\tilde{S}_{ij}, \tilde{R}_{ij}) D_{ij} \times \left\{ \Delta_{ij} - \frac{\ln(\tilde{S}_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})) - (\alpha-1)}{2 \ln(\tilde{S}_{XY}(\tilde{X}_{ij}, \tilde{Y}_{ij})) - (\alpha-1)} \right\} = 0$

After estimating α , we use the copula-graphic estimator adapted by Lakhali & Abdous (2008) for the survival function associated to the non-terminal event. The idea is to construct an estimator for $S_X(t)$ with similar properties to the K-M estimator, i.e., that jumps when a non-terminal event is observed. In order to do so, we need to consider $Z = \min(X, Y)$ and it is important to notice that $S_Z(t) = P(Z > t) = P(X > t, Y > t) = \phi_\alpha^{-1} \{ \phi_\alpha(S_X(t)) + \phi_\alpha(S_Y(t)) \}$. If t_0 is an instant when a non-terminal event (X) was observed, then it is also an instant that Z was observed. Therefore, the Kaplan-Meier estimator of $S_Z(t)$, which we denote by $\hat{S}_Z(t)$, should jump at t_0 and

$$\begin{aligned} \phi_{\hat{\alpha}}\{\hat{S}_Z(t_0)\} &= \phi_{\hat{\alpha}}\{\hat{S}_X^{CG}(t_0)\} + \phi_{\hat{\alpha}}\{\hat{S}_Y(t_0)\} \\ \phi_{\hat{\alpha}}\{\hat{S}_Z(t_0)\} - \phi_{\hat{\alpha}}\{\hat{S}_Z(t_0^-)\} &= \phi_{\hat{\alpha}}\{\hat{S}_X^{CG}(t_0)\} - \phi_{\hat{\alpha}}\{\hat{S}_X^{CG}(t_0^-)\} + \phi_{\hat{\alpha}}\{\hat{S}_Y(t_0)\} - \phi_{\hat{\alpha}}\{\hat{S}_Y(t_0^-)\}, \end{aligned}$$

where $\hat{S}_X^{CG}(t_0)$ is the copula-graphic estimator. Assuming continuous X and Y , the two events cannot occur at the same time and $\phi_{\hat{\alpha}}\{\hat{S}_Z(t_0)\} - \phi_{\hat{\alpha}}\{\hat{S}_Z(t_0^-)\} = \phi_{\hat{\alpha}}\{\hat{S}_X^{CG}(t_0)\} - \phi_{\hat{\alpha}}\{\hat{S}_X^{CG}(t_0^-)\}$. For an arbitrary t , summing this expression for all $t_0 < t$, we have

$$\sum_{S_i \leq t; \delta_{xzi}=1} \phi_{\hat{\alpha}}\{\hat{S}_Z(S_i)\} - \phi_{\hat{\alpha}}\{\hat{S}_Z(S_i^-)\} = \phi_{\hat{\alpha}}\{\hat{S}_X^{CG}(t)\}$$

and the copula-graphic estimator is given by

$$\hat{S}_X^{CG}(t) = \phi_{\hat{\alpha}}^{-1} \left[\sum_{S_i \leq t; \delta_{xzi}=1} \phi_{\hat{\alpha}}\{\hat{S}_Z(S_i)\} - \phi_{\hat{\alpha}}\{\hat{S}_Z(S_i^-)\} \right]. \tag{5}$$

Finally, we can find the semiparametric estimator of the bivariate survival function (1) with the marginals (5) for $S_X(x)$ and K-M Estimator for $S_Y(y)$. Large sample properties of the resulting estimator have been established (weak convergence to a Gaussian process) using martingale representation (Rivest & Wells 2001).

Furthermore, from the bivariate survival function, some other useful quantities can be obtained (Lakhali & Abdous 2008), such as the probability that a patient survives at y given that he has not progression at time x , given by

$$\hat{P}(Y > y | X > x; Y > x) = \frac{\hat{S}(x, y)}{\hat{S}(x, x)},$$

and the survival probability at y given that the patient has a progression at time x and is alive at time $t > x$:

$$\hat{P}(Y > y|X = x; Y > t) = \frac{\phi'_\alpha(C_\alpha\{\hat{S}_X(x), \hat{S}_Y(t)\})}{\phi'_\alpha(C_\alpha\{\hat{S}_X(x), \hat{S}_Y(y)\})}$$

These estimators can be easily obtained for the four families in Table 1; for more details see Patiño (2012).

Regarding computational implementation, only the estimation of the dependence parameter is computationally intensive. As already mentioned, for the Clayton copula there is no iterative method because an analytical solution is available; for the other copulas, there is a sum over $\binom{n}{2}$ terms in the equations and this may be the computational lengthiest step in estimation for large n , but this does not represent a real problem in applications.

3. Illness-Death Model

Generally in applied survival analysis, it is of interest to determinate the effect of covariates on the risks occurrence of the event of interest. For this purpose we use the Illness-Death model with shared frailty, proposed by Xu et al. (2010).

In this model, the semicompeting risks data can be described by a three-state process, the illness and death process, which is a process widely discussed in the literature (see, Fix & Neyman 1951, Fiocco, Putter & van Houwelingen 2008). The state occupied by the individual at time $t = 0$ is called on study (state 0) and it is assumed the patient remains in this state until the intermediate event or the death occurs. When the intermediate event (not terminal) occurs, the individual condition is called relapse or progression (state 1) and the death is the final condition or absorbing state (state 2). Notice that it is not possible a transition from the intermediate state to the initial state. The main interest is to model the transition intensities functions, i.e., the hazard rates associated with the transitions between states (See Figure 1).

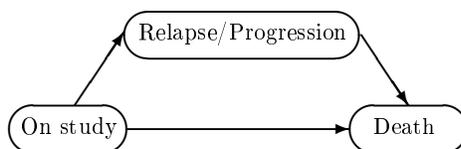


FIGURE 1: Illness-Death model for Semicompeting risks data.

The model proposed by Xu et al. (2010) is based on the hazard ratios functions of transitions between the states 0 (under study), 1 (relapse/progression) and 2 (death) defined by

$$\begin{aligned}\lambda_1(x) &= \lim_{\Delta \rightarrow 0} \frac{P(x \leq X < x + \Delta | X \geq x, Y \geq x)}{\Delta}, \\ \lambda_2(y) &= \lim_{\Delta \rightarrow 0} \frac{P(y \leq Y < y + \Delta | X \geq y, Y \geq y)}{\Delta} \text{ and} \\ \lambda_{12}(y|x) &= \lim_{\Delta \rightarrow 0} \frac{P(y \leq Y < y + \Delta | X = x, Y \geq y)}{\Delta},\end{aligned}\tag{6}$$

with $0 < x < y$.

Notice that the hazards rates $\lambda_1(x)$ and $\lambda_2(y)$ are not equal to the usual hazard rates defined in survival analysis; they correspond to the specific cause risk function used in the context of competing risks (Klein & Moeschberger 2003).

As mentioning before, in semicompeting risks, it is not appropriate to assume independence between the event X and Y . Therefore, in order to introduce a possible dependence of the transition times between the states progression and death, we consider the use of a shared frailty model, denoting the frailty by a random variable γ . The authors propose to model the conditional transitions hazard rates given γ and include this random variable multiplicatively. Since each transition hazard rate must be non-negative, we must have $\gamma > 0$ and it is convenient to assume $\gamma \sim \Gamma(1/\theta; 1/\theta)$ such that $E(\gamma) = 1$ and the variance is θ (which is unknown and must be estimated). The gamma distribution for the random effects is convenient for semi parametric proportional hazards models mainly due to mathematical convenience (for details see Wienke 2011) and it is, therefore, a natural choice for a first approach. However, models with different frailty distributions should be considered in future research.

Thus, the conditional hazard functions defined above (6) are given by

$$\begin{aligned}\lambda_1(x|\gamma) &= \gamma\lambda_{01}(x), \quad x > 0, \\ \lambda_2(y|\gamma) &= \gamma\lambda_{02}(y), \quad y > 0, \\ \lambda_{12}(y|\gamma, x) &= \gamma\lambda_{03}(y), \quad 0 < x < y.\end{aligned}$$

However, in a general model, covariates (\mathbf{z}) are incorporated similar to the Cox proportional hazards model, so that

$$\begin{aligned}\lambda_1(x|\gamma, \mathbf{z}) &= \gamma\lambda_{01}(x) \exp\{\boldsymbol{\beta}_1^t \mathbf{z}\}, \quad x > 0, \\ \lambda_2(x|\gamma, \mathbf{z}) &= \gamma\lambda_{02}(y) \exp\{\boldsymbol{\beta}_2^t \mathbf{z}\}, \quad y > 0, \\ \lambda_{12}(y|\gamma, x, \mathbf{z}) &= \gamma\lambda_{03}(y) \exp\{\boldsymbol{\beta}_3^t \mathbf{z}\}, \quad 0 < x < y.\end{aligned}\tag{7}$$

We are mainly interested in estimating the coefficients $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$, which contains information on the effects of covariates on the intensity functions conditional on frailty. Xu et al. (2010) do not make any assumption for the baseline functions $\lambda_{01}(x)$, $\lambda_{02}(y)$ and $\lambda_{03}(y)$ and non-parametric estimators are considered.

3.1. Estimation

Maximum likelihood can be used to obtain the estimators of the model parameters in (7). From this model, notice that the probability density function (and

survival function) depends on the baseline transition intensity functions $\lambda_{01}(k)$, $\lambda_{02}(k)$ and $\lambda_{03}(k)$, which are unknown and must be estimated. Following the idea of several non-parametric estimators of cumulative hazard functions (see Breslow (1972), Breslow (1975)), we can approximate the cumulative hazard function by a function that jumps when a failure or event is observed. Assuming that we observe m intermediate events, estimation of $\lambda_{01}(x)$ reduces to the problem of estimation m parameters. Denoting by $t_{r1}, t_{r2}, \dots, t_{rm}$ the instants in which a non-terminal event was observed, $t_{d1}, t_{d2}, \dots, t_{df}$ the times when a terminal event without progression was observed and $t_{rd1}, t_{rd2}, \dots, t_{rdg}$ the times in which a terminal event occurred after the progression, we can write

$$\Lambda_{01}(k) = \sum_{j=1}^m \lambda_{01j} I_{\{t_{rj} < k\}}, \quad \Lambda_{02}(k) = \sum_{j=1}^f \lambda_{02j} I_{\{t_{dj} < k\}} \quad \text{and} \quad \Lambda_{03}(k) = \sum_{j=1}^g \lambda_{03j} I_{\{t_{rdj} < k\}}.$$

Let $\boldsymbol{\lambda}_{01} = (\lambda_{011}, \dots, \lambda_{01m})^\top$, $\boldsymbol{\lambda}_{02} = (\lambda_{021}, \dots, \lambda_{02f})^\top$ and $\boldsymbol{\lambda}_{03} = (\lambda_{031}, \dots, \lambda_{03g})^\top$, where $\lambda_{01j} = d\Lambda_{01}(t_{rj}), j = 1 \dots, m$, $\lambda_{02j} = d\Lambda_{02}(t_{dj}), j = 1 \dots, f$ and $\lambda_{03j} = d\Lambda_{03}(t_{rdj}), j = 1 \dots, g$.

With this notation, the vector of parameters is

$$\boldsymbol{\eta} = (\theta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\lambda}_{01}, \boldsymbol{\lambda}_{02}, \boldsymbol{\lambda}_{03}) = (\theta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, d\boldsymbol{\Lambda}_{01}, d\boldsymbol{\Lambda}_{02}, d\boldsymbol{\Lambda}_{03}),$$

For the estimation of $\boldsymbol{\eta}$, it is necessary to obtain the joint survival function and the corresponding density probability function of (X, Y) to determine the contribution of an individual in each case on the likelihood $L(\boldsymbol{\eta})$. In semicompeting risks structure, it is not straightforward to compute the bivariate probability density function of (X, Y) from the model specification (7) because we have the restriction that $X < Y$ and X may not be observed.

In order to compute the bivariate density function, we assume conditional independence of X and Y given the frailty. Due to relation between the hazard, density and survival functions $h(\cdot) = f(\cdot)/s(\cdot)$ from (7) we can write

$$f_{X|\Gamma}(x|\gamma, \mathbf{z}) = \gamma \lambda_{01}(x) \exp\{\boldsymbol{\beta}_1^\top \mathbf{z} - \gamma \Lambda_{01}(x) \exp\{\boldsymbol{\beta}_1^\top \mathbf{z}\}\}$$

and

$$f_{Y|X,\Gamma}(y|x, \gamma, \mathbf{z}) = \lambda^*(y|x, \gamma, \mathbf{z}) S^*(y|x, \gamma, \mathbf{z}),$$

where

$$\lambda^*(y|x, \gamma, \mathbf{z}) = \begin{cases} \gamma \lambda_{03}(y) \exp\{\boldsymbol{\beta}_3^\top \mathbf{z}\}, & x \leq y, \\ \gamma \lambda_{02}(y) \exp\{\boldsymbol{\beta}_2^\top \mathbf{z}\}, & \text{otherwise (non-terminal event not observed)} \end{cases}$$

and

$$S^*(y|x, \gamma, \mathbf{z}) = \begin{cases} \exp\{-\gamma(\Lambda_{02}(x) \exp\{\boldsymbol{\beta}_2^\top \mathbf{z}\} + \Lambda_{03}(y) \exp\{\boldsymbol{\beta}_3^\top \mathbf{z}\} - \Lambda_{03}(x) \exp\{\boldsymbol{\beta}_3^\top \mathbf{z}\})\}, & x \leq y, \\ \exp\{-\gamma \Lambda_{02}(y) \exp\{\boldsymbol{\beta}_2^\top \mathbf{z}\}\}, & \text{otherwise.} \end{cases}$$

Using the fact that $f_{X,Y}(x, y|z) = \int_0^\infty f_{Y|X,\Gamma}(y|x, \gamma, z) f_{X|\Gamma}(x|\gamma, z) f_\Gamma(\gamma) d\gamma$, we can show that

$$f_{X,Y}(x, y|z) = \begin{cases} (1 + \theta)\lambda_{01}(x)e^{\beta_1^\top z} \lambda_{03}(y)e^{\beta_3^\top z} \\ (1 + \theta\Lambda_{01}(x)e^{\beta_1^\top z} + \theta\Lambda_{02}(x)e^{\beta_2^\top z} + \theta(\Lambda_{03}(y) - \Lambda_{03}(x))e^{\beta_3^\top z})^{-(1/\theta)-2}, & x \leq y, \\ (1 + \theta)\lambda_{01}(x)e^{\beta_1^\top z} \lambda_{02}(y)e^{\beta_2^\top z} [1 + \theta\Lambda_{01}(x)e^{\beta_1^\top z} + \theta\Lambda_{02}(y)e^{\beta_2^\top z}]^{-(1/\theta)-2}, & \text{otherwise.} \end{cases} \tag{8}$$

In expression (8), notice that the situation in which the non-terminal event is not observed depends on x , which is not appropriate. Xu et al. (2010) assume that this situation correspond to $X = \infty$ and compute

$$f_\infty(y|z) = \int_y^\infty (1 + \theta)\lambda_{01}(x)e^{\beta_1^\top z} \lambda_{02}(y)e^{\beta_2^\top z} [1 + \theta\Lambda_{01}(x)e^{\beta_1^\top z} + \theta\Lambda_{02}(y)e^{\beta_2^\top z}]^{-(1/\theta)-2} dx = \lambda_{02}(y)e^{\beta_2^\top z} [1 + \theta\Lambda_{01}(y)e^{\beta_1^\top z} + \theta\Lambda_{02}(y)e^{\beta_2^\top z}]^{-(1/\theta)-1}. \tag{9}$$

Expressions (8) for $x < y$ and (9) are the probability densities functions needed for the likelihood construction. Assuming that our data is subject to right censoring, we must compute the survival distributions for censored observations. As discussed before, due to censoring, we have four different possibilities and the contribution for the likelihood is different for observations in each case: for observations in case 1, the contribution to the likelihood is the density $f_{XY}(x, y)$, $x < y$; in case 2, the contribution is given by expression (9), i.e., $f_\infty(y|z)$; for case 3, we must compute $S_{Y|X=x}(y)f_X(x)$; and, finally, for case 4, we need the survival function $S_{XY}(x, y)$ (with $x = y$). Table 4 presents the expressions of the contribution for the likelihood of each case.

TABLE 4: Contribution to the likelihood construction for each possible situation for right censored data.

Contribution to the likelihood function	
Case 1	$(1 + \theta)\lambda_{01}(x)e^{\beta_1^\top z} \lambda_{03}(y)e^{\beta_3^\top z} \times [1 + \theta\Lambda_{01}(x)e^{\beta_1^\top z} + \theta\Lambda_{02}(x)e^{\beta_2^\top z} + \theta(\Lambda_{03}(y) - \Lambda_{03}(x))e^{\beta_3^\top z}]^{-\frac{1}{\theta}-2};$
Case 2	$\lambda_{02}(y)e^{\beta_2^\top z} [1 + \theta\Lambda_{01}(y)e^{\beta_1^\top z} + \theta\Lambda_{02}(y)e^{\beta_2^\top z}]^{-\frac{1}{\theta}-1};$
Case 3	$\lambda_{01}(x)e^{\beta_1^\top z} \times [1 + \theta\Lambda_{01}(x)e^{\beta_1^\top z} + \theta\Lambda_{02}(x)e^{\beta_2^\top z} + \theta(\Lambda_{03}(c) - \Lambda_{03}(x))e^{\beta_3^\top z}]^{-\frac{1}{\theta}-1};$
Case 4	$[1 + \theta\Lambda_{01}(c)e^{\beta_1^\top z} + \theta\Lambda_{02}(c)e^{\beta_2^\top z}]^{-\frac{1}{\theta}}$

Thus, with the observable variables $D_i = \{(R_i, S_i, \delta_{y_i}, \delta_{z_i}, \delta_{xz_i}), i = 1, \dots, n\}$ described in 2.1 the likelihood function is given by

$$\begin{aligned}
 L(\boldsymbol{\eta}) = & \prod_{i=1}^n \lambda_{01}(R_i)^{\delta_{xz_i}} \lambda_{02}(S_i)^{\delta_{y_i}(1-\delta_{xz_i})} \lambda_{03}(S_i)^{\delta_{xz_i}\delta_{y_i}} \\
 & \times \exp\{\delta_{xz_i}\boldsymbol{\beta}_1^\top \mathbf{z}_i + (\delta_{y_i}(1-\delta_{xz_i})\boldsymbol{\beta}_2^\top \mathbf{z}_i + \delta_{xz_i}\delta_{y_i}\boldsymbol{\beta}_3^\top \mathbf{z}_i)\} \times (1+\theta)^{\delta_{xz_i}\delta_{y_i}} \\
 & \times \{1 + \theta[\Lambda_{01}(S_i)e^{\boldsymbol{\beta}_1^\top \mathbf{z}} + \Lambda_{02}(S_i)e^{\boldsymbol{\beta}_2^\top \mathbf{z}} \\
 & + (\Lambda_{03}(R_i) - \Lambda_{03}(S_i))e^{\boldsymbol{\beta}_3^\top \mathbf{z}}]\}^{-\frac{1}{\theta} - \delta_{xz_i} - \delta_{y_i}}.
 \end{aligned} \tag{10}$$

Observe that in cases 2 and 4 the term $\Lambda_{03}(R_i) - \Lambda_{03}(S_i)$ in (10) vanishes, since $R_i = S_i$. For more details, see Patiño (2012).

Finally, the estimators are solution of score equation $U(\boldsymbol{\eta}) = \frac{\partial \log(L(\boldsymbol{\eta}))}{\partial \boldsymbol{\eta}} = \mathbf{0}$. The log likelihood function and the components of score vector are presented in the Appendix A. Xu et al. (2010) specify some conditions to establish asymptotic properties of the estimators. They discuss identifiability of the model and asymptotic convergence of estimators. The numerical approach proposed by the authors can be summarized as follows:

1. Let $\theta^{(0)}, \boldsymbol{\beta}_j^{(0)}, \Lambda_{0j}^{(0)}(k), j = 1, 2, 3$ be initial estimates. We can set $\theta^{(0)} = 0, \boldsymbol{\beta}_j^{(0)} = \mathbf{0}$ and for $\Lambda_{0j}^{(0)}(k), j = 1, 2, 3$ we use the Nelson-Aalen type estimate of the respective cumulative hazard functions.
2. Given $\Lambda_{0j}^{(0)}(k), j = 1, 2, 3$, we update estimates of $\theta, \boldsymbol{\beta}_j$ to $\theta^{(1)}, \boldsymbol{\beta}_j^{(1)}, j = 1, 2, 3$, from U_1, U_2, U_3, U_4 where $U_i = \frac{\partial \log(L(\boldsymbol{\eta}))}{\partial \beta_j}, i = 1, 2, 3$.
3. Given $\theta^{(1)}, \boldsymbol{\beta}_j^{(1)}, j = 1, 2, 3$, we solve the equations $U_5 = 0, U_6 = 0$ and $U_7 = 0$ for updated estimated $\Lambda_{0j}^{(1)}(k), j = 1, 2, 3$.
4. Finally, we update the initial estimates in (1) and we repeat steps (2)-(4) until convergence.

Let $\hat{I}(\hat{\boldsymbol{\eta}}) = -\frac{\partial U(\hat{\boldsymbol{\eta}})}{\partial \hat{\boldsymbol{\eta}}}$ be the observed information matrix for $\hat{\boldsymbol{\eta}}$ given by the second derivatives of log-likelihood function evaluated at maximum likelihood estimators (MLE) for $\boldsymbol{\eta}$. For large sample sizes, $\hat{I}(\hat{\boldsymbol{\eta}})^{-1}$ can be used to estimate asymptotic variance of $\boldsymbol{\alpha}^\top = (\theta, \boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top)$.

For the estimation of parameters of this model, we use the R package `nqslv` for the numeric solutions of the score equations. For estimation of $\theta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ we use a convergence criteria $\epsilon = 0.001$ and for $\lambda_{01}, \lambda_{02}, \lambda_{03}$ we consider a less restrictive criteria $\delta = 0.01$. We did not have convergence problems and the procedure is not much influenced by initial values (for regression parameters, initial values were obtained by a Cox model).

3.2. Simulation Study

A simulation study was conducted mainly in order to compare the illness-death frailty model with a simple regression model that does not consider the dependence of the observed events.

Considering the model (7), we simulated the frailty with mean 1 and variance θ from a gamma distribution with shape and scale parameters $1/\theta$. We added a random variable $Z \sim \text{Normal}(0,1)$ representing a covariate. We generated n observations T_{1i} and T_{2i} from the Weibull distribution for both non-terminal and terminal events: $T_{1i} \sim \text{Weibull}$ with hazard function $\lambda_{1i} = \gamma_i e^{\alpha + \beta_1 x_i}$ and $T_{2i} \sim \text{Weibull}$ with hazard function $\lambda_{2i} = \gamma_i e^{\alpha + \beta_2 x_i}$, $i = 1 \dots n$, respectively. Fixing $\alpha = -10$, we have $T_{ki} \sim \text{Weibull}(a_i, b_{ki})$, with $a_i = 2$ and $b_{ki} = \exp\{-1/a_i(\log \gamma_i + \alpha + \beta_k x_i)\}$ for $k = 1, 2$.

Since our data is subject to right-censoring, we added a proportion p_c of censored observations. As it is usual in real data applications, we considered both type I and random censoring: we assume that the study ends at time τ (type I censoring) and, if p_c is not achieved, we added a random censoring with $C \sim \text{Unif}(0, \tau)$.

We report results from 500 replications for $n = 100$, $n = 500$ and $n = 800$ with $\beta_1 = 1$, $\beta_2 = 1$, frailty variance given by $\theta = 0, 5$, $\theta = 1$ and $\theta = 2$ and, finally, fixed proportions of censoring p_c equal to 40%, 60%, 70%. We chose a follow up time $\tau = 1095$ days motivated by our real data application. All computer codes were developed for R software (for both simulation and application) and are available upon request.

Results of mean and standard error of estimated parameters from 500 replications are summarized in Figures 2 and 3. The horizontal line corresponds to true value of parameters β_1 and β_2 . In general, it is possible to observe that for most situations the illness death model was less biased than the naive Weibull model, and it is interesting to notice that, as the variance of frailty increases, the bias also increases. Therefore, one can conclude that the illness death model performs very favourably in terms of bias, although the variance of estimators (and hence the mean squared error) may be greater than the corresponding variance of the naive Weibull model for small samples. This is expected, since there are more parameters to be estimated in the frailty model considered.

4. Kidney Disease Data Analysis

As previously discussed, we are interested in analysing a dataset with 1253 end-stage renal disease (ESRD), stages IV and V, of CKD patients, followed from 2009 to 2011 by Fresenius Medical Care in Colombia and Laboratory Medical Echavarría. Patients included in the analysis started dialysis treatment and were periodically examined. Tests were performed to obtain the levels of calcium, inorganic phosphorus, serum albumin, serum creatinine, among other blood tests. It is of interest to study survival time of patients from beginning of dialysis treatment, but it is also of interest to study the disease's progression. In this first approach, it was considered that a patient experiences progression when their level of phosphorus in the blood is greater than 5.6 mg/dl.

All patients are from five cities in Colombia: Manizales (280 patients), Montería (370 patients), Rionegro (131), Sincelejo (334) and Tunja (138) and 61.2% are males. Regarding age, people between 51 and 65 years represent 34% of the data and those over 66 years represent 30.6%.

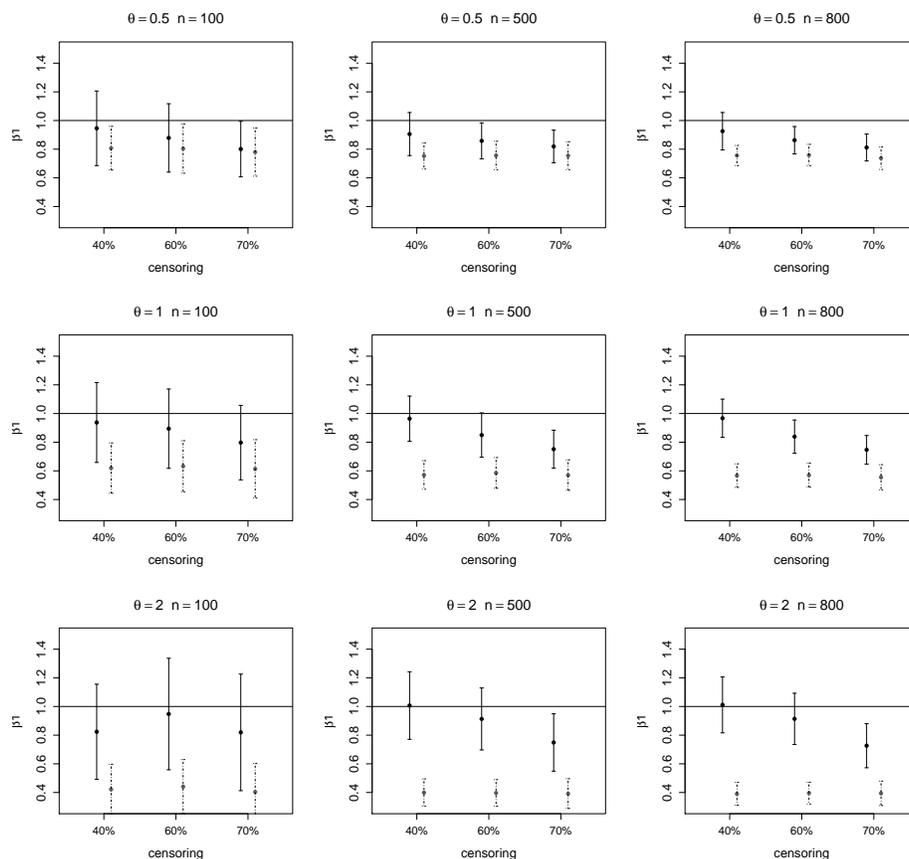


FIGURE 2: Simulations results for $\beta_1 = 1$. Continuous lines: Illness-Death model. Discontinuous lines: weibull model.

An initial descriptive analysis of data was done using copula graphic estimators. We computed the copula graphic estimator using the four copulas considered in this work. Table 5 shows the estimates of the copula parameter and Kendall's tau coefficient. An important aspect in application is the choice of the copula for the copula graphic estimator. We suggest a scatter plot of time to progression and time to death of uncensored observations as a first guide and also to use several copulas and compare the resulting plots. In our example, the estimated dependence is positive and small for most copulas; only family B copula had a different result. Although the estimated dependence for Family B copula is different from other copulas, copula graphic estimates for the survival function are very similar, as shown in Figure 4.

From the K-M plot of the survival distribution of patients and the copula graphic estimator of the survival function of the progression, we can conclude that after two years we still have around 60% of patients alive and also 60% of patients without progression of the disease. It is important to notice that we have only two

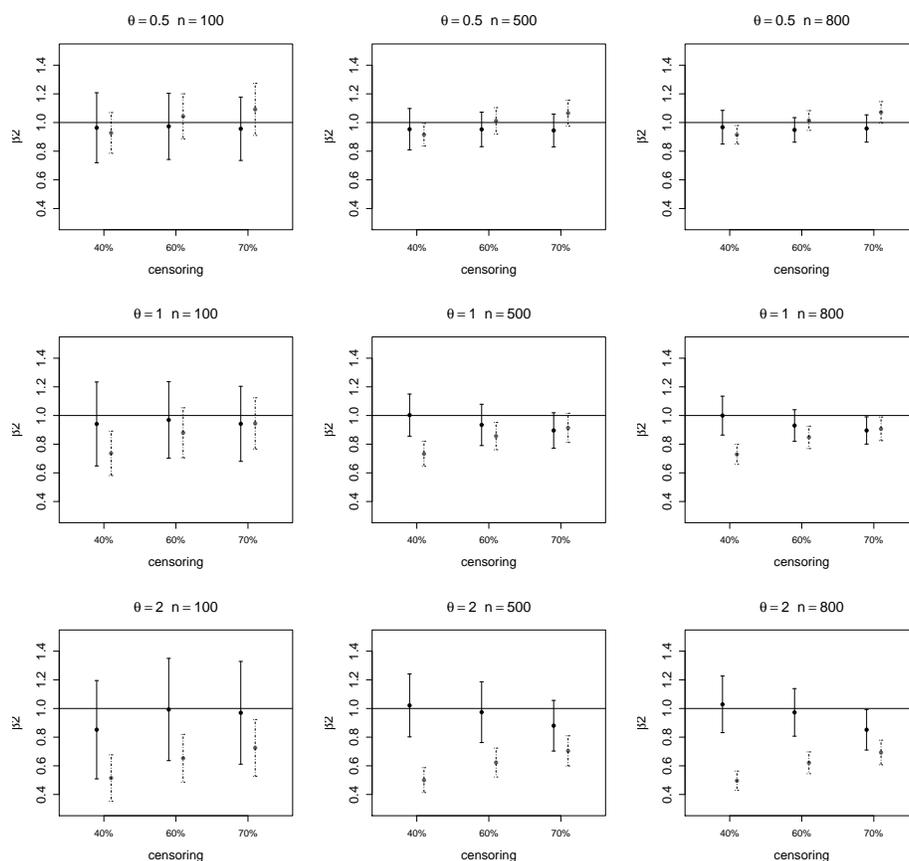


FIGURE 3: Simulations results for $\beta_2 = 1$. Continuous lines: Illness-Death model. Discontinuous lines: weibull model.

years of follow up, and the graphics show that a longer follow up time would be valuable for a better understanding of disease progression.

TABLE 5: Estimated copula parameter and dependence coefficient, $\hat{\alpha}$ and $\hat{\tau}$.

Copula	$\hat{\alpha}$ (se)	$\hat{\tau}$ (se)
Family A - Clayton copula	0,586 (0,142)	0,227 (0,042)
Family B	1,559 (0,022)	-0,282 (0,018)
Family C - Frank copula	1,387 (0,295)	0,151 (0,031)
Family D - Gumbel copula	1,141 (0,038)	0,124 (0,029)

For a multivariate analysis, we fitted the model described in Section 3 and four covariates were included: treatment (peritoneal dialysis/hemodialysis), sex (male/female), age (≤ 30 years/ ≥ 31 and ≤ 50 years/ ≥ 51 and ≤ 65 years/ ≥ 66

years) and city (Manizales/ Monteria/ Rionegro/ Sincelejo/ Tunja). The results of estimated coefficients shown in Table 6 and Table 7.

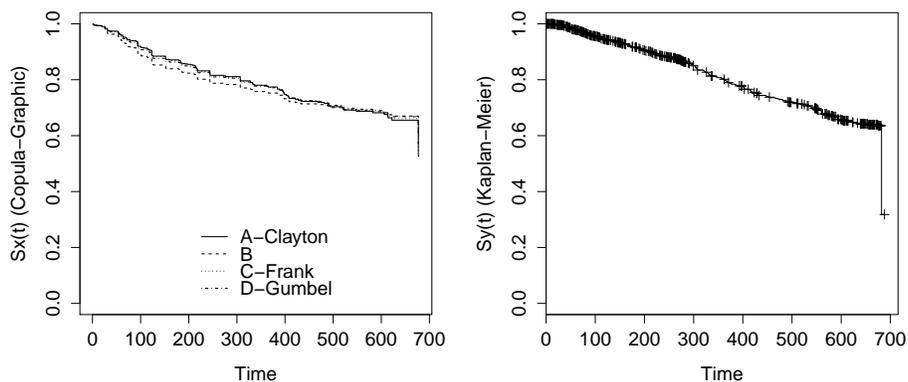


FIGURE 4: Estimated marginal survival functions for disease progression (left) and death (right).

TABLE 6: Estimates of Illness-death model for kidney disease analysis (death).

Covar./parameter	Time to death without progression			Time to death after progression		
	Estimate	SE	P value	Estimate	SE	P value
Treat. peritoneal	-	-	-	-	-	-
hemodialysis	-0,354	0,168	0,035	0,017	0,320	0,958
Sex female	-	-	-	-	-	-
male	-0,162	0,157	0,301	-0,154	0,264	0,558
Age ≤ 30	-	-	-	-	-	-
(years) ≥ 31 and ≤ 50	-0,108	0,279	0,697	-0,161	0,425	0,705
≥ 51 and ≤ 65	0,541	0,237	0,023	0,335	0,384	0,382
≥ 66	0,998	0,233	0,000	0,833	0,386	0,031
City Manizales	-	-	-	-	-	-
Monteria	0,263	0,206	0,200	0,755	0,335	0,024
Rionegro	0,259	0,382	0,499	0,917	0,560	0,101
Sincelejo	0,782	0,228	0,000	1,549	0,359	0,000
Tunja	0,633	0,302	0,036	0,538	0,454	0,236

The treatment was significant for time to progression and time to death without progression. Patients treated with hemodialysis have a higher risk of progression, however a lower risk of death when compared to patients having peritoneal disease. This result seems consistent with the fact that there was a smaller proportion of patients with death without progression (67.2 %) than with progression (80.2 %) and death after recurrence (80.6 %) in the group treated with hemodialysis.

Age is an important covariate to be included (it is significant), and the results are as expected. It is interesting to notice that age group 4 has a lower risk of progression, however a high risk of death. Regarding the city, Monteria and

TABLE 7: Estimates of Illness-death model for kidney disease analysis (progression).

Covar./parameter	Time to progression			
	Estimate	SE	P value	
Treat.	peritoneal dialysis	-	-	-
	hemodialysis	0,439	0,134	0,001
Sex	female	-	-	-
	male	0,019	0,111	0,867
Age (years)	≤ 30	-	-	-
	≥ 31 and ≤ 50	0,074	0,162	0,646
	≥ 51 and ≤ 65	0,045	0,152	0,767
	≥ 66	-0,508	0,175	0,004
City	Manizales	-	-	-
	Monteria	-0,466	0,144	0,001
	Rionegro	0,471	0,177	0,008
	Sincelejo	-0,338	0,160	0,034
	Tunja	0,197	0,184	0,284
θ	0,109	0,222	-	-

Rionegro showed significant progression and death after progression effect, which show us that the local where the patient is treated is an important factor that must be included in the analysis. The covariate sex is not significant in any transition rate and the estimated dependence is positive, but not strong ($\hat{\tau} = 0.05$).

It is important to notice that our approach was the use of hypothesis testing for covariate selection, however it would be useful to implement a discrimination criteria to help the decision of the best model, as pointed out by Lakhali & Abdous (2008). Also goodness-of-fit methods need to be developed to both assess the overall quality of fit in the regression model and assist the selection of a parametric family of Archimedean copulas (although in our application the conclusions were similar for different copulas).

5. Discussion

In this work, the semicompeting risks structure approach was used to analyse chronic kidney disease data. We focused on two events: the progression of disease (considered as a non-terminal or intermediate event) and death (terminal event). Although this is a preliminary analysis, the results will provide useful information for decision-making purposes and planning politics for patient care centers.

The first difficulty that arises in the analysis of semicompeting risks data is to obtain estimators of the survival function for a simple descriptive analysis or data visualization. Given the possible dependence between the two events, the Kaplan-Meier estimator is not suitable for estimating the marginal survival function observed when the intermediate event occurs. The approach proposed by Lakhali

& Abdous (2008) using the copula-graphic estimator in this situation seems to have good results. The use of the Archimedean copula family proved to be quite convenient for estimating the bivariate survival function. In general, for an initial descriptive analysis data, the approach with copula is a very good alternative. The selection of a particular copula of the Archimedean family can indeed influence the estimates and, although the differences among different copulas observed in the application are not significant, it would be interesting to develop a methodology to evaluate somehow which is the most convenient copula for a given application.

When the interest is focused on effects of covariates and a regression model is needed, the model proposed by Xu et al. (2010) is a very interesting alternative. The estimation by maximum likelihood and the iterative process is quite fast, obtaining convergence in few iterations. In this case, the baseline hazard rate functions are not specified and an usual approximation was used for estimation (the cumulative basal hazard function is approximated by a step function that jumps when an event is observed) which depends on the size of the sample. A parametric approach could also be developed in order to decrease the number of parameters to be estimated. Also, asymptotic properties of estimators have been established, however there are not any results for small sample sizes, so we do not recommend the use of this methodologies when the sample size is not large.

In general, the results obtained are consistent with expected by researchers and hemodialysis results to be a good alternative in many situations. Our data had patients from five different cities and Manizales is known to have a wide trajectory in the treatment of renal patients. Our results showed that Manizales is the city with higher survival probabilities (for both death without progression and death after progression). Rionegro and Sincelejo renal units are newer, and our results found less favourable conditions. Regarding the fragility parameter (i.e., the variance of the random effect), the estimative obtained from the general model, $\hat{\theta} = 0.109$ (0.222), indicates a weak association between progression and death. It is also an expected result considering that the observation time of patients with this disease was short. This is an ongoing study and a study with longer follow up time is needed. Researches suggest to continue this study to evaluate the survival of patients with CKD with a follow-up longer than two years and including clinic covariates like co-morbidity, type of primary disease and manifestations or not of extrarenal complications.

Acknowledgements

We are grateful to Clara Inés Salazar from Fresenius Medical Care of Colombia, Blanca Susana Ramirez and Maria Alejandra Restrepo from the medical laboratory Echavarria.

[Received: November 2017 — Accepted: August 2018]

References

- Breslow, N. (1972), ‘Contribuição para a discussão do artigo de d. r. cox’, *Journal of the Royal Statistical Society, Ser. B* **34**, 216–217.
- Breslow, N. (1975), ‘Analysis of survival data under the proportional hazards model’, *International Statistical Review* **43**, 45–58.
- Campbell, G. (1981), ‘Nonparametric bivariate estimation with randomly censored data.’, *Biometrika* **68**, 417–422.
- Campbell, G. & Foldes, A. (1980), *Large sample properties of nonparametric bivariate estimators with censored data.*, Technical report, Indiana. USA.
- Clayton, D. (1978), ‘A model for association in bivariate life tables and its application in epidemiological studies of family tendency in chronic disease incidence’, *Biometrika* **65**, 141–151.
- Cuenta del Alto Costo (2014), *Fondo Colombiano de Enfermedades de Alto Costo. Enfermedad Renal Crónica en Colombia. Resolución 4700 de 2008.*, Bogotá.
- Dabrowska, D. M. (1988), ‘Kaplan-meier estimate on the plane.’, *Annals of Statistics* **16**, 1475–1489.
- Fine, J., Jiang, H. & Chappell, R. (2001), ‘On semicompeting risk’, *Biometrika* **88**, 907–919.
- Fiocco, M., Putter, H. & van Houwelingen, H. (2008), ‘Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models’, *Statistics in Medicine* **27**, 4340–4358.
- Fix, E. & Neyman, J. (1951), ‘A simple stochastic model for recovery, relapse, death and loss of patients’, *Human Biology* **23**, 205–241.
- Gómez, R. A. (2006), ‘Renal disease in colombia’, *Renal failure* **28**, 643–647.
- Hanley, J. A. & Parnes, M. N. (1983), ‘Nonparametric estimation of a multivariate distribution in the presence of censoring.’, *Biometrics* **39**, 129–139.
- Hsieh, J., Wang, W. & Ding, A. (2008), ‘Regression analysis based on semicompeting risk data.’, *Journal of the Royal Statistical Society, Ser. B* **70**, 3–20.
- Jiang, H., Fine, J., Kosorok, R. & Chappell, R. (2005), ‘Pseudo self-consistent estimation of a copula model with informative censoring.’, *Scandinavian Journal of Statistics* **32**, 1–20.
- Kaplan, E. & Meier, P. (1958), ‘Nonparametric estimation from incomplete observations’, *Journal of the American Statistical Association* **53**, 457–481.

- Klein, J. & Moeschberger, M. (2003), *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd edn, Springer.
- Lakhal, L. and Rivest, L. & Abdous, B. (2008), 'Estimating survival and association in a semicompeting risk model', *Biometrics* **64**, 180–188.
- Lin, D. Y. & Ying, Z. (1993), 'A simple nonparametric estimator of the bivariate survival function under univariate censoring', *Biometrika* **80**, 573–582.
- Martín-Cleary, C. & Ortiz, A. (2014), 'Ckd hotspots around the world: where, why and what the lessons are. a ckj review series', *Clinical Kidney Journal* **7**, 519–523.
- Nelsen, R. (2006), *An Introduction to Copulas*, 2nd edn, Springer.
- Oakes, D. (1989), 'Bivariate survival models induced by frailties', *Journal of the American Statistical Association* **84**, 487–492.
- Patiño, E. G. (2012), Análise de dados com riscos semicompetitivos, Dissertação de Mestrado, Universidade de São Paulo, Instituto de Matemática e Estatística. Departamento de Estatística, São Paulo.
- Prentice, R. L. & Cai, J. (1992), 'Covariance and survival function estimation using censored multivariate failure time data.', *Biometrika* **79**, 495–512.
- Rivest, L.-P. & Wells, M. (2001), 'A martingale approach to the copula-graphic estimator for the survival function under dependent censoring', *Journal of Multivariate analysis* **79**, 138–155.
- Romeo, J. S. (2005), Modelagem Bayesiana para dados de sobrevivência bivariados através de cópulas, Tese de Doutorado, Universidade de São Paulo, Instituto de Matemática e Estatística. Departamento de Estatística, São Paulo.
- Tsai, W. Y., Leurgans, S. & Crowley, J. (1986), 'Nonparametric estimation of a bivariate survival function in the presence of censoring.', *Annals of Statistics* **14**, 1351–1365.
- Wienke, A. (2011), *Frailty Models in Survival Analysis*, 1st edn, CRC Press, Gūstrow.
- Xu, J., Kalbfleisch, J. & Tai, B. (2010), 'Statistical analysis of illness-death processes and semicompeting risks data', *Biometrics* **66**, 716–725.
- Zheng, M. & Klein, J. (1982), 'Estimates of marginal survival for dependent competing risk based on an assumed copula', *Biometrika* **82**, 127–138.

Appendix A. The log likelihood function and score vector of illness-death model

The log likelihood function of model (7) is

$$\begin{aligned} \log(L(\boldsymbol{\eta})) &= \sum_{i=1}^n \delta_{x_i} \log\{\lambda_{01}(S_i)\} + \delta_{y_i}(1 - \delta_{x_i}) \log\{\lambda_{02}(R_i)\} + \delta_{x_i} \delta_{y_i} \log\{\lambda_{03}(R_i)\} \\ &\quad + \delta_{x_i} \boldsymbol{\beta}_1^\top \mathbf{z}_i + \delta_{y_i}(1 - \delta_{x_i}) \boldsymbol{\beta}_2^\top \mathbf{z}_i + \delta_{x_i} \delta_{y_i} \boldsymbol{\beta}_3^\top \mathbf{z}_i + \delta_{x_i} \delta_{y_i} \log\{(1 + \theta)\} \\ &\quad - ((1/\theta) + \delta_{x_i} - \delta_{y_i}) \log\{1 + \theta[\Lambda_{01}(S_i)e^{\boldsymbol{\beta}_1^\top \mathbf{z}_i} + \Lambda_{02}(S_i)e^{\boldsymbol{\beta}_2^\top \mathbf{z}_i} \\ &\quad + (\Lambda_{03}(R_i) - \Lambda_{03}(S_i))e^{\boldsymbol{\beta}_3^\top \mathbf{z}_i}]\}. \end{aligned} \quad (\text{A.1})$$

In order to obtain the components of the vector score of partial derivatives from $\boldsymbol{\lambda}_{0R}$, $\boldsymbol{\lambda}_{0D}$ and $\boldsymbol{\lambda}_{0RD}$, notice that $\Lambda_{01}(k)$, $\Lambda_{02}(k)$ and $\Lambda_{03}(k)$ represent the cumulative hazard functions at time k in the sample for the time until progression, time until death without progression and time to death after progression, respectively. Thereby, it is possible to write

$$\Lambda_{01}(k) = \lambda_{011}I_{\{t_{r1} < k\}} + \lambda_{012}I_{\{t_{r2} < k\}} + \cdots + \lambda_{01m}I_{\{t_{rm} < k\}} = \sum_{j=1}^m \lambda_{01j}I_{\{t_{rj} < k\}}.$$

Analogously, $\Lambda_{02}(k) = \sum_{j=1}^f \lambda_{02j}I_{\{t_{dj} < k\}}$ and $\Lambda_{03}(k) = \sum_{j=1}^g \lambda_{03j}I_{\{t_{rdj} < k\}}$. Then, the log likelihood function $\log(L(\boldsymbol{\eta}))$ given in (A.1) can be written as

$$\begin{aligned} \log(L(\boldsymbol{\eta})) &= \sum_{i=1}^n \delta_{x_i} \log\{\lambda_{01}(S_i)\} + \delta_{y_i}(1 - \delta_{x_i}) \log\{\lambda_{02}(R_i)\} + \delta_{x_i} \delta_{y_i} \log\{\lambda_{03}(R_i)\} \\ &\quad + \delta_{x_i} \boldsymbol{\beta}_1^\top \mathbf{z}_i + \delta_{y_i}(1 - \delta_{x_i}) \boldsymbol{\beta}_2^\top \mathbf{z}_i + \delta_{x_i} \delta_{y_i} \boldsymbol{\beta}_3^\top \mathbf{z}_i + \delta_{x_i} \delta_{y_i} \log\{(1 + \theta)\} \\ &\quad - ((1/\theta) + \delta_{x_i} - \delta_{y_i}) \log \left\{ 1 + \theta \left[\sum_{j=1}^m \lambda_{01j} I_{\{t_{rj} < S_i\}} e^{\boldsymbol{\beta}_1^\top \mathbf{z}_i} + \sum_{j=1}^f \lambda_{02j} \right. \right. \\ &\quad \left. \left. \times I_{\{t_{dj} < S_i\}} e^{\boldsymbol{\beta}_2^\top \mathbf{z}_i} + \sum_{j=1}^g \lambda_{03j} I_{\{R_i \leq t_{rdj} < S_i\}} e^{\boldsymbol{\beta}_3^\top \mathbf{z}_i} \right] \right\}. \end{aligned}$$

To simplify the expression of partial derivatives of each component of score vector $U(\boldsymbol{\eta})$, let be $A_i = A_{1i} + A_{2i} + A_{3i}$, with

$$A_{1i} = \Lambda_{01}(S_i) \exp\{\boldsymbol{\beta}_1^\top \mathbf{z}_i\} = \sum_{j=1}^m \lambda_{01j} I_{\{t_{rj} < S_i\}} e^{\boldsymbol{\beta}_1^\top \mathbf{z}_i},$$

$$A_{2i} = \Lambda_{02}(S_i) \exp\{\boldsymbol{\beta}_2^\top \mathbf{z}_i\} = \sum_{j=1}^f \lambda_{02j} I_{\{t_{dj} < S_i\}} e^{\boldsymbol{\beta}_2^\top \mathbf{z}_i} \text{ and}$$

$$A_{3i} = (\Lambda_{03}(R_i) - \Lambda_{03}(S_i)) \exp\{\boldsymbol{\beta}_3^\top \mathbf{z}_i\} = \sum_{j=1}^g \lambda_{03j} I_{\{R_i \leq t_{rdj} < S_i\}} e^{\boldsymbol{\beta}_3^\top \mathbf{z}_i}.$$

Thus, the components of score vector $U(\boldsymbol{\eta})$ are:

$$\begin{aligned}
 U_1 &= \frac{\partial}{\partial \theta} \log(L(\boldsymbol{\eta})) \\
 &= \sum_{i=1}^n \left\{ \frac{\delta_{x_i} \delta_{y_i}}{1 + \theta} + \frac{n}{\theta^2} \log\{1 + \theta A_i\} - \left(\frac{A_i}{1 + \theta A_i} \left(\frac{1}{\theta} + \delta_{x_i} + \delta_{y_i} \right) \right) \right\}, \\
 U_2 &= \frac{\partial}{\partial \beta_1} \log(L(\boldsymbol{\eta})) \\
 &= \sum_{i=1}^n \left\{ \delta_{x_i} \mathbf{z}_i - \left(\frac{1}{\theta} + \delta_{x_i} + \delta_{y_i} \right) \frac{\theta A_{1i} \mathbf{z}_i}{1 + \theta A_i} \right\}, \\
 U_3 &= \frac{\partial}{\partial \beta_2} \log(L(\boldsymbol{\eta})) \\
 &= \sum_{i=1}^n \left\{ (1 - \delta_{x_i}) \delta_{y_i} \mathbf{z}_i - \left(\frac{1}{\theta} + \delta_{x_i} + \delta_{y_i} \right) \frac{\theta A_{2i} \mathbf{z}_i}{1 + \theta A_i} \right\} \mathbf{e} \\
 U_4 &= \frac{\partial}{\partial \beta_3} \log(L(\boldsymbol{\eta})) \\
 &= \sum_{i=1}^n \left\{ \delta_{x_i} \delta_{y_i} \mathbf{z}_i - \left(\frac{1}{\theta} + \delta_{x_i} + \delta_{y_i} \right) \frac{\theta A_{3i} \mathbf{z}_i}{1 + \theta A_i} \right\}.
 \end{aligned}$$

Finally, the j -th components of U_5 , U_6 and U_7 are

$$\begin{aligned}
 U_{5j} &= \sum_{i=1}^n \frac{d_{tr_j}}{\lambda_{01j}} - \left(\frac{1}{\theta} + \delta_{x_i} + \delta_{y_i} \right) \frac{I_{\{t_{r_j} < S_i\}} \theta \exp(\boldsymbol{\beta}_1^\top \mathbf{z}_i)}{1 + \theta A_i}, \quad j = 1 \cdots m, \\
 U_{6j} &= \sum_{i=1}^n \frac{d_{t_{a_j}}}{\lambda_{02j}} - \left(\frac{1}{\theta} + \delta_{x_i} + \delta_{y_i} \right) \frac{I_{\{t_{a_j} < S_i\}} \theta \exp(\boldsymbol{\beta}_2^\top \mathbf{z}_i)}{1 + \theta A_i}, \quad j = 1 \cdots f, \\
 U_{7j} &= \sum_{i=1}^n \frac{d_{tr_j}}{\lambda_{03j}} - \left(\frac{1}{\theta} + \delta_{x_i} + \delta_{y_i} \right) \frac{I_{\{S_i < t_{r_j} \leq R_i\}} \theta \exp(\boldsymbol{\beta}_3^\top \mathbf{z}_i)}{1 + \theta A_i}, \quad j = 1 \cdots g.
 \end{aligned}$$

Appendix B. Simulations Results

TABLE B.1: Summaries from 500 replications for the illness-death model (Par: parameter to be estimated, p_c : percentage of censoring) with different frailty variance ($\theta = 0.5, \theta = 1, \theta = 2$).

Par	p_c	N=100			N=500			N=800		
		40%	60%	70%	40%	60%	70%	40%	60%	70%
$\theta = 0.5$										
θ	Bias	-0.124	-0.204	-0.285	-0.137	-0.206	-0.243	-0.114	-0.209	-0.233
	SD	0.408	0.313	0.267	0.213	0.152	0.130	0.176	0.094	0.108
	MSE	0.182	0.140	0.153	0.064	0.066	0.076	0.044	0.053	0.066
β_1	Bias	-0.054	-0.121	-0.199	-0.094	-0.142	-0.181	-0.075	-0.137	-0.188
	SD	0.260	0.239	0.194	0.150	0.126	0.114	0.130	0.095	0.094
	MSE	0.071	0.072	0.077	0.031	0.036	0.046	0.023	0.028	0.044
β_2	Bias	-0.036	-0.027	-0.044	-0.047	-0.048	-0.055	-0.033	-0.052	-0.042
	SD	0.244	0.231	0.221	0.144	0.120	0.114	0.118	0.086	0.095
	MSE	0.061	0.054	0.051	0.023	0.017	0.016	0.015	0.010	0.011
β_3	Bias	0.071	0.208	0.293	-0.012	0.013	0.139	-0.030	-0.015	0.115
	SD	0.359	0.434	0.508	0.192	0.217	0.254	0.155	0.174	0.194
	MSE	0.134	0.231	0.344	0.037	0.047	0.084	0.025	0.030	0.051
$\theta = 1$										
θ	Bias	-0.193	-0.200	-0.326	-0.081	-0.318	-0.472	-0.064	-0.316	-0.466
	SD	0.617	0.641	0.612	0.344	0.325	0.222	0.279	0.263	0.140
	MSE	0.418	0.451	0.480	0.125	0.207	0.272	0.082	0.169	0.237
β_1	Bias	-0.063	-0.106	-0.203	-0.036	-0.150	-0.249	-0.033	-0.162	-0.253
	SD	0.279	0.277	0.260	0.158	0.155	0.133	0.133	0.116	0.100
	MSE	0.082	0.088	0.109	0.026	0.047	0.080	0.019	0.040	0.074
β_2	Bias	-0.059	-0.031	-0.058	0.003	-0.066	-0.104	-0.001	-0.070	-0.104
	SD	0.292	0.267	0.262	0.147	0.144	0.123	0.136	0.110	0.095
	MSE	0.089	0.072	0.072	0.022	0.025	0.026	0.018	0.017	0.020
β_3	Bias	-0.008	0.175	0.234	-0.027	-0.080	-0.040	-0.020	-0.103	-0.097
	SD	0.395	0.504	0.608	0.213	0.210	0.248	0.175	0.163	0.189
	MSE	0.156	0.285	0.424	0.046	0.050	0.063	0.031	0.037	0.045
$\theta = 2$										
θ	Bias	-0.587	-0.036	-0.236	0.115	-0.108	-0.687	0.149	-0.102	-0.756
	SD	1.090	1.444	1.463	0.811	0.847	0.675	0.687	0.700	0.541
	MSE	1.532	2.085	2.198	0.672	0.730	0.927	0.494	0.500	0.865
β_1	Bias	-0.176	-0.052	-0.181	0.007	-0.087	-0.251	0.011	-0.086	-0.274
	SD	0.332	0.389	0.407	0.236	0.216	0.200	0.195	0.179	0.154
	MSE	0.141	0.154	0.198	0.056	0.054	0.103	0.038	0.039	0.099
β_2	Bias	-0.148	-0.007	-0.030	0.022	-0.026	-0.120	0.029	-0.027	-0.148
	SD	0.343	0.357	0.358	0.220	0.212	0.177	0.198	0.166	0.141
	MSE	0.139	0.127	0.129	0.049	0.046	0.046	0.040	0.028	0.042
β_3	Bias	-0.150	0.099	0.120	0.006	-0.061	-0.148	0.022	-0.062	-0.202
	SD	0.433	0.556	0.674	0.279	0.269	0.240	0.240	0.214	0.191
	MSE	0.210	0.319	0.469	0.078	0.076	0.079	0.058	0.050	0.078

TABLE B.2: Summaries from 500 replications for the weibull model (Par: parameter to be estimated, p_c : percentage of censoring) with different frailty variance ($\theta = 0.5, \theta = 1, \theta = 2$).

Par	p_c	N=100			N=500			N=800		
		40%	60%	70%	40%	60%	70%	40%	60%	70%
$\theta = 0.5$										
β_1	Bias	-0.193	-0.197	-0.220	-0.248	-0.244	-0.245	-0.244	-0.241	-0.263
	SD	0.152	0.171	0.168	0.090	0.099	0.098	0.070	0.075	0.078
	MSE	0.060	0.068	0.077	0.070	0.069	0.070	0.065	0.064	0.075
β_2	Bias	-0.073	0.043	0.092	-0.085	0.012	0.065	-0.086	0.014	0.071
	SD	0.142	0.157	0.182	0.079	0.092	0.091	0.065	0.069	0.075
	MSE	0.025	0.027	0.042	0.013	0.009	0.012	0.012	0.005	0.011
$\theta = 1$										
β_1	Bias	-0.381	-0.368	-0.386	-0.429	-0.415	-0.430	-0.433	-0.430	-0.445
	SD	0.176	0.178	0.204	0.100	0.107	0.106	0.082	0.083	0.088
	MSE	0.176	0.167	0.191	0.194	0.183	0.196	0.194	0.192	0.206
β_2	Bias	-0.264	-0.121	-0.055	-0.268	-0.144	-0.086	-0.271	-0.152	-0.093
	SD	0.156	0.174	0.179	0.087	0.095	0.102	0.069	0.078	0.082
	MSE	0.094	0.045	0.035	0.079	0.030	0.018	0.078	0.029	0.015
$\theta = 2$										
β_1	Bias	-0.578	-0.559	-0.597	-0.601	-0.603	-0.608	-0.610	-0.604	-0.606
	SD	0.175	0.189	0.200	0.096	0.093	0.104	0.080	0.076	0.085
	MSE	0.365	0.349	0.396	0.370	0.373	0.380	0.378	0.371	0.374
β_2	Bias	-0.485	-0.347	-0.275	-0.500	-0.377	-0.296	-0.504	-0.379	-0.307
	SD	0.162	0.167	0.198	0.087	0.101	0.105	0.067	0.076	0.086
	MSE	0.261	0.148	0.115	0.257	0.152	0.099	0.258	0.149	0.102