

# Comparison of the marginal hazard model and the sub-distribution hazard model for competing risks under an assumed copula

Statistical Methods in Medical Research

0(0) 1–21

© The Author(s) 2019

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/0962280219892295

[journals.sagepub.com/home/smm](https://journals.sagepub.com/home/smm)Takeshi Emura<sup>1</sup> , Jia-Han Shih<sup>1</sup>, Il Do Ha<sup>2</sup> and Ralf A Wilke<sup>3</sup>

## Abstract

For the analysis of competing risks data, three different types of hazard functions have been considered in the literature, namely the cause-specific hazard, the sub-distribution hazard, and the marginal hazard function. Accordingly, medical researchers can fit three different types of the Cox model to estimate the effect of covariates on each of the hazard function. While the relationship between the cause-specific hazard and the sub-distribution hazard has been extensively studied, the relationship to the marginal hazard function has not yet been analyzed due to the difficulties related to non-identifiability. In this paper, we adopt an assumed copula model to deal with the model identifiability issue, making it possible to establish a relationship between the sub-distribution hazard and the marginal hazard function. We then compare the two methods of fitting the Cox model to competing risks data. We also extend our comparative analysis to clustered competing risks data that are frequently used in medical studies. To facilitate the numerical comparison, we implement the computing algorithm for marginal Cox regression with clustered competing risks data in the R *joint.Cox* package and check its performance via simulations. For illustration, we analyze two survival datasets from lung cancer and bladder cancer patients.

## Keywords

Clustered survival data, competing risk, Cox model, frailty model, survival analysis

## 1 Introduction

Competing risks data often arise in medical follow-up studies or industrial life tests where several different types of events determine the follow-up duration of a subject. In such circumstances, understanding the effects of covariates on individual event times is essential. For modeling the effects of covariates on event times, researchers typically fit the Cox proportional hazards model<sup>1</sup> on the hazard function of a specific event time of interest. However, there are various ways to formulate the Cox model in the case of competing risks.

For the analysis of competing risks data, three different types of hazard functions have been considered, namely the cause-specific hazard, the sub-distribution hazard (subhazard), and the marginal hazard function. Modeling the effect of covariates on the cause-specific hazard functions is the traditional approach to the analysis of competing risks data.<sup>2</sup> While it has been shown that the covariate effects on the cause-specific functions and on

<sup>1</sup>Graduate Institute of Statistics, National Central University, Taiwan

<sup>2</sup>Department of Statistics, Pukyong National University, South Korea

<sup>3</sup>Department of Economics, Copenhagen Business School, Denmark

## Corresponding author:

Takeshi Emura, Graduate Institute of Statistics, National Central University, No. 300, Zhongda Road, Zhongli District, Taoyuan City 32001, Taiwan.  
Email: [takeshiemura@gmail.com](mailto:takeshiemura@gmail.com)

the sub-distribution functions are different,<sup>3,4</sup> not much is known how the effect on the marginal hazard functions relate. Unlike the cause-specific hazard and subhazard, modeling covariate effects on the marginal hazard typically requires a strong assumption, called “assumed copula”.<sup>5</sup> This means that the dependence structure between competing event times is completely known or assumed by the researcher. If this assumption is deemed acceptable, it is straightforward to model covariate effects on marginal hazards.<sup>6–9</sup> While marginal hazard regression is relatively recently developed, the exploration of the marginal effects is historically regarded as the main goal of competing risk analysis.<sup>10</sup>

To draw an appropriate conclusion from competing risks data, it is essential to understand the difference between the three types of hazard function. Several studies<sup>4,11–13</sup> compared the difference between the cause-specific hazard and the subhazard. In addition, the comparison between the cause-specific hazard and marginal hazard is relatively straightforward, and has been considered in detail.<sup>8,14,15</sup> In this paper, we compare the difference between subhazard and marginal hazard models, which has not yet been explored in the literature. The absence of such a comparative study is partially due to the technical difficulty of implementing marginal regression and the lack of an adequate software package.

This paper makes the following contributions:

- a. We establish a mathematical relationship between the subhazard and the marginal hazard function under an assumed copula (Theorems 1 and 2). The relationship is also extended to clustered competing risks data that are increasingly popular in the literature (Theorem 3).
- b. We make numerical comparisons between subhazard regression and marginal hazard regression by using two cancer datasets.
- c. We develop an R function to implement a semiparametric inference method for marginal regression analysis with clustered competing risks data. This is made available in the *joint.Cox* R package.<sup>16</sup> This tool is useful for researchers who wish to compare the results of subhazard regression and marginal hazard regression.

The paper is organized as follows. Section 2 reviews classical competing risks models. Section 3 develops a mathematical relationship between the marginal hazard and subhazard. Section 4 compares covariate effects between the marginal hazard and subhazard through the Cox models. Section 5 extends our analysis to clustered competing risks data. Section 6 concludes and discusses the main findings.

## 2 Classical competing risks models

In the classical theory of competing risks, survival time of a subject is determined by several different causes of failures.<sup>17</sup> This implies that survival time is determined by the time at which at least one of the events occurs. In many applications of the competing risks theory, survival time of a subject can be any terminal event time, not necessarily time-to-death. For instance, if the major event of interest is death, the occurrence of dropout<sup>8</sup> or liver transplantation<sup>18</sup> can be regarded as a competing risk for death. The independence among different events is not assumed in competing risks analysis.

Hereafter, we consider bivariate competing risks where one event is a major focus of analysis and the other event is of secondary importance. For instance, if researchers are interested in survival time of prostate cancer patients, death from prostate cancer is the major cause of failure, and death due to other reasons is relegated to another cause of failure.<sup>6,11</sup> For another instance, if researchers are interested in overall survival of lung cancer patients, death from any cause is the major importance, and dropout and follow-up end are combined into another cause of failure.<sup>19</sup> Conversely, researchers may treat death as a competing risk if their major interest is on time-to-relapse.<sup>20</sup>

Let  $X$  be time to “Event 1” and  $Y$  be time to “Event 2”. Under competing risks, we observe the first occurring event time  $T = \min(X, Y)$ , and the event indicator  $\delta = \mathbf{I}(T = X)$ , where  $\mathbf{I}(\cdot)$  is the indicator function. Since one cannot observe  $X$  and  $Y$  simultaneously, the pair of event times  $X$  and  $Y$  are often called “latent times”.<sup>10,21</sup>

### 2.1 Hazard functions

The marginal hazard function for Event 1 is defined as

$$\lambda_1(t) = \Pr(t < X \leq t + dt | X > t) / dt$$

The marginal hazard function describes the instantaneous risk of experiencing Event 1 given that a subject has not yet experienced Event 1 at time  $t$ . It is simply a hazard for one marginal of the pair  $(X, Y)$ . Accordingly, the marginal survival function is related to the marginal hazard function through

$$S_1(t) \equiv \Pr(X > t) = \exp \left[ - \int_0^t \lambda_1(s) ds \right]$$

If  $X$  and  $Y$  were independent, the distribution could be easily estimated by treating  $Y$  as an independent censoring variable. In dependent competing risks, the marginal distribution is not identifiable unless some assumptions are made on the joint distribution of  $(X, Y)$ .<sup>22</sup>

Other functions of interest are therefore often considered in the presence of dependent competing risks. What can be estimated without knowing or assuming the risk dependence is the cause-specific (CS) hazard function. For Event 1 it is defined as

$$\lambda_1^{CS}(t) = \Pr(t < T \leq t + dt, \delta = 1 | T > t) / dt$$

This is the instantaneous risk of experiencing Event 1 given that a subject has not yet experienced any event before time  $t$ . The CS hazard function for Event 2 ( $\delta = 0$ ) is analogously  $\lambda_2^{CS}(t)$ . By definition,  $\lambda_1^{CS}(t) + \lambda_2^{CS}(t)$  is the hazard function for  $T$ .

Another identifiable quantity in dependent competing risks is the *sub-distribution function*

$$F_1^{Sub}(t) = \Pr(T \leq t, \delta = 1)$$

This is simply the probability of experiencing Event 1 before time  $t$ , which is also known as the cumulative incidence function (CIF). Many analyses of competing risks data start by plotting a nonparametric estimate of  $F_1^{Sub}(t)$ .<sup>11</sup>

The subhazard function for Event 1 is defined as a hazard function associated with the sub-distribution function through

$$\begin{aligned} \lambda_1^{Sub}(t) &= -d \log \{ 1 - F_1^{Sub}(t) \} / dt \\ &= \Pr(t < T \leq t + dt, \delta = 1 | \{ T > t \} \cup \{ T \leq t, \delta = 0 \}) / dt \end{aligned} \quad (1)$$

Here, the conditioning event of the CS hazard function is modified by adding the event  $\{ T \leq t, \delta = 0 \}$ . While it is difficult to interpret the conditioning event, an important advantage of the subhazard is the direct link to the sub-distribution function, such that  $F_1^{Sub}(t) = 1 - \exp \left[ - \int_0^t \lambda_1^{Sub}(s) ds \right]$

The relationship between the subhazard and CS hazard is well known. It is not difficult to show  $\lambda_1^{Sub}(t) = [dF_1^{Sub}(t)/dt] / \Pr(T > t)$  and  $\Pr(T > t) = \exp[-\Lambda_1^{CS}(t) - \Lambda_2^{CS}(t)]$ , where  $\Lambda_1^{CS}(t)$  and  $\Lambda_2^{CS}(t)$  are the cumulative CS hazard functions for Events 1 and 2, respectively. With these equations, we obtain the well-known relationship

$$F_1^{Sub}(t) = \int_0^t \lambda_1^{CS}(s) \exp[-\Lambda_1^{CS}(s) - \Lambda_2^{CS}(s)] ds$$

By taking the derivative of the preceding equation, the relationship between the subhazard and CS hazard is obtained as

$$\lambda_1^{Sub}(t) = \lambda_1^{CS}(t) \exp[-\Lambda_1^{CS}(t) - \Lambda_2^{CS}(t)] / \{ 1 - F_1^{Sub}(t) \}$$

These relationships are general and do not require restrictions on the model. Similarly, we have for the relationship between the marginal hazard and the subhazard

$$\lambda_1^{Sub}(t) = \frac{f_1^{Sub}(t)}{1 - F_1^{Sub}(t)} \leq \frac{f_1(t)}{S_1(t)} = \lambda_1(t), \forall t > 0$$

which follows from

$$1 - F_1^{Sub}(t) = 1 - \Pr(T \leq t, \delta = 1) = 1 - \Pr(X \leq t, X \leq Y) \geq 1 - \Pr(X \leq t) = S_1(t)$$

and

$$f_1^{Sub}(t) = dF_1^{Sub}(t)/dt = \Pr(t < X \leq t + dt, X \leq Y)/dt \leq \Pr(t < X \leq t + dt)/dt = -dS_1(t)/dt = f_1(t)$$

where  $f_1^{Sub}(t)$  is the sub-density function and  $f_1(t)$  is the marginal density function.

Unfortunately, the mathematical equation between the marginal hazard and subhazard cannot be established unless some model assumptions are made. Indeed, the same CS hazard function can originate from many different marginal models.<sup>22</sup> This problem is known as the nonidentifiability. To derive a mathematical relationship between the marginal hazard and subhazard functions, we first consider the independent competing risks model as a special case.

## 2.2 Independence assumption

A classical assumption is the independence between  $X$  and  $Y$ , which makes the joint distribution of  $(X, Y)$  identifiable from the joint distribution of  $(T, \delta)$ .<sup>22</sup> Under the independence assumption, it is known that  $\lambda_1(t) = \lambda_1^{CS}(t)$  and  $\lambda_2(t) = \lambda_2^{CS}(t)$ . Hence, the sub-distribution function can be expressed as the marginal hazards by

$$F_1^{Sub}(t) = \int_0^t \lambda_1(s) \exp[-\Lambda_1(s) - \Lambda_2(s)] ds$$

Accordingly, the subhazard is then

$$\lambda_1^{Sub}(t) = \lambda_1(t) \frac{\exp\{-\Lambda_1(t) - \Lambda_2(t)\}}{1 - \int_0^t \lambda_1(s) \exp\{-\Lambda_1(s) - \Lambda_2(s)\} ds}$$

It is important to note that  $\lambda_1^{Sub}(t) \neq \lambda_1(t)$  even though  $\lambda_1^{CS}(t) = \lambda_1(t)$ . To see this phenomenon more closely, let us consider the exponential margin  $\Lambda_1(t) = \Lambda_2(t) = \lambda t$ . Then

$$\lambda_1^{Sub}(t) = \lambda \frac{2\exp(-2\lambda t)}{1 + \exp(-2\lambda t)}, \quad F_1^{Sub}(t) = \frac{1 - \exp(-2\lambda t)}{2}$$

Thus, the constant marginal hazard functions give a non-constant subhazard function for Event 1. Therefore, the marginal hazard and subhazard functions have different functional forms in the case of independent competing risks.

## 2.3 Assumed copula for competing risks

In competing risks analysis, the two random variables  $X$  and  $Y$  are rarely independent. Zheng and Klein<sup>5</sup> showed that the joint distribution of  $X$  and  $Y$  becomes identifiable by assuming a copula function for the dependence structure between risks. Independence between  $X$  and  $Y$  is simply a special case of assuming the independence copula.

To model the dependence between two competing event times, Escarela and Carrière<sup>6</sup> proposed a *survival copula model*

$$\Pr(X > x, Y > y) = C_\theta\{S_1(x), S_2(y)\} \quad (2)$$

where  $C_\theta : [0, 1]^2 \rightarrow [0, 1]$  is a copula function with a parameter  $\theta$ .<sup>23</sup> The copula function can be any bivariate distribution function having the uniform marginal distribution on  $(0,1)$ . Here, the assumed copula in the sense of Zheng and Klein<sup>5</sup> means the assumptions for both the parametric form of  $C_\theta$  and its parameter value  $\theta$ . The survival copula model (2) is subsequently applied to various competing risks problems in literature.<sup>8,9,19,24</sup>

We shall not use another copula model  $\Pr(X \leq x, Y \leq y) = C_\theta\{1 - S_1(x), 1 - S_2(y)\}$  that produces a different joint distribution of  $X$  and  $Y$  from the survival copula model (2).

Examples of commonly used copula functions  $C_\theta$  are as follows

- The Clayton copula:

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad \theta > 0$$

- The Gumbel copula:

$$C_\theta(u, v) = \exp\left[-\left\{(-\log u)^{\theta+1} + (-\log v)^{\theta+1}\right\}^{\frac{1}{\theta+1}}\right], \quad \theta \geq 0$$

- The Farlie-Gumbel-Morgenstern (FGM) copula:

$$C_\theta(u, v) = uv\{1 + \theta(1 - u)(1 - v)\}, \quad -1 \leq \theta \leq 1$$

- The Fréchet-Hoeffding upper bound copula:

$$C_\infty(u, v) = \min(u, v)$$

The parameter  $\theta$  is related to Kendall's tau ( $\tau$ ) between  $X$  and  $Y$ . The Clayton copula has  $\tau = \theta/(\theta + 2)$ , the Gumbel copula has  $\tau = \theta/(\theta + 1)$ , and the FGM copula has  $\tau = 2\theta/9$ . These three copulas reduce to the independence copula  $C(u, v) = uv$  when  $\theta \rightarrow 0$ . When  $\theta \rightarrow \infty$ , the Clayton and Gumbel copulas reduce to the Fréchet-Hoeffding upper bound copula whose Kendall's tau is 1.

### 3 Mathematical relationship between the marginal hazard and subhazard

To establish the relationship between the subhazard and marginal hazard, it is convenient to introduce the following notations

$$D_\theta(s, t) = C_\theta\{\exp(-s), \exp(-t)\}, \quad D_\theta^{[1,0]}(s, t) = -\frac{\partial}{\partial s}D_\theta(s, t), \quad D_\theta^{[0,1]}(s, t) = -\frac{\partial}{\partial t}D_\theta(s, t)$$

Also, we rewrite the copula model (2) as  $\Pr(X > x, Y > y) = D_\theta\{\Lambda_1(x), \Lambda_2(y)\}$ . This expression emphasizes the relationship between the joint survival function and the marginal cumulative hazard functions. For instance, the Gumbel copula has  $D_\theta(s, t) = \exp\left[-(s^{\theta+1} + t^{\theta+1})^{1/(\theta+1)}\right]$ , which implies  $\Pr(X > x, Y > y) = \exp\left[-\{\Lambda_1^{\theta+1}(x) + \Lambda_2^{\theta+1}(y)\}^{1/(\theta+1)}\right]$ .

The sub-distribution function for Event 1 is

$$F_1^{Sub}(t) = \Pr(X \leq t, Y \geq X) = \int_0^t -\frac{\partial}{\partial x}\Pr(X > x, Y \geq y)|_{x=y} dy$$

This well-known formula implies that the sub-distribution function is obtained through the joint survival function  $\Pr(X > x, Y > y)$ .<sup>6,11</sup> Under the survival copula model (2), we have

$$F_1^{Sub}(t) = \int_0^t -\frac{\partial}{\partial x}D_\theta\{\Lambda_1(x), \Lambda_2(s)\}\Big|_{x=s} ds = \int_0^t \lambda_1(s)D_\theta^{[1,0]}\{\Lambda_1(s), \Lambda_2(s)\} ds \quad (3)$$

This expression describes the link between the sub-distribution function and marginal hazard function. The integral in equation (3) cannot be computed analytically for most of the well-known copulas.

By using equation (3) and  $\lambda_1^{Sub}(t) = -d\log[1 - F_1^{Sub}(t)]/dt$ , we arrive at the following result:

**Theorem 1:** Under the copula model (2), the marginal hazard and subhazard are connected through

$$\lambda_1^{Sub}(t) = \lambda_1(t) \frac{D_\theta^{[1,0]}\{\Lambda_1(t), \Lambda_2(t)\}}{1 - \int_0^t \lambda_1(s) D_\theta^{[1,0]}\{\Lambda_1(s), \Lambda_2(s)\} ds}$$

Under the common marginal assumption  $\Lambda(t) \equiv \Lambda_1(t) = \Lambda_2(t)$ , the integrals in Theorem 1 may have explicit forms. For instance, the Gumbel copula gives

$$\lambda_1^{Sub}(t) = \lambda(t) \frac{2^{1/(\theta+1)-1} \exp[-2^{1/(\theta+1)} \Lambda(t)]}{1 + 2^{-1} (\exp[-2^{1/(\theta+1)} \Lambda(t)] - 1)} = \lambda(t) \frac{2^{1/(\theta+1)} \exp[-2^{1/(\theta+1)} \Lambda(t)]}{\exp[-2^{1/(\theta+1)} \Lambda(t)] + 1}$$

For the Clayton copula, we have

$$\lambda_1^{Sub}(t) = \lambda(t) \frac{2 \exp\{\theta \Lambda(t)\} [2 \exp\{\theta \Lambda(t)\} - 1]^{-1/\theta-1}}{1 + [2 \exp\{\theta \Lambda(t)\} - 1]^{-1/\theta}} \quad (4)$$

and for the FGM copula, we have

$$\lambda_1^{Sub}(t) = \lambda(t) \frac{2 \exp\{-2\Lambda(t)\} (1 + \theta [1 - 2 \exp\{-\Lambda(t)\}] [1 - \exp\{-\Lambda(t)\}])}{1 + \exp\{-2\Lambda(t)\} (1 + \theta [1 - \exp\{-\Lambda(t)\}]^2)} \quad (5)$$

Our numerical studies confirm  $\lambda_1^{Sub}(t) < \lambda_1(t) \forall t > 0$  under various models (details are given in the Supplementary Material). The difference  $\lambda_1(t) - \lambda_1^{Sub}(t)$  can be quite large, which depends heavily on the choice of the marginal distributions. For instance, exponential distributions in the marginal models produce a steeply decreasing subhazard function (S1, Supplementary Material). The choice of copula function also influences the difference  $\lambda_1(t) - \lambda_1^{Sub}(t)$ . We do not observe equality  $\lambda_1^{Sub}(t) = \lambda_1(t)$  in any of the numerically examined models but technically it is attained under the following necessary and sufficient conditions.

**Theorem 2:** Under the copula model (2) with continuous marginal survival functions, one has

$$\lambda_1^{Sub}(t) = \lambda_1(t) \forall t \geq 0 \text{ if and only if } \Pr(X \leq Y) = 1$$

The proof of Theorem 2 is given in Appendix 1.

Under the copula model (2), Theorem 2 reveals that the equality  $\lambda_1^{Sub}(t) = \lambda_1(t)$  holds for all  $t \geq 0$  if and only if the model is degenerated under competing risks, i.e.  $T = \min(X, Y) = X$  with probability one. Thus, we conclude that  $\lambda_1^{Sub}(t) = \lambda_1(t) \forall t \geq 0$  does not hold for any real competing risks model that gives observed values for  $Y$ .

**Example 1:** Consider the Fréchet-Hoeffding upper bound copula model

$$\Pr(X > x, Y > y) = C_\infty\{S_1(x), S_2(y)\} = \min\{S_1(x), S_2(y)\}, x, y \geq 0 \quad (6)$$

for continuous marginal survival functions  $S_1(t) < S_2(t) \forall t \geq 0$ . One can verify

$$f_1^{Sub}(t) = -\frac{d}{dx} \Pr(X > x, Y > y)|_{x=y=t} = -\frac{dS_1(t)}{dt} = f_1(t) \quad \forall t \geq 0$$

Hence,  $\lambda_1^{Sub}(t) = \lambda_1(t) \forall t \geq 0$  holds true. By Theorem 2, we have  $\Pr(X \leq Y) = 1$ . One can also verify  $\Pr(X \leq Y) = 1$  directly from model (6).

### 3.1 Covariate effects

So far, we have focused on the difference between  $\lambda_1^{Sub}(t)$  and  $\lambda_1(t)$  as a function of  $t$ . Another approach is to compare the difference in terms of covariate effects given  $t$ . To study the covariate effects on the two hazards, we assume a marginal Cox model  $\lambda_1(t|\mathbf{Z}) = \lambda_{10}(t)\exp(\boldsymbol{\beta}'_1\mathbf{Z})$  for some baseline hazard function  $\lambda_{10}(\cdot)$ , regression coefficients  $\boldsymbol{\beta}_1$ , and covariates  $\mathbf{Z}$ . By Theorem 1

$$\lambda_1^{Sub}(t|\mathbf{Z}) = \lambda_{10}(t)\exp(\boldsymbol{\beta}'_1\mathbf{Z}) \frac{D_\theta^{[1,0]}\{\Lambda_{10}(t)\exp(\boldsymbol{\beta}'_1\mathbf{Z}), \Lambda_2(t|\mathbf{Z})\}}{1 - \int_0^t \lambda_{10}(s)\exp(\boldsymbol{\beta}'_1\mathbf{Z})D_\theta^{[1,0]}\{\Lambda_{10}(s)\exp(\boldsymbol{\beta}'_1\mathbf{Z}), \Lambda_2(s|\mathbf{Z})\}ds}$$

where  $\Lambda_{10}(t) = \int_0^t \lambda_{10}(s)ds$ . Hence,  $\lambda_1^{Sub}(t|\mathbf{Z})$  does not have a proportional hazard form since the covariate effect depends on  $t$ . An exception is the case of  $t = 0$ , where for some copulas (e.g. the Clayton copula in equation (4) and FGM copula in equation (5)) give  $\lambda_1^{Sub}(0|\mathbf{Z}) = \lambda_{10}(0)\exp(\boldsymbol{\beta}'_1\mathbf{Z})$ .

We emphasize the difference between the subhazard and marginal hazard by considering the case of  $\boldsymbol{\beta}_1 = 0$ , i.e. no marginal effect on Event 1. Then, the subhazard function is

$$\lambda_1^{Sub}(t|\mathbf{Z}) = \lambda_{10}(t) \frac{D_\theta^{[1,0]}\{\Lambda_{10}(t), \Lambda_2(t|\mathbf{Z})\}}{1 - \int_0^t \lambda_{10}(s)D_\theta^{[1,0]}\{\Lambda_{10}(s), \Lambda_2(s|\mathbf{Z})\}ds}$$

Hence, even if there is no marginal effect on Event 1, the subhazard of Event 1 is influenced by the marginal effect on Event 2.

## 4 Semiparametric inference for the Cox model

We compare two inference methods for the marginal Cox model and subhazard Cox model, respectively. Let  $\hat{\boldsymbol{\beta}}_1$  be an estimator obtained by fitting a model  $\lambda_1(t|\mathbf{Z}) = \lambda_{10}(t)\exp(\boldsymbol{\beta}'_1\mathbf{Z})$  and  $\hat{\boldsymbol{\beta}}_1^{Sub}$  be an estimator obtained by fitting a model  $\lambda_1^{Sub}(t|\mathbf{Z}) = \lambda_{10}^{Sub}(t)\exp(\boldsymbol{\beta}'_1^{Sub}\mathbf{Z})$ . We wish to compare the two methods of computing  $\hat{\boldsymbol{\beta}}_1$  and  $\hat{\boldsymbol{\beta}}_1^{Sub}$  from a dataset consisting of  $(T_j, \delta_{1j}, \delta_{2j}, \mathbf{Z}_j)$ ,  $j = 1, 2, \dots, n$ , where  $T_j = \min(X_j, Y_j, C_j)$ ,  $\delta_{1j} = \mathbf{I}(T_j = X_j)$ ,  $\delta_{2j} = \mathbf{I}(T_j = Y_j)$ , and  $C_j$  is independent censoring time. Censored cases correspond to  $\delta_{1j} = \delta_{2j} = 0$ .

Below, we review two semiparametric estimators  $\hat{\boldsymbol{\beta}}_1$  and  $\hat{\boldsymbol{\beta}}_1^{Sub}$  that do not require the model specifications for baseline hazard functions.

### 4.1 Fitting the subhazard model

Fine and Gray<sup>3</sup> proposed Cox regression on the sub-distribution based on the model

$$\lambda_1^{Sub}(t|\mathbf{Z}_j) = \lambda_{10}^{Sub}(t)\exp(\boldsymbol{\beta}'_1^{Sub}\mathbf{Z}_j)$$

The estimator  $\hat{\boldsymbol{\beta}}_1^{Sub}$  is obtained by applying some weights to the partial likelihood.<sup>3</sup> Here, the weights are computed by estimating the survival function of censoring time  $C_j$ . The *cmprsk* R package<sup>25</sup> can compute  $\hat{\boldsymbol{\beta}}_1^{Sub}$  and its standard error (SE) and confidence interval (CI). The package can also estimate the covariate-specific cumulative subhazard function as  $\hat{\Lambda}_1^{Sub}(t|\mathbf{Z}) = \hat{\Lambda}_{10}^{Sub}(t)\exp(\hat{\boldsymbol{\beta}}_1^{Sub}\mathbf{Z})$  and sub-distribution function as  $\hat{F}_1^{Sub}(t|\mathbf{Z}) = 1 - \exp[-\hat{\Lambda}_1^{Sub}(t|\mathbf{Z})]$ . See Pintilie<sup>11</sup> for the review.

Some explanation is necessary to interpret the value of  $\hat{\boldsymbol{\beta}}_1^{Sub}$  as the effects of covariates. In the subhazard model, the major focus is on the effects of covariates on the sub-distribution function. In this respect, the value of  $\hat{\boldsymbol{\beta}}_1^{Sub}$  is interpreted as the acceleration factor in a complementary log–log linear model for the sub-distribution function

$$\log[-\log\{1 - F_1^{Sub}(t|\mathbf{Z})\}] = \log[-\log\{1 - F_{10}^{Sub}(t)\}] + \boldsymbol{\beta}_1^{Sub}\mathbf{Z}$$



where  $F_{10}^{Sub}(t) = F_1^{Sub}(t|\mathbf{0})$ . Hence,  $\hat{\beta}_1^{Sub}$  is linked to the observed differences among the nonparametric estimates of the sub-distribution function computed for different covariate values. While this link is advantageous, the interpretation of the subhazard itself is not straightforward due to its complex conditioning events in equation (1).

## 4.2 Fitting the marginal Cox model

We assume the two Cox models for the two marginal hazard functions such that

$$\lambda_1(t|\mathbf{Z}_j) = \lambda_{10}(t)\exp(\beta_1'\mathbf{Z}_j), \quad \lambda_2(t|\mathbf{Z}_j) = \lambda_{20}(t)\exp(\beta_2'\mathbf{Z}_j)$$

The joint survival function is defined by

$$\Pr(X_j > x, Y_j > y|\mathbf{Z}_j) = C_\theta[\exp\{-\Lambda_1(x|\mathbf{Z}_j)\}, \exp\{-\Lambda_2(y|\mathbf{Z}_j)\}]$$

where  $C_\theta$  is a copula, the parameter  $\theta$  is assumed or known, and  $\Lambda_k(t|\mathbf{Z}_j) = \int_0^t \lambda_k(u|\mathbf{Z}_j)du$  is the marginal cumulative hazard functions ( $k = 1$  and  $2$ ). The estimates  $(\hat{\beta}_1, \hat{\beta}_2, \hat{\lambda}_{10}, \hat{\lambda}_{20})$  are obtained by a semiparametric method of Chen.<sup>8</sup> Appendix 2 provides the details of this method.

## 4.3 Methodological comparison of the two Cox models

In this section we qualitatively compare the two Cox models and provide some guidance on the choice of a suitable model.

If the main interest lies in a single event, fitting the subhazard Cox model is easier and requires weaker restrictions on the model. The subhazard Cox model for a single event time does not require any assumption on the other event time. On the other hand, the marginal hazard approach needs to specify two Cox models on two event times, as well as their copula function. A minor drawback of the subhazard approach is the need to estimate the censoring distribution by applying the inverse probability of censoring weighting to the partial likelihood. Fitting the marginal Cox model does not need to estimate the censoring distribution, yet the numerical computation is demanding.<sup>8</sup>

If the interest lies in the joint assessment of two events, the marginal Cox model may be desirable. However, investigation of the dependence structure between two events is inherently difficult with competing risks data. The subhazard Cox model does not provide any parameter related to the dependence structure since the latter is not a part of the observational model. In the marginal Cox model, the copula parameter provides a tool to assess dependence. However, the current consensus is to perform a sensitivity analysis under an assumed value of a copula parameter. Often, the fitted results are examined under a few different copula parameters selected by a researcher.<sup>8,9,15</sup> We shall further demonstrate the method of assessing the dependence through a real data example.

## 4.4 Numerical comparison of the two Cox models

We suggest comparing the two Cox models with aid of graphical diagnostic tools. We consider three estimators of  $F_1^{Sub}(t|\mathbf{Z}) = \Pr(T \leq t, \delta = 1|\mathbf{Z})$

- A new estimator for equation (3) that bases on the marginal Cox model

$$F_{1,\xi}^{Sub}(t|\mathbf{Z}) = \sum_{j: T_j \leq t} \delta_{1j} \hat{\lambda}_1(T_j|\mathbf{Z}) D_\theta^{[1,0]} \left\{ \hat{\Lambda}_1(T_j|\mathbf{Z}), \hat{\Lambda}_2(T_j|\mathbf{Z}) \right\}$$

where  $\hat{\lambda}_1(\cdot|\mathbf{Z}) = \hat{\lambda}_{10}(\cdot)\exp(\hat{\beta}_1'\mathbf{Z})$ ,  $\hat{\Lambda}_k(\cdot|\mathbf{Z}) = \hat{\Lambda}_{k0}(\cdot)\exp(\hat{\beta}_k'\mathbf{Z})$ , and  $(\hat{\beta}_1, \hat{\beta}_2, \hat{\lambda}_{10}, \hat{\lambda}_{20})$  are given in Appendix 2.

- The estimator under the subhazard Cox model

$$\hat{F}_1^{Sub}(t|\mathbf{Z}) = 1 - \exp\left[-\hat{\Lambda}_1^{Sub}(t|\mathbf{Z})\right]$$



where  $\hat{\Lambda}_1^{Sub}(t|\mathbf{Z}) = \hat{\Lambda}_{10}^{Sub}(t)\exp(\hat{\boldsymbol{\beta}}_1^{Sub}\mathbf{Z})$

- *The nonparametric (model-free) estimator*

$$\hat{F}_1^{NP}(t|\mathbf{Z}) = \sum_{j: T_j \leq t, \mathbf{Z}_j = \mathbf{Z}} \hat{S}(T_j|\mathbf{Z}) \frac{\delta_{1j}}{n_{j,\mathbf{Z}}}$$

where  $\hat{S}(t|\mathbf{Z}) = \prod_{j: T_j \leq t, \mathbf{Z}_j = \mathbf{Z}} \{1 - (\delta_{1j} + \delta_{2j})/n_{j,\mathbf{Z}}\}$  and  $n_{j,\mathbf{Z}} = \sum_{i: \mathbf{Z}_i = \mathbf{Z}} \mathbf{I}(T_i \geq T_j)$

We then plot the three estimators  $F_{1,\xi}^{Sub}(t|\mathbf{Z})$ ,  $\hat{F}_1^{Sub}(t|\mathbf{Z})$ , and  $\hat{F}_1^{NP}(t|\mathbf{Z})$  to check their similarity. A discrepancy among them is a signal of inappropriate model assumptions made in one of the Cox models. The first two estimators are inconsistent if their model assumptions are wrong. On the other hand, the last (model-free) estimator represents the empirical behavior of the sub-distribution function without any model assumption. The idea was presented in Escarela and Carrière<sup>6</sup> for the marginal Cox model and Pintilie<sup>11</sup> for the subhazard Cox model. A similar plot can be made to compare  $F_{2,\xi}^{Sub}(t|\mathbf{Z})$ ,  $\hat{F}_2^{Sub}(t|\mathbf{Z})$ , and  $\hat{F}_2^{NP}(t|\mathbf{Z})$  for Event 2.

The Cramér-von Mises (CvM) distance can be used as a tool for selecting a copula parameter in marginal regression. As a measure of fit based on the CvM distance, we suggest

$$\text{CvM} = \sum_{\mathbf{Z} \in I} \left[ \sum_{k=1}^2 \left( \frac{1}{n_{k,\mathbf{Z}}^{\delta}} \sum_{j: \mathbf{Z}_j = \mathbf{Z}} \delta_{kj} \left\{ F_{k,\xi}^{Sub}(T_j|\mathbf{Z}) - \hat{F}_k^{NP}(T_j|\mathbf{Z}) \right\}^2 \right) \right]$$

where  $I$  is the set of all possible covariate values and  $n_{k,\mathbf{Z}}^{\delta} = \sum_{j=1}^n \delta_{kj} \mathbf{I}(\mathbf{Z}_j = \mathbf{Z})$  for  $k = 1$  and  $2$ . The idea follows from Shih and Emura<sup>26</sup> who present a goodness-of-fit test based on the CvM distance under a parametric model and in the absence of covariates. We suggest using a grid search to find  $\theta$  that minimizes the CvM distance. The detailed algorithms and simulation results are given in the Supplementary Material (S2, Copula parameter selection). These results show that  $\theta$  is only *weakly* identified by the minimizer of the CvM distance due to the latter often being very flat. The convergence behavior of the estimated parameters is therefore worse than if  $\theta$  was known or assumed. The identification of  $\theta$  comes from the presence of covariates and the assumptions for the marginal Cox models.<sup>19,27,28</sup> In the absence of covariates,  $\theta$  is not identified.

However, the CvM distance may not be used to select between the marginal model and subhazard model. First, the CvM distance under the marginal model could be minimized over a number of copula parameters, or different copula functions. On the other hand, the subhazard does not have such options. Second, the CvM distance does not account for the difference in the number of parameters, which may lead to a favourable result for a model with a larger number of parameters.

*Remark I:* A goodness-of-fit method for the subhazard Cox model was developed by Scheike and Zhang<sup>29</sup> and Sfumato et al.<sup>30</sup> These methods are not applicable to measure the fit of the marginal Cox model, and hence, it cannot be used to compare the two models.

*Remark II:* Since the estimator  $F_{1,\xi}^{Sub}(t|\mathbf{Z})$  is new in the literature, we have checked its accuracy by simulations given in the Supplementary Material (S2, Copula parameter selection). Our results show that  $F_{1,\xi}^{Sub}(t|\mathbf{Z})$  consistently estimates  $F_1^{Sub}(t|\mathbf{Z})$  if the value of  $\theta$  is correctly specified.

## 4.5 Data example (lung cancer data)

We analyze the data on 125 lung cancer patients of Chen et al.<sup>31</sup> In this study, the primary endpoint is overall survival (i.e., time-to-death). During the follow-up, 38 patients died, while the remaining 87 patients were censored by dropout or the end of the study. Some early dropouts were possibly related to patients' health status. Therefore, we regard this censoring as a competing risk for death, leading to bivariate competing risks involving time-to-death ( $X_i$ ) for Event 1 and time-to-censoring ( $Y_j$ ) for Event 2. There is no independent censoring (i.e.  $C_j = \infty$ ). As in Chen et al.,<sup>31</sup> we use the 63 training samples out of the 125 samples.

We are interested in how the gene expression of *ZNF264* is associated with overall survival. The values of *ZNF264* were categorized according to the quantile (taking values 1, 2, 3 and 4; see Chen et al.<sup>31</sup>). It is found in

Chen et al.<sup>31</sup> that the gene expression of *ZNF264* is significantly associated with overall survival ( $P < 0.05$ ) based on the usual Cox regression with independent censoring. However, the analysis did not allow for dependent censoring. Therefore, we adopt a dependent competing risks model.

The subhazard model for Event 1 (death) is

$$\lambda_{1j}^{Sub}(t) = \lambda_{10}^{Sub}(t) \exp(\beta_1^{Sub} \times ZNF264_j)$$

and the subhazard model for Event 2 (censoring) is

$$\lambda_{2j}^{Sub}(t) = \lambda_{20}^{Sub}(t) \exp(\beta_2^{Sub} \times ZNF264_j)$$

We fitted the data to the two models to estimate  $\beta_1^{Sub}$  and  $\beta_2^{Sub}$  using the method of Section 4.1.

The marginal models for Events 1 and 2 are specified as

$$\begin{cases} \lambda_{1j}(t) = \lambda_{10}(t) \exp(\beta_1 \times ZNF264_j) & \text{(for death),} \\ \lambda_{2j}(t) = \lambda_{20}(t) \exp(\beta_2 \times ZNF264_j) & \text{(for censoring),} \\ \Pr(X_j > x, Y_j > y) = [\exp\{\theta \Lambda_{1j}(x)\} + \exp\{\theta \Lambda_{2j}(y)\} - 1]^{-1/\theta} \end{cases}$$

where  $\theta = 0, 0.22, \dots$  and 18. The latter correspond to  $\tau = 0, 0.1, \dots$  and 0.9, respectively. For each  $\theta$ , we fitted the data to estimate  $\beta_1$  and  $\beta_2$  using the method of Section 4.2.

Table 1 summarizes the results of fitting the data. Under the subhazard model, the effect of *ZNF264* gene on overall survival is significant ( $P < 0.05$ ). Under the marginal hazard model, the effect of *ZNF264* gene on overall survival is also significant across all the selected values of  $\theta$ . However, the effect sizes in the two different models are interpreted in different ways. For instance, the value  $\hat{\beta}_1 = 0.548$  (under  $\theta = 0.00$ ) implies that the unit increase of *ZNF264* gene expression yields about 1.73 ( $= \exp(0.548)$ ) times higher instantaneous risk of death. This value is equivalent to that obtained by Chen et al.<sup>31</sup> under the usual Cox regression with independent censoring. Meanwhile, the value  $\hat{\beta}_1^{Sub} = 0.425$  yields the degree of acceleration in the sub-distribution function

$$\log[-\log\{1 - F_1^{Sub}(t|Z+1)\}] = 0.425 - \log[-\log\{1 - F_1^{Sub}(t|Z)\}]$$

Interpreting  $\exp(\hat{\beta}_1^{Sub}) = 1.57$  as a relative risk of death is difficult, since the conditioning events are complex and involve censoring.

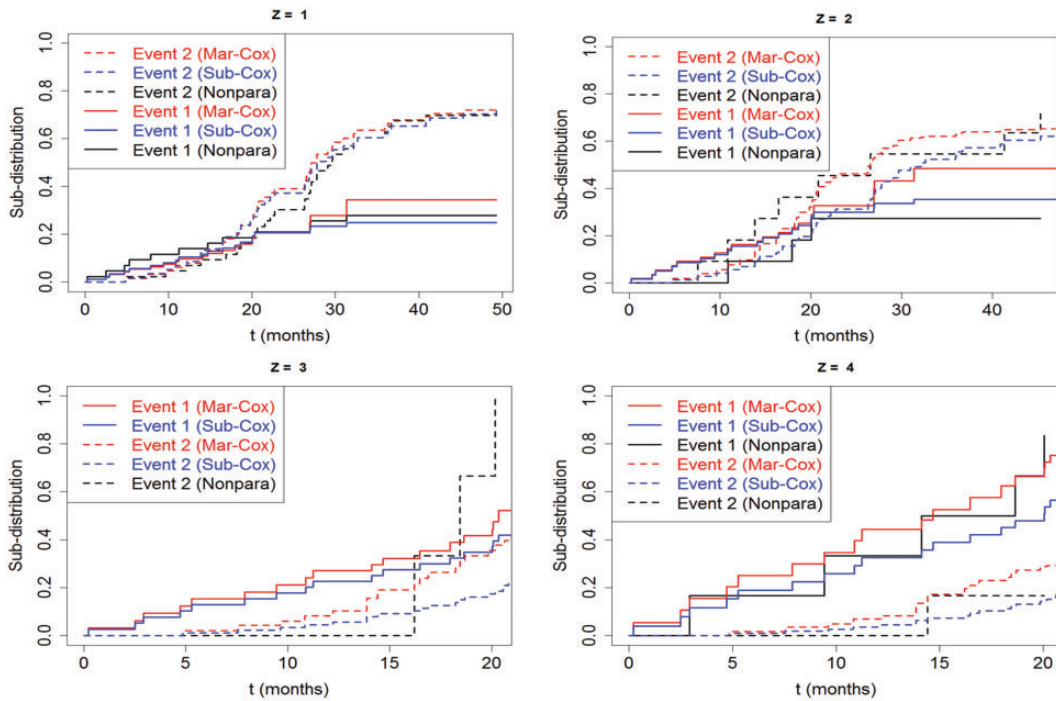
Table 1 shows some interesting differences in the regression coefficients between the subhazard model and marginal model for Event 2 (censoring); while  $\hat{\beta}_2^{Sub}$  is negative and non-significant,  $\hat{\beta}_2$  is positive for all values of  $\theta$  with a P-value  $< 0.05$  for  $\theta \geq 3$ . Thus, the overexpressed value of *ZNF264* gene expression may increase the instantaneous risk of censoring, though this effect may not become apparent from the sub-distribution-based analysis.

To draw some conclusions on the effect of *ZNF264* gene on the hazard of censoring time, we selected the copula parameter  $\theta = 8$  ( $\tau = 0.8$ ) that minimizes the CvM distance (see Section 4.4). Under this value, *ZNF264* gene is significantly associated with censoring time ( $\hat{\beta}_2 = 0.453$ , 95% CI: 0.156–0.751). On the other hand, the regression coefficient for the subhazard model lacks statistical significance ( $\hat{\beta}_2^{Sub} = -0.222$ , 95% CI:  $-0.182 \sim 0.698$ ). The difference of these conclusions comes naturally, as the two models are estimating two different quantities.

Figure 1 gives the model diagnosis plot for the subhazard model and the marginal model. Both models fit well to the data since their model-based estimators of the sub-distribution function capture the empirical behavior of the sub-distribution. While the CvM distance is smaller for the marginal model, this does not mean the marginal hazard is more suitable for the data (Section 4.4). The R codes for performing the above data analysis are available in the Supplementary Material (S3, Computer codes for the analysis of the lung cancer data).

**Table 1.** Regression coefficients obtained by fitting the lung cancer data.

Model	Event 1 (death) $\hat{\beta}_1$ (95%CI)	Event 2 (censoring) $\hat{\beta}_2$ (95%CI)
Subhazard	0.425 (0.044, 0.807)	-0.222 (-0.586, 0.143)
Marginal ( $\theta = 0.00$ ; $\tau = 0.0$ )	0.548 (0.144, 0.952)	0.259 (-0.176, 0.694)
Marginal ( $\theta = 0.22$ ; $\tau = 0.1$ )	0.560 (0.154, 0.965)	0.272 (-0.158, 0.702)
Marginal ( $\theta = 0.50$ ; $\tau = 0.2$ )	0.570 (0.162, 0.979)	0.280 (-0.143, 0.704)
Marginal ( $\theta = 0.86$ ; $\tau = 0.3$ )	0.578 (0.169, 0.988)	0.290 (-0.129, 0.710)
Marginal ( $\theta = 1.33$ ; $\tau = 0.4$ )	0.585 (0.178, 0.991)	0.311 (-0.103, 0.725)
Marginal ( $\theta = 2.00$ ; $\tau = 0.5$ )	0.593 (0.198, 0.987)	0.349 (-0.051, 0.749)
Marginal ( $\theta = 3.00$ ; $\tau = 0.6$ )	0.599 (0.229, 0.969)	0.394 (0.026, 0.762)
Marginal ( $\theta = 4.67$ ; $\tau = 0.7$ )	0.591 (0.251, 0.932)	0.432 (0.101, 0.762)
Marginal ( $\theta = 8.00$ ; $\tau = 0.8$ )	0.561 (0.251, 0.872)	0.453 (0.156, 0.751)
Marginal ( $\theta = 18.0$ ; $\tau = 0.9$ )	0.508 (0.227, 0.788)	0.455 (0.187, 0.723)



**Figure 1.** The estimated sub-distribution functions under the marginal Cox model (Mar-Cox), subhazard Cox model (Sub-Cox), and nonparametric model (Nonpara) using the lung cancer data. The plots show the estimated cumulative incidence rates for Event 1 (death) and Event 2 (censoring). The nonparametric estimator for Event 1 (death) is not available for  $Z = 3$  since this group only contains three censored samples. The marginal hazard model is fitted under  $\theta = 8$ .

## 5 Clustered competing risks data

This section extends our analysis to permit for more complex data structure, namely, clustered competing risks data. Clustered data frequently appear in medical studies, where patients are collected from different hospitals (in multi-center clinical trials) or different studies (in meta-analysis). Competing risks methods for analyzing clustered data are developed by a number of authors,<sup>24,32–38</sup> most of them using frailty to account for heterogeneity between clusters.

Consider  $G$  different clusters with the  $i$ -th cluster containing  $N_i$  subjects. Let  $X_{ij}$  be the time to Event 1,  $Y_{ij}$  be the time to Event 2 for  $i = 1, 2, \dots, G$  and  $j = 1, 2, \dots, N_i$ . Define  $T_{ij} = \min(X_{ij}, Y_{ij})$  and  $\delta_{ij} = \mathbf{I}(T_{ij} = X_{ij})$ . Independent censoring is introduced later.

To account for heterogeneity between clusters, we consider an unobserved frailty term  $u_i$  that acts on the hazard functions of  $X_{ij}$  and  $Y_{ij}$ . The frailty term  $u_i$  is considered as a realization of a positive-valued random variable. Two parametric distributions are commonly used for  $u_i$ : The gamma distribution with  $E_\eta(u_i) = 1$  and  $Var_\eta(u_i) = \eta$ , and the log-normal distribution  $u_i = \exp(v_i)$  where  $E_\eta(v_i) = 0$  and  $Var_\eta(v_i) = \eta$ . In either case, a large value of  $\eta$  implies a large amount of heterogeneity while the value of  $\eta \rightarrow 0$  implies no heterogeneity between the clusters.

The marginal hazard functions of  $X_{ij}$  and  $Y_{ij}$  given  $u_i$  are denoted as  $\lambda_{1ij}(t|u_i)$  and  $\lambda_{2ij}(t|u_i)$ , respectively. Here, the ‘‘marginal’’ refers to one marginal of the joint survival function  $\Pr(X_{ij} > x, Y_{ij} > y|u_i)$ . For instance, the marginal hazard for  $X_{ij}$  is  $\lambda_{1ij}(t|u_i) = -d\log\Pr(X_{ij} > t, Y_{ij} > 0|u_i)/dt$ . In a multi-center analysis, the frailty term  $u_i$  captures the ‘‘frailty’’ for the  $i$ -th center, acting on all the patients in the center. The corresponding marginal survival function is  $S_{1ij}(t|u_i) = \exp[-\Lambda_{1ij}(t|u_i)] = \exp\left[-\int_0^t \lambda_{1ij}(s|u_i) ds\right]$ .

As in Ha et al.,<sup>37</sup> the subhazard function for Event 1, given the frailty term, is

$$\lambda_{1ij}^{Sub}(t|u_i) = \Pr(t \leq T_{ij} < t + dt, \delta_{ij} = 1 | \{T_{ij} \geq t\} \cup \{T_{ij} < t, \delta_{ij} = 0\}, u_i) / dt$$

The subhazard and the sub-distribution function are related through  $\lambda_{1ij}^{Sub}(t|u_i) = -d\log[1 - F_{1ij}^{Sub}(t|u_i)]/dt$ , where  $F_{1ij}^{Sub}(t|u_i) = \Pr(T_{ij} \leq t, \delta_{ij} = 1|u_i)$  is the sub-distribution function for Event 1.

The next theorem extends the relationship between the marginal hazard and subhazard as given in Theorem 1 to clustered competing risks data:

**Theorem 3:** Under the joint frailty-copula model (Emura et al. 2017) of

$$\Pr(X_{ij} > x, Y_{ij} > y|u_i) = C_\theta\{S_{1ij}(x|u_i), S_{2ij}(y|u_i)\} \quad (7)$$

the marginal hazard and subhazard are connected through

$$\lambda_{1ij}^{Sub}(t|u_i) = \lambda_{1ij}(t|u_i) \frac{D_0^{[1,0]}\{\Lambda_{1ij}(t|u_i), \Lambda_{2ij}(t|u_i)\}}{1 - \int_0^t \lambda_{1ij}(x|u_i) D_0^{[1,0]}\{\Lambda_{1ij}(x|u_i), \Lambda_{2ij}(x|u_i)\} dx}$$

Assuming  $C_\theta(u, v) = uv$  in equation (7) corresponds to  $X_{ij}$  and  $Y_{ij}$  that are independent given the frailty term. In this case we have the joint frailty model of Rondeau et al.<sup>32</sup> However, assuming independence does not simplify the relationship between the marginal hazard and subhazard as in Section 3. A copula model similar to equation (7) was also considered by Rotolo et al.<sup>24</sup> for the purpose of simulating clustered competing risks data.

The implications of Theorem 3 are similar to those from Theorem 1 for non-clustered data. Specifically,  $\lambda_{1ij}^{Sub}(t|u_i) < \lambda_{1ij}(t|u_i)$  hold for  $\forall t > 0$  under various marginal models. The difference  $\lambda_{1ij}(t|u_i) - \lambda_{1ij}^{Sub}(t|u_i)$  is usually large, which depends on both marginal distributions and an assumed copula. The equality  $\lambda_{1ij}^{Sub}(t|u_i) = \lambda_{1ij}(t|u_i)$  does not hold except for a degenerated competing risks model satisfying  $\Pr(T_{ij} = X_{ij}|u_i) = 1$ . Thus, we conclude that  $\lambda_{1ij}^{Sub}(t|u_i) = \lambda_{1ij}(t|u_i)$  does not hold for any clustered competing risks data that contains observed values for  $Y_{ij}$ .

Next, we study the effect of covariates on the two hazards.

## 5.1 Semiparametric inference for the Cox models

Medical researchers are typically interested in estimating the effects of covariates  $\mathbf{Z}_j$  on one or two event times. For Event 1, we specify the covariate effects through the subhazard Cox model given  $u_i$

$$\lambda_{1ij}^{Sub}(t|u_i, \mathbf{Z}_{ij}) = u_i \lambda_{10}^{Sub}(t) \exp(\boldsymbol{\beta}_1^{Sub} \mathbf{Z}_{ij})$$

The estimator of  $\boldsymbol{\beta}_1^{Sub}$  is denoted by  $\hat{\boldsymbol{\beta}}_1^{Sub}$ . Similarly for Event 2, we consider the subhazard Cox model

$$\lambda_{2ij}^{Sub}(t|u_i, \mathbf{Z}_{ij}) = u_i \lambda_{20}^{Sub}(t) \exp(\boldsymbol{\beta}_2^{Sub} \mathbf{Z}_{ij})$$

and the estimator is denoted as  $\hat{\boldsymbol{\beta}}_2^{Sub}$ . The two estimators  $\hat{\boldsymbol{\beta}}_1^{Sub}$  and  $\hat{\boldsymbol{\beta}}_2^{Sub}$  are obtained separately by working on two different likelihood functions (e.g. Ha et al.<sup>37</sup>).

Meanwhile, we also specify the covariate effects through the marginal Cox models

$$\begin{cases} \lambda_{1ij}(t|u_i, \mathbf{Z}_{ij}) = u_i \lambda_{10}(t) \exp(\boldsymbol{\beta}'_1 \mathbf{Z}_{ij}), \\ \lambda_{2ij}(t|u_i, \mathbf{Z}_{ij}) = u_i^\alpha \lambda_{20}(t) \exp(\boldsymbol{\beta}'_2 \mathbf{Z}_{ij}) \end{cases}$$

where  $(\lambda_{10}, \lambda_{20})$  are baseline hazard functions, and  $\alpha$  can be any real number such as  $\alpha = 1$  (shared frailty case) and  $\alpha = 0$  (no frailty on Event 2). The estimators  $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$  can be obtained by maximizing a likelihood function for clustered competing risks data that was shortly mentioned in Emura et al.<sup>34</sup> To implement their methods, we make the *cmprskCox.reg* (`()`) function available in the R *joint.Cox* package. While our goal in this paper is to compare between marginal regression and subhazard regression, the R function itself can be a useful tool for researchers.

A conclusion from Theorem 3 and Section 4 is that the subhazard Cox model and marginal Cox model do not hold simultaneously. Hence,  $\hat{\boldsymbol{\beta}}_k^{Sub}$  and  $\hat{\boldsymbol{\beta}}_k$  are estimating different population values. Nonetheless, it is informative to review how these estimators can be computed from the clustered competing risks data.

Clustered competing risks data consist of  $(T_{ij}, \delta_{1ij}, \delta_{2ij}, \mathbf{Z}_{ij})$  for  $i = 1, 2, \dots, G$  and  $j = 1, 2, \dots, N_i$ , where  $T_{ij} = \min(X_{ij}, Y_{ij}, C_{ij})$ ,  $C_{ij}$  is independent censoring time,  $\delta_{1ij} = \mathbf{I}(T_{ij} = X_{ij})$  and  $\delta_{2ij} = \mathbf{I}(T_{ij} = Y_{ij})$  are event indicators. Next, we introduce the models for analyzing the data.

### 5.1.1 Fitting the subhazard model

The estimator  $\hat{\boldsymbol{\beta}}_1^{Sub}$  is obtained by maximizing the profiled h-likelihood function.<sup>13,37</sup> The *frailtyHL* R package<sup>39</sup> can compute  $\hat{\boldsymbol{\beta}}_1^{Sub}$  and its SE and CI. Note that the estimates  $\hat{\boldsymbol{\beta}}_1^{Sub}$  and  $\hat{\boldsymbol{\beta}}_2^{Sub}$  should be obtained separately by fitting two subhazard Cox models.

### 5.1.2 Fitting the marginal model

We give the methodological details for computing the marginal regression estimators  $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ . Emura et al.<sup>34</sup> derived the log-likelihood function

$$\begin{aligned} \ell(\eta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{10}, \lambda_{20} | \theta, \alpha) &= \sum_{i=1}^G \left[ \sum_{j=1}^{N_i} \{ \delta_{1ij} \log \lambda_{1ij}(T_{ij}) + \delta_{2ij} \log \lambda_{2ij}(T_{ij}) \} \right. \\ &\quad + \log \int_0^\infty u_i^{m_{1i} + \alpha m_{2i}} \left\{ \prod_{j=1}^{N_i} \psi_\theta [u_i \Lambda_{1ij}(T_{ij}), u_i^\alpha \Lambda_{2ij}(T_{ij})]^{\delta_{1ij}} \psi_\theta^* [u_i \Lambda_{1ij}(T_{ij}), u_i^\alpha \Lambda_{2ij}(T_{ij})]^{\delta_{2ij}} \right. \\ &\quad \left. \left. \times D_\theta [u_i \Lambda_{1ij}(T_{ij}), u_i^\alpha \Lambda_{2ij}(T_{ij})] \right\} f_\eta(u_i) du_i \right] \end{aligned}$$

where  $\lambda_{kij}(t) = \lambda_{k0}(t) \exp(\boldsymbol{\beta}'_k \mathbf{Z}_{ij})$ ,  $\Lambda_{kij}(t) = \int_0^t \lambda_{kij}(s) ds$ , and  $m_{ki} = \sum_{j=1}^{N_i} \delta_{kij}$  for  $k = 1$  and  $2$ . They proposed to apply a cubic M-spline function to model  $(\lambda_{10}, \lambda_{20})$ . Unlike the full nonparametric models, the spline yields smooth functions  $\lambda_{10}(t) = \sum_{\ell=1}^{L_1} g_\ell M_\ell(t)$  and  $\lambda_{20}(t) = \sum_{\ell=1}^{L_2} h_\ell M_\ell(t)$ , where  $M_\ell(t)$  is a cubic M-spline basis function,<sup>40</sup> and  $g_\ell \geq 0$  and  $h_\ell \geq 0$  are unknown parameters. Here,  $\psi_\theta = D_\theta^{[1,0]}/D_\theta$  and  $\psi_\theta^* = D_\theta^{[0,1]}/D_\theta$ .

To compute the M-spline basis functions, one needs to specify the number and locations of knots. As in literature,<sup>34,38</sup> we adopt equally spaced knots  $\zeta_1 < \zeta_2 < \zeta_3$ , where  $\zeta_1 = \min(T_{ij})$ ,  $\zeta_3 = \max(T_{ij})$ , and  $\zeta_2 = (\zeta_1 + \zeta_3)/2$ . The functions  $M_\ell(t)$ 's for  $L_1 = L_2 = 5$  are defined on  $t \in [\zeta_1, \zeta_3]$ .

We develop the R function *cmprskCox.reg* (`()`) in the *joint.Cox* R package that automatically obtains  $(\hat{\eta}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\lambda}_{10}, \hat{\lambda}_{20})$ . This function applies a Newton-type optimization to the penalized log-likelihood

$$\ell(\eta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{g}, \mathbf{h} | \theta, \alpha) - \kappa_1 \int \left\{ \frac{d^2}{dt^2} \lambda_{10}(t) \right\}^2 dt - \kappa_2 \int \left\{ \frac{d^2}{dt^2} \lambda_{20}(t) \right\}^2 dt \quad (8)$$



**Table 2.** Regression coefficients obtained by fitting the bladder cancer data.

Covariate (taking 0 or 1)	Model	Event 1 (recurrence) $\hat{\beta}_1$ (95%CI)	Event 2 (death) $\hat{\beta}_2$ (95%CI)
Chemo (1 for chemotherapy)	Subhazard	-0.70 (-1.04, -0.36)	0.64 (-0.09, 1.37)
	Marginal ( $\theta=0$ ; $\tau=0$ )	-0.55 (-0.91, -0.20)	0.34 (-0.38, 1.06)
	Marginal ( $\theta=0.5$ ; $\tau=0.2$ )	-0.52 (-0.87, -0.17)	0.19 (-0.48, 0.86)
	Marginal ( $\theta=2$ ; $\tau=0.5$ )	-0.51 (-0.86, -0.16)	-0.27 (-0.77, 0.23)
	Marginal ( $\theta=8$ ; $\tau=0.8$ )	-0.30 (-0.63, 0.04)	-0.18 (-0.53, 0.18)
Age (1 for >65 years)	Subhazard	-0.22 (-0.50, 0.06)	0.93 (0.43, 1.43)
	Marginal ( $\theta=0$ ; $\tau=0$ )	-0.10 (-0.39, 0.18)	0.73 (0.21, 1.26)
	Marginal ( $\theta=0.5$ ; $\tau=0.2$ )	-0.07 (-0.36, 0.21)	0.66 (0.16, 1.17)
	Marginal ( $\theta=2$ ; $\tau=0.5$ )	-0.04 (-0.31, 0.23)	0.37 (-0.02, 0.76)
	Marginal ( $\theta=8$ ; $\tau=0.8$ )	-0.05 (-0.30, 0.20)	0.08 (-0.20, 0.36)

where  $\theta$ ,  $\alpha$ , and  $\kappa_k > 0$  are given values. The R function *cmprskCox.reg* ( ) automatically selects the best values of  $\kappa_1$  and  $\kappa_2$  by a cross-validation while the values  $\theta$  and  $\alpha$  have to be pre-specified. We suggest considering  $\alpha = 0$ ,  $\alpha = 1$ , and other plausible values, and then choose the one that gives the best (largest) likelihood value in equation (8). This is straightforward by checking the output of “*cmprskCox.reg*( ,*alpha*=)”. The choice  $\alpha = 0$  implies that the risks of Event 2 are homogeneous across clusters. The choice  $\alpha = 1$  gives a shared frailty model for Events 1 and 2 so that the between-cluster heterogeneity in two events is similar. We have examined the performance of the inference procedure via simulations (Appendix 3).

## 5.2 Data example (bladder cancer)

We consider the bladder cancer data collected from 21 centers that participated in the EORTC trial 30791.<sup>41</sup> We use the subset of the bladder cancer data as considered in Section 1.2.4 of Ha et al.,<sup>13</sup> consisting of 396 patients with bladder cancer from 21 centers. We consider two competing endpoints, i.e. time to first bladder recurrence (Event 1) and time to death prior to recurrence (Event 2). Of 396 patients, 200 (50.51%) had recurrence of bladder cancer, 81 (20.45%) died prior to recurrence, and 115 (29.04%) were censored without recurrence or death at the last date of the follow-up. The numbers of patients per center varied from 3 to 78, with the mean of 18.9 and the median of 14. Two covariates are considered: *Chemotherapy* (0 = No vs. 1 = Yes) and *Age* (0 if age  $\leq 65$  years vs. 1 if age  $> 65$  years).

The subhazard model for Event  $k$  is

$$\lambda_{kij}^{Sub}(t|u_i) = u_i \lambda_{k0}^{Sub}(t) \exp\left(\beta_{k1}^{Sub} \times Chemo_{ij} + \beta_{k2}^{Sub} \times Age_{ij}\right)$$

for  $k = 1$  (Recurrence) and  $k = 2$  (Death). The two models are fitted separately to compute  $\hat{\beta}^{Sub}$ s.

The marginal model for the two events is jointly specified as

$$\begin{cases} \lambda_{1ij}(t|u_i) = u_i \lambda_{10}(t) \exp(\beta_{11} \times Chemo_{ij} + \beta_{12} \times Age_{ij}) & \text{(for recurrence),} \\ \lambda_{2ij}(t|u_i) = u_i^2 \lambda_{20}(t) \exp(\beta_{21} \times Chemo_{ij} + \beta_{22} \times Age_{ij}) & \text{(for death),} \\ \Pr(X_{ij} > x, Y_{ij} > y|u_i) = [\exp\{\theta \Lambda_{1ij}(x|u_i)\} + \exp\{\theta \Lambda_{2ij}(y|u_i)\} - 1]^{-1/\theta} \end{cases}$$

We set  $\theta = 0, 0.5, 2$ , or  $8$ , which corresponds to  $\tau = 0, 0.2, 0.5$ , or  $0.8$ , respectively. We report the fitted results for  $\alpha = 1$  (shared frailty model) as it gave a greater likelihood value than other values of  $\alpha$  (e.g.  $\alpha = 0$ ,  $\alpha = 2$ ).

Table 2 summarizes the results of fitting the two different Cox models. For the subhazard model, the effect of *Chemo* on time-to-recurrence is significant ( $\hat{\beta}^{Sub} = -0.70$ ;  $P$ -value  $< 0.05$ ) but its effect on time-to-death is non-significant ( $\hat{\beta}^{Sub} = 0.64$ ;  $P$ -value  $> 0.05$ ). The result implies that the chemotherapy is effective for reducing the event rate of recurrence, in the sense that the sub-distribution function is decreased as

$$\log[-\log\{1 - F_1^{Sub}(t|Chemo = \text{Yes})\}] = -0.70 + \log[-\log\{1 - F_1^{Sub}(t|Chemo = \text{No})\}]$$

**Table 3.** R functions in available R packages for competing risks.

	R package	No covariate	Covariates	Covariate & cluster
Subhazard	<i>cmprsk</i>	cuminc	crr	–
	Gray <sup>25</sup>			
	<i>timereg</i>	–	comp.risk	comp.risk
	Scheike et al. <sup>42</sup>		prop.odds.subdist	prop.odds.subdist
	<i>frailtyHL</i>	–	–	hlike.frailty
	Ha et al. <sup>39</sup>			
Marginal hazard	<i>Gofte</i>	–	prop.crr	–
	Sfumato et al. <sup>30</sup>		fcov.crr	
	<i>joint.Cox</i>	–	–	cmprskCox.reg
	Emura <sup>16</sup>			
	<i>compound.Cox</i>	–	dependCox.reg	–
	Emura et al. <sup>43</sup>		dependCox.reg.CV	

See Appendix 2 for some descriptions about the *compound.Cox* package.

However, the effect of *Chemo* on time-to-death is non-significant, possibly due to toxicity or side effects.

Under the marginal hazard model, the effect of *Chemo* on time-to-recurrence is significant ( $P$ -value  $< 0.05$ ) but its effect on time-to-death is non-significant ( $P$ -value  $> 0.05$ ). Hence, the significance of *Chemo* is confirmed on both the marginal hazard and subhazard. However, since the values of  $\hat{\beta}^{Sub}$  and  $\hat{\beta}$  are targeted to measure different risks, they must be interpreted differently. To interpret the results for the marginal regression, we choose  $\theta = 2$  ( $\tau = 0.5$ ) corresponding to a moderate amount of dependence between time-to-recurrence and time-to-death. Under this value, *Chemo* reduced the marginal hazard rate of recurrence by  $\exp(\hat{\beta}) = \exp(-0.51) = 0.60$ . Normally, researchers do not interpret  $\hat{\beta} = -0.51$  as the decrease in the survival function.

Now we compare the effects of *Age* on the two endpoints (Table 2). Under the subhazard model, *Age* significantly influences time-to-death. This means that older patients tend to have higher death rate in terms of their increased sub-distribution function. While this appears to be a plausible finding, the marginal model with  $\theta = 2$  ( $\tau = 0.5$ ) and  $\theta = 8$  ( $\tau = 0.8$ ) suggests a different conclusion: *Age* does not have a significant effect on time-to-death. A possible explanation of the effect of *Age* on subhazard model is the presence of some indirect influence of the effect of *Age* on time-to-recurrence. That is, even if *Age* has no marginal effect on time-to-death, it influences the subhazard of time-to-death through its effect on time-to-recurrence (Section 3).

## 6 Conclusion and discussion

In this article, we investigated in detail the relationship between the subhazard function and marginal hazard function in the analysis of competing risks data. We have adopted the assumed copula model of Zheng and Klein<sup>5</sup> to establish a mathematical relationship between the two hazard functions (Theorem 1). We have also adopted the joint frailty-copula model<sup>34,38</sup> to extend the mathematical result to clustered competing risks data (Theorem 3). One conclusion from Theorems 1 and 3 is that, even if there is no marginal effect on one event time, the subhazard of this event time is influenced by the marginal effect on the other event. Hence, even if a fitted covariate for a marginal model does not show any significance on one event, the covariate can still be significant on the subhazard model.

Besides these theoretical findings, we have developed an R function for marginal regression with clustered competing risks data: *cmprskCox.reg* (`()`), and checked the validity by means of simulations (Appendix 3). This adds a new tool to the existing R functions for fitting the marginal hazard model for competing risks data (Table 3). While we do not intend to compare all the available packages, the new R function allows researchers to compare the results of subhazard regression and marginal hazard regression for their own data. An R package for subhazard regression using interval censored data is only recently developed.<sup>44</sup>

One has to recognize that the marginal hazard function and subhazard function describe different population quantities. We have shown that a constant marginal hazard function produces a steeply decreasing subhazard function. In the analysis of the two cancer datasets, we have seen that regression coefficients fitted by the marginal



hazard model and the subhazard model can have different signs. These findings emphasize the fact that the two hazard functions measure different natures of event risks.

We introduced a Cramér-von Mises distance for selecting a copula parameter for the marginal hazard model. While we developed and tested the method for discrete or categorical covariates, an extension to continuous covariate might be possible by following the idea of Escarela and Carrière.<sup>6</sup> An extension of this method to clustered competing risks data is challenging, because of the lack of a nonparametric estimator for the sub-distribution function in the presence of heterogeneity and/or correlation caused by clustering.

### Acknowledgements

The authors kindly thank two anonymous referees for their valuable suggestions that improved the paper. This work was initiated in July 2018 when Emura T visited Department of Statistics, Pukyong National University, South Korea.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The research of Il Do Ha was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science & ICT (No.NRF-2017R1E1A1A03070747). The travel costs of Emura T have been financed by this research grant. The research of Emura T is funded by the grant from the Ministry of Science and Technology of Taiwan (MOST 107-2118-M-008 -003-MY3).

### ORCID iD

Takeshi Emura  <https://orcid.org/0000-0002-3904-4014>

### Supplemental material

Supplemental material for this article is available online.

### References

1. Cox DR. Regression models and life-tables (with discussion). *J Royal Stat Soc Ser B* 1972; **34**: 187–220.
2. Kalbfleisch JD and Prentice RL. The statistical analysis of failure time data. 2nd ed. New York, NY: John Wiley and Sons.
3. Fine JP and Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999; **94**: 548–560.
4. Jeong JH and Fine J. Direct parametric inference for the cumulative incidence function. *J Royal Stat Soc Ser C (Appl Stat)* 2006; **55**: 187–200.
5. Zheng M and Klein JP. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* 1995; **82**: 127–138.
6. Escarela G and Carriere JF. Fitting competing risks with an assumed copula. *Stat Meth Med Res* 2003; **12**: 333–349.
7. Braekers R and Veraverbeke N. A copula-graphic estimator for the conditional survival function under dependent censoring. *Can J Stat* 2005; **33**: 429–447.
8. Chen YH. Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *J Royal Stat Soc Ser B* 2010; **72**: 235–251.
9. Lo SM and Wilke RA. A copula model for dependent competing risks. *J Royal Stat Soc Series C (Appl Stat)* 2010; **59**: 359–376.
10. Crowder MJ. *Classical competing risks*. New York, NY: Chapman and Hall/CRC, 2001.
11. Pintilie M. *Competing risks: a practical perspective*. New York, NY: John Wiley & Sons, 2006.
12. Bakoyannis G and Touloumi G. Practical methods for competing risks data: a review. *Stat Meth Med Res* 2012; **21**: 257–272.
13. Ha ID, Jeong JH and Lee Y. *Statistical modelling of survival data with random effects: h-likelihood approach*. Singapore: Springer, 2017.
14. Emura T and Chen YH. *Analysis of survival data with dependent censoring, copula-based approaches, JSS research series in statistics*. Singapore: Springer, 2018.
15. Rivest LP and Wells MT. A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *J Multivariate Analys* 2001; **79**: 138–155.

16. Emura T. joint.Cox: joint frailty-copula models for tumour progression and death in meta-analysis, CRAN, version 3.6. 2019.
17. Cox DR and Oakes D. *Analysis of survival data*. New York, NY: CRC Press, 1984.
18. de Uña-Álvarez J and Veraverbeke N. Generalized copula-graphic estimator. *Test* 2013; **22**: 343–360.
19. Emura T and Chen YH. Gene selection for survival data under dependent censoring, a copula-based approach. *Stat Meth Med Res* 2016; **25**: 2840–2857.
20. Klein JP and Moeschberger ML. *Survival analysis: techniques for censored and truncated data*. New York, NY: Springer, 2003.
21. Crowder MJ. *Multivariate survival analysis and competing risks*. New York, NY: CRC Press, 2012.
22. Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci USA* 1975; **72**: 20–22.
23. Nelsen RB. *An introduction to copulas*. 2nd ed. New York, NY: Springer Series in Statistics, Springer-Verlag, 2006.
24. Rotolo F, Legrand C and Van Keilegom I. A simulation procedure based on copulas to generate clustered multi-state survival data. *Comput Meth Programs Biomed* 2013; **109**: 305–312.
25. Gray RJ. cmprsk: subdistribution analysis of competing risks, CRAN, version 2.2-7. 2017.
26. Shih J-H and Emura T. Likelihood-based inference for bivariate latent failure time models with competing risks under the generalized FGM copula. *Computat Stat* 2018; **33**: 1293–1323.
27. Heckman JJ and Honoré BE. The identifiability of the competing risks model. *Biometrika* 1989; **76**: 325–330.
28. Abbring JH and Van den Berg GJ. The identifiability of the mixed proportional hazards competing risks model. *J Royal Stat Soc: Series B (Statistical Methodology)* 2003; **65**: 701–710.
29. Scheike TH and Zhang MJ. Flexible competing risks regression modeling and goodness-of-fit. *Lifetime Data Analys* 2008; **14**: 464.
30. Sfumato P, Filleron T, et al. Gofcrr: A R package for assessing goodness-of-fit in proportional (sub) distributions hazards regression models. *Comput Meth Program Biomed* 2019; **177**: 269–275
31. Chen HY, Yu SL, Chen CH, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *New Engl J Med* 2007; **356**: 11–20.
32. Rondeau V, Pignon JP and Michiels S. A joint model for dependence between clustered times to tumour progression and deaths: a meta-analysis of chemotherapy in head and neck cancer. *Stat Meth Med Res* 2015; **24**: 711–729.
33. Lee M, Ha ID and Lee Y. Frailty modeling for clustered competing risks data with missing cause of failure. *Stat Meth Med Res* 2017; **26**: 356–373.
34. Emura T, Nakatochi M, Murotani K, et al. A joint frailty-copula model between tumour progression and death for meta-analysis. *Stat Meth Med Res* 2017; **26**: 2649–2666.
35. Peng M, Xiang L and Wang S. Semiparametric regression analysis of clustered survival data with semi-competing risks. *Computat Stat Data Analys* 2018; **124**: 53–70.
36. Brazauskas R and Le-Rademacher J. Methods for generating paired competing risks data. *Comput Meth Programs Biomed* 2016; **135**: 199–207.
37. Ha ID, Christian NJ, Jeong JH, et al. Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties. *Stat Meth Med Res* 2016; **25**: 2488–2505.
38. Emura T, Matsui S and Rondeau V. *Survival analysis with correlated endpoints, joint frailty-copula models, JSS Research series in statistics*. Singapore: Springer, 2019.
39. Ha ID, Noh M, Kim J, et al. frailtyHL: frailty models using h-likelihood, CRAN, version 2.1. 2018, <http://CRAN.Rproject.org/package=frailtyHL>.
40. Ramsay J. Monotone regression spline in action. *Stat Sci* 1988; **3**: 425–461.
41. Sylvester RJ, van der Meijden AP, Oosterlinck W, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur Urology* 2006; **49**: 466–477.
42. Scheike T. timereg: flexible regression models for survival data, CRAN, version 1.9.4. 2019.
43. Emura T, Matsui S and Chen HY. compound.Cox: univariate feature selection and compound covariate for predicting survival. *Comput Meth Programs Biomed* 2019; **168**: 21–37.
44. Park J, Bakoyannis G and Yiannoutsos C. Semiparametric competing risks regression under interval censoring using the R package intccr. *Comput Meth Programs Biomed* 2019; **173**: 167–176.
45. Sabatier R, Finetti P, Adelaide J, et al. Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer. *PLoS One* 2011; **6**: e27656.

## Appendix I. Proof of Theorem 2

We first rewrite the condition  $\lambda_1^{Sub}(t) = \lambda_1(t) \forall t \geq 0$  as  $f_1^{Sub}(t)S_1(t) = f_1(t)\{1 - F_1^{Sub}(t)\} \forall t \geq 0$ .  
The right-hand-side of this equation can be written as

$$\begin{aligned} f_1(t)\{1 - F_1^{Sub}(t)\} &= f_1(t)\{S_1(t) + \Pr(X \leq t, X > Y)\} \\ &= f_1^{Sub}(t)S_1(t) + \{f_1(t) - f_1^{Sub}(t)\}S_1(t) + f_1(t)\Pr(X \leq t, X > Y) \end{aligned}$$

Hence,  $\lambda_1^{Sub}(t) = \lambda_1(t) \forall t \geq 0$  is equivalent to

$$\{f_1(t) - f_1^{Sub}(t)\}S_1(t) + f_1(t)\Pr(X \leq t, X > Y) = 0 \quad \forall t \quad (9)$$

Since  $S_1(t)$  is non-increasing in  $t$ ,  $\exists t^* \in [0, \infty]$  such that  $S_1(t) = 0$  for  $\forall t \geq t^*$ . Note that  $S_1(t) = 0$  implies  $f_1(t) = 0$ . Hence equation (9) holds for  $\forall t \geq t^*$ . On the other hand, for  $\forall t < t^*$ , we have a positive value  $S_1(t) > 0$ . Thus, a necessary condition for equation (9) is  $\{f_1(t) - f_1^{Sub}(t)\} = 0 \forall t < t^*$ . This is also a sufficient condition for equation (9) since

$$\frac{d}{dt}\Pr(X \leq t, X > Y) = f_1(t) - f_1^{Sub}(t) = 0 \quad \forall t < t^* \quad \text{iff} \quad \Pr(X \leq t, X > Y) = 0 \quad \forall t < t^*$$

Hence,  $\lambda_1^{Sub}(t) = \lambda_1(t) \forall t \geq 0$  is equivalent to  $\Pr(X \leq t, X > Y) = 0 \quad \forall t < t^*$ . The proof complete since

$$\Pr(X \leq t, X > Y) = 0 \quad \forall t < t^* \quad \text{iff} \quad \Pr(X > Y) = 0 \quad \text{iff} \quad \Pr(X \leq Y) = 1$$

## Appendix 2. Semiparametric MLE under the marginal model

We introduce the marginal regression method of Chen<sup>8</sup> to compute the MLE  $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\lambda}_{10}, \hat{\lambda}_{20})$  based on the dataset  $(T_j, \delta_{1j}, \delta_{2j}, \mathbf{Z}_j), j = 1, 2, \dots, n$ . Under the marginal Cox models of section 4.2, the CS hazard functions are

$$\begin{aligned} \lambda_1^{CS}(t|\mathbf{Z}_j) &= \lambda_{10}(t)\exp(\boldsymbol{\beta}'_1\mathbf{Z}_j)\eta_{1j}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_{10}, \Lambda_{20}|\theta), \\ \lambda_2^{CS}(t|\mathbf{Z}_j) &= \lambda_{20}(t)\exp(\boldsymbol{\beta}'_2\mathbf{Z}_j)\eta_{2j}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_{10}, \Lambda_{20}|\theta) \end{aligned}$$

where

$$\begin{aligned} \eta_{1j}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_{10}, \Lambda_{20}|\theta) &= \frac{D_\theta^{[1,0]}\{\Lambda_{1j}(t), \Lambda_{2j}(t)\}}{D_\theta\{\Lambda_{1j}(t), \Lambda_{2j}(t)\}}, \\ \eta_{2j}(t; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_{10}, \Lambda_{20}|\theta) &= \frac{D_\theta^{[0,1]}\{\Lambda_{1j}(t), \Lambda_{2j}(t)\}}{D_\theta\{\Lambda_{1j}(t), \Lambda_{2j}(t)\}} \end{aligned}$$

where  $\Lambda_{kj}(t) \equiv \Lambda_k(t|\mathbf{Z}_j)$  for  $k = 1$  and  $2$ . As in Chen,<sup>8</sup> we obtain the log-likelihood

$$\begin{aligned} \ell(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{10}, \lambda_{20}|\theta) &= \sum_j \delta_{1j}[\boldsymbol{\beta}'_1\mathbf{Z}_j + \log\eta_{1j}(T_j; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_{10}, \Lambda_{20}|\theta) + \log\lambda_{10}(T_j)] \\ &\quad + \sum_j (1 - \delta_{2j})[\boldsymbol{\beta}'_2\mathbf{Z}_j + \log\eta_{2j}(T_j; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_{10}, \Lambda_{20}|\theta) + \log\lambda_{20}(T_j)] \\ &\quad - \sum_j D_\theta[\Lambda_{1j}(T_j), \Lambda_{2j}(T_j)] \end{aligned}$$

For a given  $\theta$ , the MLE is obtained as  $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\lambda}_{10}, \hat{\lambda}_{20}) = \text{argmax}_{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{10}, \lambda_{20}} \ell(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{10}, \lambda_{20}|\theta)$ , where we treat  $\Lambda_{k0}$  as an increasing step function with jumps sizes  $\lambda_{10}(T_j) = d\Lambda_{10}(T_j) = \Lambda_{10}(T_j) - \Lambda_{10}(T_j - dt)$  for  $\delta_{1j} = 1$  and  $\lambda_{20}(T_j) = d\Lambda_{20}(T_j) = \Lambda_{20}(T_j) - \Lambda_{20}(T_j - dt)$  for  $\delta_{2j} = 0$ . Currently, there seems no software package to implement the computation of the MLE.

For a special case of  $C_j = \infty$  and one-dimensional covariate  $\mathbf{Z}_j = Z_j$ , one can compute the MLE by the R function `dependCox.reg()` in `compound.Cox` R package.<sup>43</sup> The value of  $\theta$  should be given by user. Multiple covariates and cross-validated estimate of  $\theta$  can be handled by `dependCox.reg.CV()`; however, the estimates are obtained from univariate analyses and they are not equal to  $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\lambda}_{10}, \hat{\lambda}_{20})$ . See Chapter 5 of

Emura and Chen<sup>14</sup> for details. We illustrate the `dependCox.reg()` function through the lung cancer data. Given below are the input and output in the R console.

```
> data(Lung)
>
> t.vec = Lung$t.vec[Lung$strain == TRUE]
> d.vec = Lung$d.vec[Lung$strain == TRUE]
> X.vec = Lung$ZNF264[Lung$strain == TRUE]
>
> theta = 8
> dependCox.reg(t.vec,d.vec,X.vec,censor.reg = TRUE,alpha = theta,baseline = TRUE)

$surv.reg
beta      SE      Z      P
0.5613137475 0.1583489783 3.5447891957 0.0003929272

$cursor.reg
beta      SE      Z      P
0.453470429 0.151979571 2.983759106 0.002847308

$baseline
[1] 0.006091766 0.012249470 0.018474552 0.025089564 0.025089564 0.032470851
...
[61] 1.006589532 1.006589532 1.006589532

$scen.baseline
[1] 0.00000000 0.00000000 0.00000000 0.00000000 0.01387123 0.01387123
...
[61] 1.83754289 2.08436566 2.72019869
```

We see the estimate  $\hat{\beta}_1 = 0.5613137475$  (along with the SE, Z-value, and P-value), a regression coefficient for Event 1 (death). Similarly, we see the estimate  $\hat{\beta}_2 = 0.453470429$ , a regression coefficient for Event 2 (censoring). We also see the cumulative baseline hazard function  $\hat{\Lambda}_{k0}(T_j)$  for  $j = 1, 2, \dots, n$  and Event  $k$ . Since the sample size is  $n = 63$ , we have  $\hat{\Lambda}_{k0}(T_{(j)})$  for 63 ordered time points  $T_{(1)} < \dots < T_{(63)}$ . The value of  $\theta$  must be given by user as seen from the input “theta = 8”.

### Appendix 3. Simulations for the marginal regression method

This appendix conducts simulation studies to demonstrate that the marginal regression methods for clustered competing risks data (section 5.1) are reliable. To this end, we generate data from a joint frailty-copula model and estimate the parameters. We then check the consistency between the estimates and the parameters for 500 repetitions.

#### A.3.1 Simulation designs

Following Emura et al.,<sup>34</sup> we consider two different scenarios:

**Scenario (I):**  $G = 5$  and  $N_i = 100$  (or 200) for  $i = 1, 2, \dots, 5$ .

**Scenario (II):**  $G = 30$  and  $N_i = 10$  (or 20) for  $i = 1, 2, \dots, 30$ .

The case of  $G = 5$  corresponds to small number of clusters (studies), which is common in meta-analyses. For example, Sabatier et al.<sup>45</sup> combined  $G = 6$  independent studies to examine the effect of *ECRG4* expression on survival. In the ovarian cancer research, Emura et al.<sup>34</sup> analyzed the effect of *CXCL12* expression to both tumor recurrence and death based on  $G = 4$  independent studies. The case of  $G = 30$  corresponds to a larger amount of clusters with smaller number of subjects. Our bladder cancer data example (section 5.2) is  $G = 21$ .

**Table 4.** Simulation results under Scenario (I) ( $G = 5$  studies) based on 500 replications.

	Parameter	$N_i = 100$				$N_i = 200$			
		Mean	SD	SE	CP	Mean	SD	SE	CP
CEN = 16% $\theta = 2$	$\beta_1 = 1$	1.003	0.212	0.217	0.95	1.003	0.148	0.151	0.96
	$\beta_2 = 1$	1.013	0.210	0.217	0.96	1.003	0.146	0.151	0.97
	$\eta = 0.5$	0.400	0.263	0.245	0.89	0.396	0.282	0.239	0.83
	$\kappa_1$	60.020	185.132	–	–	28.880	108.023	–	–
	$\kappa_2$	79.420	218.246	–	–	26.480	86.495	–	–
CEN = 32% $\theta = 2$	$\beta_1 = -1$	-1.002	0.254	0.249	0.95	-1.000	0.169	0.174	0.95
	$\beta_2 = -1$	-0.995	0.253	0.248	0.95	-0.995	0.172	0.174	0.96
	$\eta = 0.5$	0.397	0.264	0.244	0.89	0.395	0.283	0.239	0.83
	$\kappa_1$	264.340	397.644	–	–	126.320	284.601	–	–
	$\kappa_2$	229.340	373.205	–	–	105.720	245.626	–	–
CEN = 18% $\theta = 6$	$\beta_1 = 1$	1.007	0.193	0.198	0.95	1.005	0.139	0.138	0.95
	$\beta_2 = 1$	1.015	0.191	0.197	0.96	1.006	0.138	0.138	0.95
	$\eta = 0.5$	0.402	0.266	0.247	0.88	0.399	0.285	0.241	0.83
	$\kappa_1$	117.480	281.349	–	–	55.800	178.341	–	–
	$\kappa_2$	136.080	292.132	–	–	52.200	175.123	–	–
CEN = 36% $\theta = 6$	$\beta_1 = -1$	-1.002	0.233	0.231	0.95	-0.999	0.157	0.161	0.95
	$\beta_2 = -1$	-0.997	0.230	0.231	0.95	-0.998	0.158	0.161	0.95
	$\eta = 0.5$	0.400	0.268	0.248	0.89	0.396	0.281	0.240	0.84
	$\kappa_1$	306.740	420.877	–	–	126.040	278.419	–	–
	$\kappa_2$	285.940	412.046	–	–	127.040	277.603	–	–

CEN: the censoring percentage; SD: the sample standard deviation of the estimates; SE: the average of the standard errors; CP: the coverage probability of the 95% CIs.

**Table 5.** Simulation results under Scenario (II) ( $G = 30$  studies) based on 500 replications.

	Parameter	$N_i = 10$				$N_i = 20$			
		Mean	SD	SE	CP	Mean	SD	SE	CP
CEN = 16% $\theta = 2$	$\beta_1 = 1$	0.993	0.306	0.296	0.95	1.019	0.193	0.201	0.96
	$\beta_2 = 1$	0.991	0.313	0.295	0.93	1.012	0.204	0.202	0.94
	$\eta = 0.5$	0.489	0.165	0.158	0.95	0.482	0.144	0.136	0.93
	$\kappa_1$	73.320	176.789	–	–	29.520	64.846	–	–
	$\kappa_2$	75.720	190.648	–	–	27.520	57.491	–	–
CEN = 32% $\theta = 2$	$\beta_1 = -1$	-1.015	0.360	0.331	0.91	-0.991	0.221	0.229	0.96
	$\beta_2 = -1$	-1.024	0.354	0.331	0.92	-0.994	0.228	0.229	0.95
	$\eta = 0.5$	0.493	0.176	0.168	0.97	0.480	0.147	0.139	0.94
	$\kappa_1$	262.420	397.434	–	–	151.000	311.598	–	–
	$\kappa_2$	267.220	401.863	–	–	143.200	299.575	–	–
CEN = 18% $\theta = 6$	$\beta_1 = 1$	0.993	0.287	0.274	0.94	1.013	0.178	0.185	0.96
	$\beta_2 = 1$	0.993	0.289	0.274	0.93	1.011	0.180	0.185	0.96
	$\eta = 0.5$	0.489	0.165	0.158	0.96	0.484	0.144	0.135	0.93
	$\kappa_1$	137.120	285.962	–	–	49.440	143.696	–	–
	$\kappa_2$	126.920	273.864	–	–	54.840	143.462	–	–
CEN = 36% $\theta = 6$	$\beta_1 = -1$	-1.023	0.342	0.307	0.91	-0.999	0.206	0.213	0.96
	$\beta_2 = -1$	-1.029	0.341	0.308	0.91	-1.001	0.207	0.213	0.95
	$\eta = 0.5$	0.494	0.177	0.171	0.97	0.482	0.148	0.141	0.94
	$\kappa_1$	343.880	435.267	–	–	209.380	363.457	–	–
	$\kappa_2$	341.680	440.912	–	–	198.780	354.868	–	–

CEN: Censoring percentage; SD: sample standard deviation of the estimates; SE: average of the standard errors; CP: coverage probability of the 95% CIs.

For each cluster  $i = 1, 2, \dots, G$ , a frailty term  $u_i$  follows a gamma distribution with mean 1 and variance  $\eta = 0.5$ . For each subject  $j = 1, 2, \dots, N_i$ , a covariate  $Z_{ij}$  follows a uniform distribution on the unit interval  $(0, 1)$ . Given  $u_i$  and  $Z_{ij}$ , the joint distribution of  $X_{ij}$  (Event 1) and  $Y_{ij}$  (Event 2) follows a joint frailty-copula model specified by

$$\Pr(X_{ij} > x, Y_{ij} > y|u_i) = [\exp\{\theta\Lambda_{1ij}(x|u_i)\} + \exp\{\theta\Lambda_{2ij}(y|u_i)\} - 1]^{-1/\theta}$$

where  $\Lambda_{1ij}(x|u_i) = u_i\Lambda_{10}(x)\exp(\beta_1 Z_{ij})$  and  $\Lambda_{2ij}(y|u_i) = u_i^2\Lambda_{20}(y)\exp(\beta_2 Z_{ij})$  are the cumulative hazards with  $d\Lambda_{10}(x)/dx = \lambda_{10}(x) = 1$  and  $d\Lambda_{20}(y)/dy = \lambda_{20}(y) = 1$ . Here, we introduce dependence between Events 1 and 2 by choosing  $\theta = 2$  ( $\theta = 6$ ) that corresponds to Kendall's tau equals to 0.5 (0.75). To avoid the non-identifiability issue on the competing risks data,<sup>22</sup> we assumed the dependence parameter  $\theta$  to be known. We also assumed  $\alpha = 1$  is known and not estimated. The independent censoring time  $C_{ij}$  follows a uniform distribution on the interval  $(0, 5)$  that yields around 16–36% censored subjects. After generating data, we fit the joint model by utilizing the R function `cmprskCox.reg()` in the `joint.Cox` R package. Our simulations are based on 500 replications.

### A.3.2 Simulation results

Table 4 shows the simulation results under Scenario (I). The parameter estimates are nearly unbiased for regression parameters  $\beta_1$  and  $\beta_2$ . The standard deviations (SDs) of the estimates decrease as the numbers of subjects increase from  $N_i = 100$  to 200. In addition, the average standard errors (SEs) accurately approximate the SDs. Consequently, the coverage probabilities (CPs) of the 95% CIs are all close to the nominal confidence level of 0.95.

Table 4 reveals that the estimates of the frailty parameter  $\eta$  are biased. Also, the average SEs are lower than the SDs. Consequently, the resulting CPs of the 95% CIs do not reach the nominal confidence level of 0.95. These problems do not vanish even if the number of subjects increases from  $N_i = 100$  to 200 (Table 4). This is because the bias is due to the small number of clusters ( $G = 5$ ). Fortunately, the problems of the bias and CPs vanish when the number of clusters is large ( $G = 30$ , Table 5).

Table 5 shows the simulation results under Scenario (II). Similar conclusions can be drawn for the estimation of  $\beta_1$  and  $\beta_2$  from those of Scenario (I). On the other hand, unlike Scenario (I), the estimates of  $\eta$  are nearly unbiased. The average SEs are close to the SDs, and hence the CPs of the 95% CIs are close to the nominal confidence level of 0.95.