Dependent competing risks: Cause elimination and its impact on survival

Dimitrina S. Dimitrova*, Steven Haberman and Vladimir K. Kaishev

Cass Business School, City University London, UK

Abstract

The dependent competing risks model of human mortality is considered, assuming that the dependence between lifetimes is modelled by a multivariate copula function. The effect on overall survival of removing one or more causes of death is explored under two alternative definitions of removal, *ignoring* the causes and *eliminating* them. Under the two definitions of removal, expressions for the overall survival functions in terms of the specified copula (density) and the net (marginal) survival functions, are given. The net survival functions are obtained as a solution to a system of non-linear differential equations, which relates them through the specified copula (derivatives) to the crude (sub-) survival functions, estimated from data. The overall survival functions in a model with four competing risks, cancer, cardiovascular diseases, respiratory diseases and all other causes grouped together have been implemented and evaluated, based on cause specific mortality data for England and Wales published by the Office for National Statistics, for year 2007. We show that the two alternative definitions of removal of a cause of death have different effects on the overall survival and in particular on the life expectancy at birth and at age 65, when one, two or three of the competing causes are removed. An important conclusion, is that the eliminating definition is better suited for practical use in competing risks applications, since it is more intuitive, and it suffices to consider only positive dependence between the lifetimes which is not the case under the alternative ignoring definition.

Keywords: dependent competing risks model, lifetimes, failure times, overall survival function, copula functions, cause elimination, cause removal

*Faculty of Actuarial Science and Insurance, Cass Business School, City University London, 106 Bunhill Row, EC1Y 8TZ London, UK. Email: d.dimitrova@city.ac.uk

1 Introduction

In the competing risks model, a group of individuals (units) is subject to the simultaneous operation of a set of competing risks which cause death (failure). It is assumed that each individual can die from any one of the causes and that there are corresponding lifetime random variables attached to him/her at birth. This model has been widely studied in the (bio)statistical, medical, actuarial and demographic literature, under the assumption of independence of the corresponding lifetimes. Important contributions to the subject, to mention only a few, are the books by Pintilie (2006), Kalbfleisch and Prentice (2002), Crowder (2001), Lawless (2003), Bowers et al. (1997) and Elandt-Johnson and Johnson (1980), the recent overview by Lindqvist (2007) and papers by Solari et al. (2008), Salinas-Torres et al. (2002) and Bryant and Dignam (2004), where various aspects and problems related to the competing risks model such as statistical methods for estimating (sub-) survival functions, marginal survival functions and related inference are considered.

A considerable amount of work has been devoted to the competing risks model and its application in economics, reliability, medicine and actuarial science, under the assumption of dependence of the competing risks lifetimes. Important early contribution in this strand of literature are the papers by Elandt-Johnson (1976), and also by Yashin et al. (1986) who consider conditional independence of the times to death, given an assumed stochastic covariate process. Tsiatis (1975) shows that it is impossible to identify the dependence structure underlying the (dependent) joint distribution of the competing risks failure times and their (marginal) distributions, based on observed data. This is the well-known, unresolvable problem of identifiability. It has been overcome in more recent work by simply assuming that the dependence structure is known. With this approach, Zheng and Klein (1995) propose the so called copula-graphic estimator of the marginal distributions for dependent competing risks, assuming dependence is represented by a known copula with known parameters. Recently, under the similar assumption of a completely specified underlying copula, Chen (2010) develops a non-parametric maximum likelihood estimation of the marginal semiparametric transformation models. Lo and Wilke (2010) apply a risk pooling approach combined with the two-dimensional copula-graphic estimator of

Zheng and Klein (1995) in order to estimate the marginal survival functions in a multivariate dependent competing risks model with an assumed Archimedean copula. They test their model on unemployment duration data. EM-based estimation of sub-distribution functions under the assumption that some of the competing causes are masked, has been considered by Craiu and Reiser (2006). Bounds in a dependent competing risks models with interval outcome data have been derived by Honoré and Lleras-Muney (2006), who apply their model in estimating changes in cancer and cardiovascular mortality in USA. Recently, Lindqvist and Skogsrud (2009) has focused at modelling dependent competing risks in reliability, by considering first passage times of Wiener processes. A useful survey of statistical methods for dependent competing risks is provided by Moeschberger and Klein (1995).

The dependent competing risks model of human mortality, under the assumption of a (known) underlying copula function, has been considered by Carriere (1994, 1995) and Escarela and Carriere (2003) and more recently by Kaishev et al. (2007). Carriere (1994) and Escarela and Carriere (2003) have modelled dependence between two failure times by a two dimensional copula. In Escarela and Carriere (2003), the bi-variate Frank copula was fitted to a prostate cancer data set. Carriere (1994) was the first to use a bi-variate Gaussian copula in order to model the effect of complete removal of one of two competing causes of death on human mortality. However, the mortality data used by Carriere (1994) was not complete with respect to older ages and therefore, it was not possible to calculate such important survival characteristics as expected lifetimes and draw relevant conclusions.

This deficiency has been overcome in the paper by Kaishev et al. (2007) who close the life table by applying a method of spline extrapolation up to a limiting age 120. They have extended further the work of Carriere (1994), considering a multidimensional copula model for the joint distribution of the lifetimes. The model has been tested on the example of up to four competing causes of death, (cancer, heart diseases, respiratory diseases and other causes grouped together), based on the US general population cause specific mortality data set, provided by the National Center for Health Statistics, NCHS (1999). Several alternative four dimensional copula models underlying the joint distribution of the life times have been explored: the Gaussian copula, the Student t-copula, the Frank copula and the Plackett copula.

The impact of removal of one, two or three of the competing causes of death on the overall survival function and the life expectancy, which have utmost importance in medical, biostatistical and actuarial applications, has been studied.

In the paper by Kaishev et al. (2007), as well as in the earlier paper by Carriere (1994), it has been assumed that deaths by a cause are removed by simply ignoring that cause, i.e., by omitting the corresponding lifetime random variable from the vector of lifetimes considered. For this reason, removal of a cause of death under this definition, can be described more precisely as *ignoring* the cause. However, as pointed out by Kaishev et al. (2007) and also earlier, by Elandt-Johnson (1976), an alternative definition of removal of a certain cause may be given by considering the limiting distribution of the vector of lifetimes, given that the lifetime with respect to the removed cause tends to infinity, or more realistically to the limiting age. In other words, under this definition, it is assumed that deaths from the removed cause would not occur and all individuals would survive an infinitely long time (in reality up to the limiting age) with respect to that cause. In what follows, we will call this type of removal of deaths from a particular cause, *elimination* of that cause. As pointed out by Kaishev et al. (2007), this alternative definition is more intuitive and easy to interpret, but leads to more complex expressions for the limiting survival distribution, under the assumption that dependence is modelled by a suitable copula.

The purpose of this paper is to explore the two alternative definitions of *ignoring* a cause and *eliminating* that cause, within the multivariate copula dependent competing risks model. We compare and contrast the two definitions, based on UK cause specific mortality data for year 2007, provided by the Office for National Statistics, ONS (2008), which includes deaths from cancer, heart disease, respiratory diseases and all other causes grouped together. We show that the choice of definition of cause removal has a significant effect on the overall survival function and the life expectancy at birth and at age 65, in the cases where one, two or three of the competing causes of death are simultaneously removed. It is demonstrated that the *eliminating* definition is easier and more intuitive to interpret and does not necessarily require the use of comprehensive copulas and also that the complexity related to its implementation can be overcome without difficulty. Therefore, an important conclusion of the current work is that the *eliminating* defi-

nition is preferable for practical use compared to the *ignoring* definition, studied earlier in the papers by Carriere (1994) and Kaishev et al. (2007).

A second purpose of the paper is to demonstrate that, given a known copula, the approach of estimating the net survival functions by solving a system of differential equations, first considered by Carriere (1994) in the two dimensional case, and later extended by Kaishev et al. (2007) to the multivariate case, is numerically accurate and viable. Recently, this has been questioned by Lo and Wilke (2010) who have instead used the copula-graphic estimator of Zheng and Klein (1995) to estimate the net survival functions in the special case of (exchangeable) multivariate Archimedean copulas. It can be argued that in practice it is restrictive to assume symmetry in the dependence structure of competing risks failure times. Contrary to this, our approach is general and allows to incorporate any copula model for the competing-risk failure times distribution.

The paper is organized as follows. In section 2, we introduce the dependent competing risks model under the assumption that dependence between the competing risks lifetimes is modelled by a suitable copula function. We summarize the methodology for obtaining net survival functions, given estimates of the crude survival functions, considered earlier by Carriere (1994) and Kaishev et al. (2007). In section 3, we give two alternative definitions of removal of a cause of death, *ignoring* and *eliminating* and provide expressions for the overall survival functions when one or more causes are removed. In section 4, we implement the definitions numerically and compare the effect they have on the overall survival and on the life expectancy. Section 5 provides some conclusions and comments.

2 The dependent competing risks model

As pointed out by a number of authors, see e.g., Hooker and Longley-Cook (1957), Carriere (1994), Kalbfleisch and Prentice (2002), Valdez (2001), Fukumoto (2005), Lindqvist (2007), Lindqvist and Skogsrud (2009), risks in many real life applications tend to be dependent. In particular, as established in studying disease interactions (see e.g., Kaput et al. 1994, Weir 2005, Lobo 2008), diseases may be jointly caused by the interaction of particular genes. For example, as pointed out by Kaput et al. (1994), high levels of dietary fat, regulated and characterized by certain genes, jointly enhance

the severity of certain cancers, obesity and cardiovascular diseases. Therefore, successful treatment of obesity, may lead to considerable reduction in the number of deaths from certain types of cancer and atherosclerosis. Weir (2005) has studied the interaction between cardiovascular disease (CVD) and chronic kidney disease (CKD) in patients with CKD and has explained the increased risk for CVD in patients with CKD. The paper by Lobo (2008) is devoted to understanding epistatic interactions between genes as the key to understanding complex diseases, such as Alzheimer's disease, diabetes, cardiovascular disease, and cancer. These and other studies in the medical literature suggest that, by reducing (or completely removing) deaths from one disease, it is possible to significantly improve mortality rates from the related (interacting) disease. In terms of lifetimes, this means that the lifetimes of interacting diseases are related (mutually dependent), and this dependence, which characterizes the overall survival from such causes, can be represented and studied under the copula-dependent competing risks model considered in this section.

The copula-dependent competing risks model of human mortality has recently been considered by Kaishev et al. (2007) where a detailed account of its properties, model assumptions and parameter estimation can be found. For our purpose of considering the model uncertainty with respect to the definition of cause elimination, we will briefly introduce the model and recall its basic characteristics.

Consider a group of individuals, exposed to m competing causes of death. It is assumed that each individual may die from any single one of the$m$ causes. To make the problem more formally tractable it is assumed that, at birth, each individual is assigned a vector of potential life times $T_1, \ldots, T_m$, $0 \leq T_j < \infty$, $j = 1, \ldots, m$, if he/she were to die from each one of the m causes. Obviously, the actual lifetime span is the minimum of all the $T_1, \ldots, T_m$. Thus, it is clear that under this model the lifetimes $T_1, \ldots, T_m$ are unobservable and we can only observe the $\min(T_1, \ldots, T_m)$. In the classical competing risks model the random variables $T_1, \ldots, T_m$ are assumed independent, whereas here we will be interested in their (dependent) joint survival distribution function

$$S(t_1, \ldots, t_m) = \Pr(T_1 > t_1, \ldots, T_m > t_m) \tag{1}$$

which is assumed absolutely continuous and where $t_j \geq 0$, for $j = 1, \ldots, m$. In what follows, we will also need the marginal survival functions $S'^{(j)}(t) = \Pr(T_j > t)$, $j = 1, \ldots, m$, associated with $S(t_1, \ldots, t_m)$, which we call *net* survival functions. As we will see, $S'^{(j)}(t)$ are the target quantities in our study since, if we know them we can identify and calculate the joint survival function $S(t_1, \ldots, t_m)$ and hence, evaluate the overall survival function $S(t, \ldots, t)$, under some appropriate assumptions on the dependence structure underlying (1). Note that, $S'^{(j)}(t)$, $j = 1, \ldots, m$ are not observable. Let us recall that the classical model of independence of the r.v.s $T_1, \ldots, T_m$ implies that

$$S(t_1, \ldots, t_m) = S'^{(1)}(t_1) \times \ldots \times S'^{(m)}(t_m).$$

The overall survival of an individual, under the dependent competing risks model assumptions, is defined by the random variable $T = \min(T_1, \ldots, T_m)$, and we will be interested in modelling the overall survival function,

$$S(t, \ldots, t) = \Pr(T_1 > t, \ldots, T_m > t) = \Pr(T > t)$$

where $t \geq 0$. In order to do so, one can apply the celebrated theorem of Sklar and express the survival function $S(t_1, \ldots, t_m)$ in terms of the net (marginal) survival functions $S'^{(j)}(t)$ and a suitable copula function, $C(u_1, \ldots, u_m)$, $0 \leq u_i \leq 1$, $i = 1, \ldots, m$ which captures the dependence structure, underlying the multivariate survival distribution of the random vector $T_1, \ldots, T_m$.

Copula functions have become a well established tool for modelling stochastic dependence and their properties are well documented in the monographs by Nelsen (2006), Joe (1997) and Cherubini et al. (2004). There are numerous copula related papers scattered throughout the statistical, financial and actuarial journals and some relevant references can be extracted from the CopulaWiki web page `http://140.78.127.5/mediawiki/index.php/Main$\_$Page`. For a concise summary of the main properties of copulas, relevant to the multivariate dependent competing risks model of human mortality, see Kaishev et al. (2007).

Having fixed a suitable copula, we can write

$$S(t_1, \ldots, t_m) = C\left(S'^{(1)}(t_1), \ldots, S'^{(m)}(t_m)\right), \tag{2}$$

from where we can also evaluate the overall survival function

$$S(t,\ldots,t) = C\left(S^{'(1)}(t),\ldots,S^{'(m)}(t)\right), \tag{3}$$

if the net survival functions $S^{'(j)}(t_j)$, $j = 1,\ldots,m$ were known. In order to find them, we may use the relationship between $S^{'(j)}(t)$ and the so called crude survival functions, $S^{(j)}(t)$, $j = 1,\ldots,m$. The crude survival function $S^{(j)}(t)$ is defined as the survival function with respect to the j-th cause of death, due to which death actually occurs, i.e.,

$$S^{(j)}(t) = \Pr\left(\min\left(T_1,\ldots,T_m\right) > t, \min\left(T_1,\ldots,T_m\right) = T_j\right)$$

The survival function $S^{(j)}(t)$ is called *crude*, since it reflects the observed mortality of an individual and hence, may be estimated, from the observed mortality data of a population, as will be illustrated in section 4. In the biostatistics literature the crude survival function $S^{(j)}(t)$ is sometimes called the *sub-survival function* and the related cumulative distribution function is named *sub-distribution function* or *cumulative incidence function*. A considerable amount of literature exists which focuses at the use and estimation of the latter functions in the context of competing risks (see e.g. Gaynor et al. 1993, Lin 1997, Gooley 1999, Kalbfleisch and Prentice 2002, Lawless 2003, Craiu and Reiser 2006, Jeong and Fine 2007), where further references can be traced down.

It is not difficult to see that

$$S(t,\ldots,t) = S^{(1)}(t) + \ldots. + S^{(m)}(t) \tag{4}$$

since the events $\min\left(T_1,\ldots,T_m\right) = T_j$, $j = 1,\ldots,m$ are mutually exclusive. This obviously suggests that $S^{(j)}(0) < 1$, $j = 1,\ldots,m$ and the crude survival functions are defective.

As shown by Carriere (1994), under the assumption of differentiability of $C\left(u_1,\ldots,u_m\right)$ with respect to $u_j \in (0,1)$ and of $S^{'(j)}\left(t_j\right)$ with respect to $t_j > 0$, for$t > 0$, the following system of differential equations relates the

crude and net survival functions

$$\begin{aligned}
\frac{d}{dt}S^{(1)}(t) &= C_1\left(S'^{(1)}(t),\ldots,S'^{(m)}(t)\right)\times\frac{d}{dt}S'^{(1)}(t)\\
\frac{d}{dt}S^{(2)}(t) &= C_2\left(S'^{(1)}(t),\ldots,S'^{(m)}(t)\right)\times\frac{d}{dt}S'^{(2)}(t)\\
&\vdots\\
\frac{d}{dt}S^{(m)}(t) &= C_m\left(S'^{(1)}(t),\ldots,S'^{(m)}(t)\right)\times\frac{d}{dt}S'^{(m)}(t)
\end{aligned} \tag{5}$$

where $C_j\left(u_1,\ldots,u_m\right)=\frac{\partial}{\partial u_j}C\left(u_1,\ldots,u_m\right)$, $j=1,\ldots,m$.

It is important to note that (5) is a system of nonlinear, differential equations which may be solved with respect to the net survival functions $S'^{(j)}(t)$, given a suitable copula and estimates of the crude survival functions $S^{(j)}(t)$, $j=1,\ldots,m$. The (approximate) numerical solution of (5) has been considered by Carriere (1994) in the two dimensional case, $m=2$. The choice of the copula function, C, the (spline) estimation of $S^{(j)}(t)$, $j=1,\ldots,m$ and the efficient numerical solution of (5) in the multivariate case, $m>2$ using *Mathematica* has been considered by Kaishev et al. (2007).

The derivatives with respect to time of the crude and net survival functions in (5) are actually the crude and net probability density functions of the r.v.s $T_1,T_2,\ldots,T_m$. We will denote these densities as $f^{(j)}(t)$ and $f'^{(j)}(t)$, $j=1,\ldots,m$, respectively.

Let us also note that equality (4) can be used as a check on the solution of (5). For this purpose, we can apply (3) to express the overall survival function on the left-hand side of (4) as

$$C\left(S'^{(1)}(t),\ldots,S'^{(m)}(t)\right)=S^{(1)}(t)+\ldots.+S^{(m)}(t)$$

where , $0\leq t\leq 120$.

Once the net survival functions are obtained, one can use (3) and evaluate the overall survival function which is of major interest in our investigation. More precisely, we will be interested in studying the effect of removal of a cause of death on the overall survival function, under two alternative definitions of removal, which will be introduced in the next section.

3 Removal of a cause of death

Our main interest in the paper is to investigate the effect of removing a cause of death, say indexed j, on the overall survival function $S(t,\ldots,t)$. This effect depends on the definition of removal and, as mentioned in the introduction, one can consider two alternative definitions, either *ignore* the cause or *eliminate* it. The two alternatives have been highlighted already in the early paper by Elandt-Johnson (1976). Under the first approach, deaths arising from the j-th cause are removed by simply ignoring the j-th cause and considering a modified version of the lifetime random variable T, defined as

$$T^{(-j)}_{\text{ignore}} = \min\left(T_1,\ldots,T_{j-1},T_{j+1},\ldots,T_m\right)$$

i.e., considering the marginal distribution

$$F\left(t_1,\ldots,t_{j-1},t_{j+1},\ldots,t_m\right) = \Pr\left(T_1 \le t_1,\ldots,T_{j-1} \le t_{j-1},T_{j+1} \le t_{j+1},\ldots,T_m \le t_m\right)$$

with overall survival function

$$\begin{aligned} S^{(-j)}_{\text{ignore}}(t) &= S(t,\ldots,t,0,t,\ldots,t) \\ &= \Pr\left(T_1 > t_1,\ldots,T_{j-1} > t_{j-1},T_{j+1} > t_{j+1},\ldots,T_m > t_m\right) \\ &= \Pr\left(T^{(-j)}_{\text{ignore}} > t\right), \end{aligned} \tag{6}$$

where $t = 0$ appears on the j-th position. Similarly, ignoring two causes, say the j-th and the k-th ones, $j \neq k$, would lead to considering the survival function

$$S^{(-j,-k)}_{\text{ignore}}(t) = S(t,\ldots,t,0,t,\ldots,t,0,t,\ldots,t). \tag{7}$$

Alternatively, the j-th cause of death, may be eliminated by considering the limiting distribution, conditional on $T_j \uparrow \infty$, of surviving from all other causes. Under this definition, the overall survival distribution function becomes

$$S^{(-j)}_{\text{eliminate}}(t) = \lim_{t_j\to\infty} \frac{S\left(t,\ldots,t,t_j,t,\ldots,t\right)}{S'^{(j)}\left(t_j\right)}. \tag{8}$$

Similarly, eliminating the j-th and the k-th cause, $j \neq k$, may be defined as considering the survival function

$$S^{(-j,-k)}_{\text{eliminate}}(t) = \lim_{t_j\to\infty,t_k\to\infty} \frac{S\left(t,\ldots,t,t_j,t,\ldots,t,t_k,t,\ldots,t\right)}{S'^{(j,k)}\left(t_j,t_k\right)}, \tag{9}$$

where $S'^{(j,k)}(t_j, t_k)$, is the marginal survival function with respect to the j-th and the k-th causes. Note that both expressions (7) and (9) directly generalize to the case of removing more than two competing risks.

The elimination definition allows for a more natural interpretation of the dependence between lifetimes and of the elimination of their corresponding causes, as will be illustrated numerically in the next section. To see this, assume that the j-th cause is strongly positively correlated with, say, the k-th cause. In this case, eliminating the j-th cause will mean that an individual is much more likely to survive to a longer time-horizon with respect to the k-th cause and more precisely, under perfect positive correlation, $T_j \uparrow \infty$ would lead to $T_k \uparrow \infty$, which is intuitive. On the other extreme, if T_j and T_k are perfectly negatively correlated, if $T_j \uparrow \infty$, then $T_k \downarrow 0$ which could be described as: elimination of the j-th cause would lead to increased mortality with respect to the k-th cause and hence, to decreased overall survival. Clearly, this is of little practical relevance since removal of a cause of death usually leads to improvement of the overall survival and for this reason, *elimination* should be considered only under non-negative correlation. Let us note that this is not the case for the alternative ignoring definition, under which both negative or positive correlations between lifetimes may produce improvements in the overall mortality, and worse mortality is not achievable, as confirmed numerically in section 4 in this paper on the example of the UK cause specific mortality data and also in Kaishev et al. (2007) for US data. Therefore, the requirements with respect to the copula functions are more stringent under the ignoring definition, since in order to cover the whole range, from perfectly negative to perfectly positive correlation, only comprehensive copulas may be used. It is also more difficult to give a meaningful interpretation of ignoring a cause under both negative and positive correlation between competing lifetimes.

It has to be noted that the elimination approach is confronted with the difficulty that the limiting conditional distributions in (8) and (9), may not always exist and if they exist, the evaluation of the overall survival function may in general be more complex. Based on a particular selection of copulas, we show in section 4 that, the numerical complexity added due to the change of definition of elimination may be successfully overcome. Let us also note that, in the case when $T_1, \ldots, T_m$ are assumed independent, the two approaches are equivalent (see Elandt-Johnson 1976).

While the somewhat simpler approach of ignoring a cause has been implemented and explored further in the papers by Carriere (1994) and more recently by Kaishev et al. (2007), to the best of our knowledge, the alternative approach of eliminating a cause of death has not been implemented and studied previously.

Our major goal in this paper will be to find representations, in terms of a suitable copula, of the survival functions $S^{(-j)}_{\text{ignore}}(t)$, $S^{(-j,-k)}_{\text{ignore}}(t)$,... and $S^{(-j)}_{\text{eliminate}}(t)$, $S^{(-j,-k)}_{\text{eliminate}}(t)$,... under the two alternative definitions of removal of a cause of death. This will allow us to quantify and compare the effect of removal of one or more causes, under the two alternative definitions, on life expectancy at birth and at age 65.

Applying Sklar's theorem one can express $S^{(-j)}_{\text{ignore}}(t)$ as defined in (6), in terms of a copula function as

$$S^{(-j)}_{\text{ignore}}(t) = C(S'^{(1)}(t), \ldots, S'^{(j-1)}(t), 1, S'^{(j+1)}(t), \ldots, S'^{(m)}(t)\Big), \quad (10)$$

where the marginal (net) survival function $S'^{(j)}(t) = \Pr(T_j > t)$ due to cause j, are found as solutions to the system of differential equations, (5), following the methodology described in section 2.

Similarly, one can write

$$S^{(-j,-k)}_{\text{ignore}}(t) = C\Big(S'^{(1)}(t), \ldots, S'^{(j-1)}(t), 1, S'^{(j+1)}(t), \\ \ldots, S'^{(k-1)}(t), 1, S'^{(k+1)}(t), \ldots, S'^{(m)}(t)\Big). \quad (11)$$

Alternatively, under the elimination approach, applying definition (8) the following expression for the overall survival function, given the j-th cause has been eliminated can be written

$$S^{(-j)}_{\text{eliminate}}(t) = \int_t^\infty \ldots \int_t^\infty c\left(S'^{(1)}(t_1), \ldots, S'^{(j-1)}(t_{j-1}), 0, S'^{(j+1)}(t_{j+1}), \right. \\ \left. \ldots, S'^{(m)}(t_m)\right) \times \prod_{i=1, i\neq j}^{m} f'^{(i)}(t) dt_1 \ldots dt_m, \quad (12)$$

where $c(u_1, \ldots, u_{j-1}, 0, u_{j+1}, \ldots, u_m)$, is the copula density, $f'^{(i)}(t)$, $i = 1, \ldots, m$ are the marginal (net) probability density functions, corresponding to each cause of death and the integral in (12) has dimension $m-1$.

Similarly, following (9),

$$S^{(-j,-k)}_{\text{eliminate}}(t) = \int_t^\infty \dots \int_t^\infty \frac{1}{C_{jk}(0,0)} c\left(S'^{(1)}(t_1), \dots, S'^{(j-1)}(t_{j-1}), 0, S'^{(j+1)}(t_{j+1}),\right.$$
$$\left. \dots, S'^{(k-1)}(t_{k-1}), 0, S'^{(k+1)}(t_{k+1}), \dots, S'^{(m)}(t_m)\right) \times \prod_{i=1, i\neq j, i\neq k}^{m} f'^{(i)}(t) dt_1 \dots dt_m, \tag{13}$$

where

$$C_{jk}(0,0) = \frac{\partial}{\partial u_j}\frac{\partial}{\partial u_k} C(1, \dots, 1, u_j, 1, \dots, 1, u_k, 1, \dots, 1)\,|_{u_j=0, u_k=0}.$$

It is easy to see how (11) and (13) generalize directly to the case of eliminating more than two causes, and therefore we will omit the corresponding formulae.

Comparing expressions (10) and (11), with (12) and (13), it can be seen that the latter are more complex and more difficult to evaluate. In order to evaluate (10) and (11), it is sufficient to compute the copula function C whereas, in order to evaluate (12) and (13) one would need to compute a multiple integral of a relatively complex integrand function. In order to produce a simpler expression for $S^{(-j)}_{\text{eliminate}}(t)$ and $S^{(-j,-k)}_{\text{eliminate}}(t)$, one may consider either simplifying (12) and (13) or finding explicitly the limits in (8) and (9). In general, both approaches are confronted with difficulties. One of them is that the marginal densities, $f'^{(i)}(t)$, $i = 1, \dots, m$, are not in analytic form but are derived from the numerical solution of (5), so direct integration in (12) and (13) is not plausible even for copulas with simpler representation, such as Frank or Plackett copulas. Furthermore, directly finding the limits in (8) and (9) is difficult since, the denominator, $S'^{(j)}(s)$ is obtained as a numerical solution of (5), and it tends to zero as $s \to \infty$. However, as is established in section 4, definitions (8) and (9) lead to a more efficient numerical implementation than the more involved integral expressions (12) and (13). The implementation of the competing risks model under both the *ignoring* and the *eliminating* definition, is illustrated in the next section 4.

4 Numerical results

In this section, we apply the methodology described earlier to UK cause specific mortality data for year 2007, published by the ONS (2008), which includes deaths from cancer, heart disease, respiratory diseases and all other causes grouped together. The classification of causes of death is according to the 10th revision of the International Classification of diseases (ICD-10). For ease of presentation, we consider the two dimensional and the multidimensional competing risk models separately. The numerical implementation of the methodology has been performed using *Mathematica* 7.

4.1 Two causes of death

We consider here the simplest case of only two competing causes of death, one due to cancer (ICD-10 codes C00-D48), and a second one due to all other, non-cancer causes, pooled together. Thus, here $m = 2$ and we denote by T_c and T_o the lifetime random variables for the cancer and non-cancer causes of death and by $S^{(c)}(t)$, $S'^{(c)}(k)$, and $S^{(o)}(t)$, $S'^{(o)}(k)$, the crude and net survival functions for cancer and non-cancer respectively. As noted in section 2, it is possible to estimate crude survival functions based on an appropriate set of cause specific, mortality data. In order to estimate the crude survival functions for cancer and other (non-cancer) causes, we have used a two decrement life table, obtained on the basis of England and Wales cause specific female mortality data for year 2007, published by the ONS (2008); for further details see Table 5, therein. For more details on how the two decrement life table was obtained, see the Appendix. The two decrement life table data are presented in 5 year age intervals and cover the age range from 0 to 95+ years. We have fitted a cubic spline function to the observed crude survival data for ages from 0 to 100, noting that the values published by ONS (2008) have already been smoothed and no further smoothing was required. In order to obtain a "closed" mortality model up to a limiting age of 120, we have extrapolated the fitted cubic spline functions $S^{(c)}(t)$ and $S^{(o)}(t)$, for the cancer and the other (non-cancer) causes, over the 100-120 age range, under the condition that $S^{(c)}(120) = S^{(o)}(120) = 10^{-10}$. For further details regarding the method and formulas used to obtain the observed and extrapolated values of the crude survival functions, we refer to the Appendix. For a summary on different methods which can be used

to extrapolate and close a life table, we refer to Kaishev et al. (2007) and Buettner (2004).

The fitted cubic spline survival functions $S^{(c)}(t)$ and $S^{(o)}(t)$, $0 \leq t \leq 120$ and their densities are given in Fig. 1.

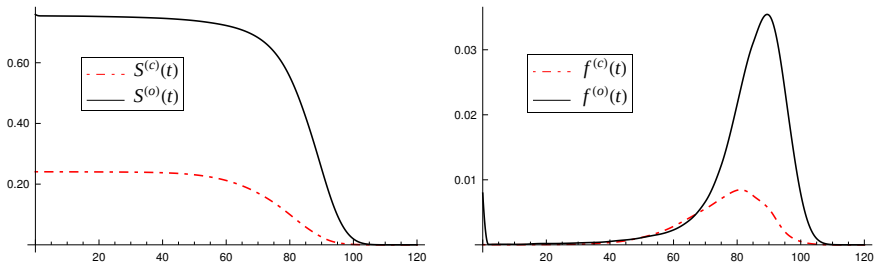

Figure 1: Interpolated crude survival functions (left panel) and their densities (right panel) for 'cancer' and 'other' causes of death.

Having estimated the crude survival functions $S^{(c)}(t)$ and $S^{(o)}(t)$, $0 \leq t \leq 120$, we obtain the net survival functions $S'^{(c)}(t)$ and $S'^{(o)}(t)$, $0 \leq t \leq 120$, by solving the system (5), using three different type of copulas, namely Gaussian, Frank and Plackett copulas. The solutions $S'^{(c)}(t), S'^{(o)}(t)$, $0 \leq t \leq 120$, obtained from (5) have been checked applying equation (4) for the case $m = 2$. As can be seen from Fig. 1, the crude survival functions, $S^{(c)}(t)$ and $S^{(o)}(t)$ are both close to zero in the age range $100 \leq t \leq 120$, therefore numerical solutions of (5), $S'^{(o)}(t)$ and $S'^{(c)}(t)$ may not possess the *Mathematica* precision for the built-in function NDSolve in that range. Another important point is that both, $S'^{(o)}(t)$ and $S'^{(c)}(t)$, are influenced by the extrapolated sections of the crude survival functions not only for $100 \leq t \leq 120$ but within the entire age range $0 \leq t \leq 120$. This in turn means that the exact figures presented in the tables later in this section, depend on the extrapolation that has been carried out. However, the general results and conclusions with respect to survival under the dependent competing risks model are still valid. Furthermore, the life expectancy at birth for females of 81.66 years, reported by the ONS (see the 2005-07 Interim Life Table for England and Wales) coincides with the corresponding figure obtained by integrating the (extrapolated) overall survival function given by (4).

The net survival functions, obtained as a solution of (5), using the Gaussian copula, $C^{\mathrm{Ga}}(u_1, u_2)$, with values of ρ corresponding to five different values of Kendall's τ are plotted in Fig. 2 (so that $\tau = 0.91$ corresponds to

$\rho = 0.99$, $\tau = 0.35$ corresponds to $\rho = 0.52$ and so on). Let us recall that the (non-linear) dependence between the causes of death is measured by the Kendall's τ which in the case of a Gaussian copula is expressed through the linear correlation parameter ρ as $\tau(T_c, T_o) = \frac{2}{\pi}\arcsin(\rho(T_c, T_o))$. The linear correlation ρ is considered as a free parameter, by means of which different degrees of association, between the cancer and non-cancer modes of death, are preassigned. Thus, the system (5) has been solved for values of ρ equal to $-0.99, -0.52, 0.00, 0.52, 0.99$ and the obtained net survival functions $S'^{(o)}(t)$ and $S'^{(c)}(t)$, $0 \leq t \leq 120$, are given in the left and right panel in Fig. 2. The corresponding densities, $f'^{(o)}(t)$ and $f'^{(c)}(t)$ are plotted in Fig. 3. Plots for $S'^{(o)}(t)$ and $S'^{(c)}(t)$, assuming Frank and Plackett copulas are very similar and therefore have been omitted.

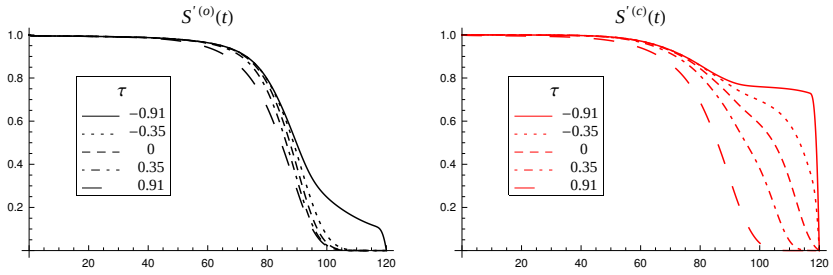

Figure 2: The survival functions $S'^{(o)}(t) \equiv S^{(-c)}_{\text{ignore}}(t)$, $0 \leq t \leq 120$ (left panel) and $S'^{(c)}(t) \equiv S^{(-o)}_{\text{ignore}}(t)$ (right panel), assuming a Gaussian copula.

In the remainder of this section, we will compare and analyze the numerical results of survival under the two alternative definitions of removal of a cause of death, the ignoring and the eliminating definitions given by (6) and (8) in section 3.

In the bi-variate case, under the ignoring definition, for fixed ρ, the net survival function, $S'^{(o)}(t)$, $0 \leq t \leq 120$, coincides with the overall survival function, $S^{(-c)}_{\text{ignore}}(t)$, $0 \leq t \leq 120$, when cancer has been removed, i.e., $S'^{(o)}(t) \equiv S^{(-c)}_{\text{ignore}}(t)$ and $f'^{(o)}(t) \equiv f^{(-c)}_{\text{ignore}}(t)$, where $f^{(-c)}_{\text{ignore}}(t) = -\frac{dS^{(-c)}_{\text{ignore}}(t)}{dt}$. Obviously, if cancer is ignored in the bi-variate decrement model, the overall survival will entirely be determined by the only remaining cause of death, that of non-cancer, and vice-versa. Therefore, in order to study the overall survival, when cancer is ignored, we may directly study the non-cancer net survival function $S'^{(o)}(t)$ and its corresponding density, $f'^{(o)}(t)$, given in the

left panel of Fig. 3.

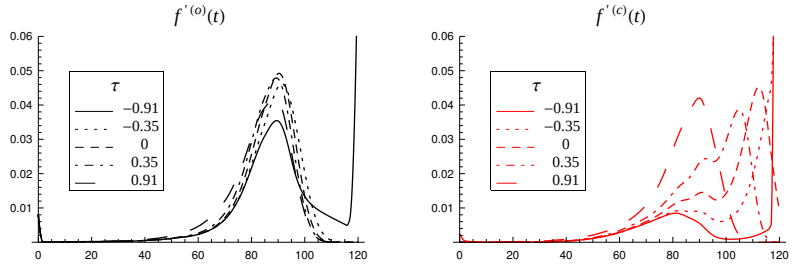

Figure 3: The density functions, $f^{'(o)}(t) \equiv f^{(-c)}_{\text{ignore}}(t)$, $0 \leq t \leq 120$, (left panel) and $f^{'(c)}(t) \equiv f^{(-o)}_{\text{ignore}}(t)$, $0 \leq t \leq 120$, (right panel), assuming a Gaussian copula.

As can be seen from the left panel of Fig. 2, ignoring cancer affects survival most significantly when Kendall's $\tau = -0.91$ ($\rho_S = -0.99$), which corresponds to the case of extreme negative dependence. This effect of rectangularization of the overall survival function is seen even more clearly on the right panel of Fig. 2, where the 'other' cause of death has been removed. In addition, we note that in the case of negative dependence or even independence between T_c and T_o, the trend of the overall survival curves suggests that the limiting age lies somewhere beyond 120 and it would not be natural to expect the old age survivors to die almost simultaneously at 120.

Survival under the eliminating definition of removal of a cause is illustrated for the three different choices of copula, Gaussian, Frank and Plackett copulas, in Fig. 4-6, respectively. It is worth noting that, contrary to the ignoring definition, the net survival function, $S^{'(o)}(t)$, $0 \leq t \leq 120$, does not coincide with the overall survival function, $S^{(-c)}_{\text{eliminate}}(t)$, $0 \leq t \leq 120$, when cancer has been eliminated, i.e., $S^{'(o)}(t) \neq S^{(-c)}_{\text{eliminate}}(t)$. The overall survival function, $S^{(-c)}_{\text{eliminate}}(t)$, has been computed based on definition (8) (see section 3) for the case $m = 2$, with $s \to \infty$ replaced by $s \to 120$, in which case $S^{'(j)}(s) \to 0$ has been replaced by $S^{'(j)}(s) \to 10^{-10}$, i.e. (8) simplifies to

$$S^{(-j)}_{\text{eliminate}}(t) = C\left(S^{'(1)}(t), \ldots, S^{'(j-1)}(t), 10^{-10}, S^{'(j+1)}(t), \ldots, S^{'(m)}(t)\right) \times 10^{10}. \tag{14}$$

It has to be noted that expression (12) can be used as an alternative to (14), however, its evaluation is much more time consuming and in the case

of the Gaussian and t-copulas, for which considerable probability mass is located at the origin, this leads to unstable computations. In contrast, the evaluation of (14) is stable and requires only a few seconds in the case of $m = 2$, for any of the three copulas selected.

We are interested in assessing the gain in life expectancy due to a cause removal and hence, in what follows we will compare the corresponding values with the life expectancy (81.66 years) in the case when none of the causes has been removed. The life expectancy at birth, $\overset{\circ}{e}{}_0^{(-j)}$, and at age 65, $\overset{\circ}{e}{}_{65}^{(-j)}$, when the j-th cause is removed can be expressed as

$$\overset{\circ}{e}{}_0^{(-j)} = \int_0^{120} S^{(-j)}(t)\, dt \tag{15}$$

and as

$$\overset{\circ}{e}{}_{65}^{(-j)} = \int_{65}^{120} \frac{S^{(-j)}(t)}{S^{(-j)}(65)}\, dt, \tag{16}$$

where the survival function $S^{(-j)}(t)$ is substituted from expressions (10) and (14) for the ignore and eliminate definitions of removal, respectively.

As can be seen from Fig. 4-6, for negative values of τ, the overall survival function, $S^{(-c)}_{\text{eliminate}}(t)$, when cancer is eliminated, suggests very poor survival from the remaining cause (all other causes pooled together) for all three copula choices. This is confirmed by the negative values of the gain in the life expectancies at birth, $\overset{\circ}{e}{}_0^{(-c)}$, and at age 65, $\overset{\circ}{e}{}_{65}^{(-c)}$, calculated according to (15) and (16) respectively with $j = c$, presented in Tables 1-3 for $\tau = -0.91$ and $\tau = -0.35$. The latter phenomenon is observed because, under strong negative correlation i.e., $\tau = -0.91$, individuals survive to 120 from cancer (when it is eliminated) and hence, they will tend to die from the remaining competing cause already at birth, due to the assumed strong negative correlation of the corresponding lifetimes. Clearly, under the eliminating definition, such negative correlation makes little sense, since it suggests that improvement of mortality with respect to one cause would lead to increasing the mortality from the remaining cause. Such a setting is of little relevance when the competing risks are critical illnesses, since what is important in the context of medical, demographic and actuarial applications is how life expectancy and other survival characteristics are affected if mortality improves as a result of successful elimination of any of the main causes of death. Therefore, under the eliminating definition of removal of a

cause of death, it is sufficient to study only the range of positive correlation between the competing lifetime random variables. Hence, copula functions which are not necessarily comprehensive can be used. The latter is particularly important in the truly multivariate case, $m > 2$, where the number of comprehensive copulas is limited.

As can be seen from the left panel of Fig. 4, assuming almost perfect positive correlation and eliminating cancer, i.e. achieving perfect survival with respect to it, naturally leads to perfect survival with respect to the only remaining competing risk (all other causes pooled together), and hence leads to perfect overall survival, given cancer is eliminated. This is clearly illustrated by the curve, $S^{(-c)}_{\text{eliminate}}(t)$ for $\tau = 0.91$, which is almost rectangular.

Comparing Fig. 2 and Fig. 4, and also the columns "Ignore" and "Eliminate" of Table 1, which summarizes the values of $\overset{\circ}{e}{}^{(-c)}_0$ and $\overset{\circ}{e}{}^{(-c)}_{65}$, under both the ignoring and eliminating definitions, it can be seen that survival under the two alternative definitions is quite different. Thus, under the ignoring definition, improvement in mortality is achieved for all values of $\tau \in (-1, 1)$, whereas under the eliminating definition, mortality improvement is achieved only for non-negative values of $\tau \in [0, 1)$. On the other hand, looking at Fig. 2 and Fig. 4, it can be seen that the overall survival function under the ignoring definition varies within a relatively small range and is bounded from above by the curve for $\tau = -0.91$ which is nearly the best possible mortality improvement, attained in the limit, as $\tau \to -1$, in which case the Gaussian copula converges to the lower Fréchet-Hoeffding bound. In contrast to the ignoring case, under the elimination definition survival is very sensitive with respect to the value of τ and can vary within the entire range, from zero life span to 120 years life span, as seen from Fig. 4and the values for $\overset{\circ}{e}{}^{(-c)}_0$ and $\overset{\circ}{e}{}^{(-c)}_{65}$, presented in Table 1. Also, contrary to the ignoring definition, under elimination, survival depends significantly on the choice of the copula modelling the dependence between the lifetimes, as can be seen comparing the survival functions in Figures 4, 5 and 6, and the numbers for $\overset{\circ}{e}{}^{(-c)}_0$ and $\overset{\circ}{e}{}^{(-c)}_{65}$, presented in Tables 1-3. Comparing the curves in Fig. 2 and Fig. 4, it can be verified that the two definitions are equivalent in the independent case $\tau = 0$, as noted in section 3.

What can also be observed, comparing the survival curves in Fig. 4, 5 and 6 is that the curves corresponding to the Frank and Plackett copulas

are relatively much closer to each other then to the curves for the Gaussian copula case. This is consistent also with the numerical results for $\overset{\circ}{e}{}_0^{(-c)}$ and $\overset{\circ}{e}{}_{65}^{(-c)}$, presented in the columns "Eliminate" of Tables 2 and 3, which are close to each other for most of the values of τ. It can also be seen from Fig. 5 and 6 that for both copulas, improvement of survival is somewhat more limited and rectangularization for $\tau = 0.91$ is not achieved, in contrast to the case of Gaussian copula, given in Fig. 4. Comparing the numerical values for $\overset{\circ}{e}{}_0^{(-c)}$ and $\overset{\circ}{e}{}_{65}^{(-c)}$, summarized in Table 1 with those given in Tables 2 and 3, one can conclude that, under the eliminating definition, the results are more sensitive both with respect to the value of τ and the choice of copula, than under the ignoring definition.

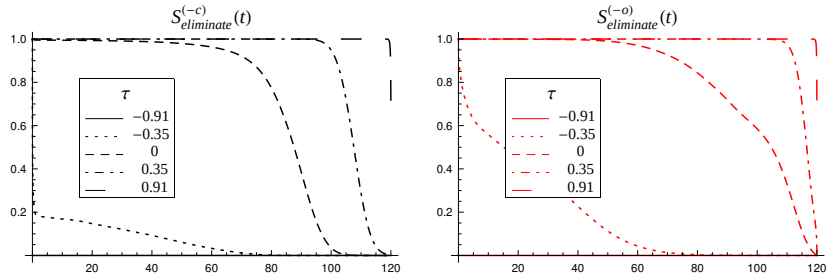

Figure 4: Overall survival functions, $S^{(-c)}_{\text{eliminate}}(t)$, given cancer eliminated (left panel) and $S^{(-o)}_{\text{eliminate}}(t)$, given all other causes eliminated (right panel), assuming Gaussian copula dependence.

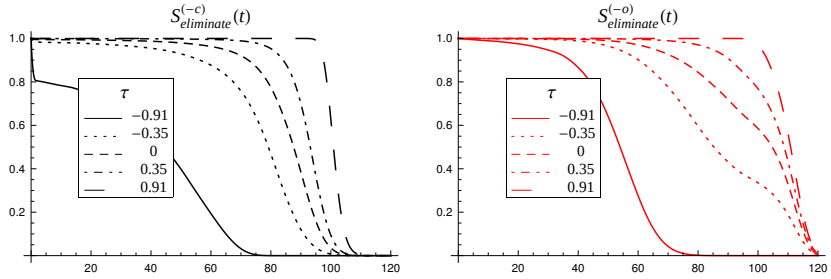

Figure 5: Overall survival functions, $S^{(-c)}_{\text{eliminate}}(t)$, given cancer eliminated (left panel) and $S^{(-o)}_{\text{eliminate}}(t)$, given all other causes eliminated (right panel), assuming Frank copula dependence.

Although our focus so far has been at the changes in the overall survival function $S(t)$, $0 \leq t \leq 120$, under the two alternative definitions of ignoring

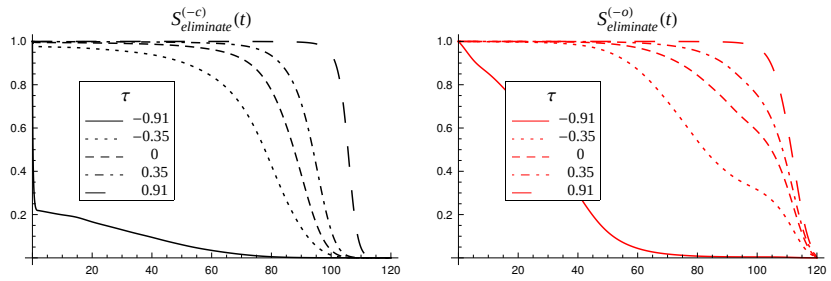

Figure 6: Overall survival functions, $S^{(-c)}_{\text{eliminate}}(t)$, given cancer eliminated (left panel) and $S^{(-o)}_{\text{eliminate}}(t)$, given all other causes eliminated (right panel), assuming Plackett copula dependence.

Table 1: Gaussian copula results.

τ	$N(\rho)$	$\overset{\circ}{e}{}^{(-c)}_0$	[gain]	$\overset{\circ}{e}{}^{(-c)}_{65}$	[gain]
		Ignore	Eliminate	Ignore	Eliminate
-0.91	$\rho = -0.99$	90.56 [8.90]	0 [-81.66]	27.94 [7.93]	0 [-20.01]
-0.35	$\rho = -0.52$	86.39 [4.74]	7.37 [-74.29]	23.55 [3.54]	6.63 [-13.38]
0	$\rho = 0$	85.21 [3.55]	85.21 [3.55]	22.33 [2.31]	22.33 [2.31]
0.35	$\rho = 0.52$	84.07 [2.41]	107.48 [25.82]	21.27 [1.25]	42.48 [22.47]
0.91	$\rho = 0.99$	82.02 [0.36]	119.96 [38.30]	20.06 [0.05]	54.96 [34.95]

Table 2: Frank copula results.

τ	$F(\theta)$	$\overset{\circ}{e}_0^{(-c)}$ [gain]		$\overset{\circ}{e}_{65}^{(-c)}$ [gain]	
		Ignore	Eliminate	Ignore	Eliminate
-0.91	$\theta = -44.88$	89.74 [8.08]	37.87 [-43.79]	27.08 [7.07]	3.97 [-16.04]
-0.35	$\theta = -3.46$	86.28 [4.62]	75.32 [-6.34]	23.43 [3.42]	16.34 [-3.67]
0.35	$\theta = 3.46$	84.15 [2.49]	92.24 [10.58]	21.28 [1.27]	27.54 [7.53]
0.91	$\theta = 44.88$	82.20 [0.54]	101.05 [19.39]	19.94 [-0.07]	36.05 [16.04]

Table 3: Plackett copula results.

τ	$P(\theta)$	$\overset{\circ}{e}_0^{(-c)}$ [gain]		$\overset{\circ}{e}_{65}^{(-c)}$ [gain]	
		Ignore	Eliminate	Ignore	Eliminate
-0.91	$\theta = \frac{1}{735.8}$	88.93 [7.27]	8.36 [-73.30]	26.22 [6.21]	9.77 [-10.24]
-0.35	$\theta = \frac{1}{5.022}$	86.30 [4.64]	73.53 [-8.13]	23.46 [3.45]	16.42 [-3.59]
0.35	$\theta = 5.022$	84.17 [2.51]	92.42 [10.76]	21.32 [1.30]	27.94 [7.93]
0.91	$\theta = 735.8$	82.12 [0.46]	105.28 [23.62]	20.10 [0.08]	40.29 [20.28]

and eliminating a cause of death, the joint survival function of T_c and T_o, $S(t_1, t_2) = \Pr(T_c > t_1, T_o > t_2)$, $0 \leq t_j \leq 120$, $j = 1, 2$, is also of interest. However, since either one of the causes leads to death, and the other lifetime remains latent, probabilistic inference related to the joint distribution of T_c and T_o is somewhat artificial. Nevertheless, it is instructive and in Fig. 7-9 we have plotted the joint density of T_c and T_o, in case of the bi-variate Gaussian, Frank and Plackett copulas for Kendall's $\tau = 0.35$. For any bi-variate copula, the joint density of T_c and T_o can be calculated from (2) as

$$\frac{\partial^2}{\partial t_1 \partial t_2} S(t_1, t_2) = c\left(S'^{(c)}(t_1), S'^{(o)}(t_2)\right) \times f'^{(c)}(t_1) \times f'^{(o)}(t_2) \qquad (17)$$

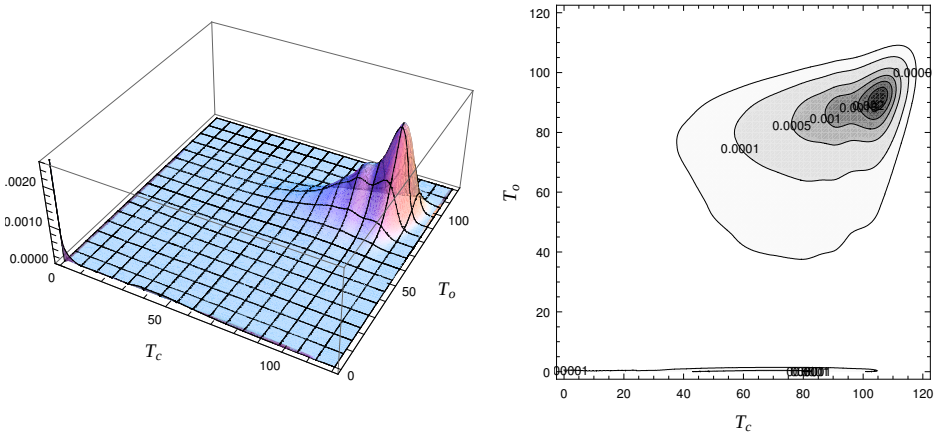

Figure 7: A 3D plot and a contour plot of the joint density of T_c and T_o, expressed through the Gaussian copula, for Kendall's $\tau = 0.35$.

As seen from Fig. 7-9, under this assumption of positive dependence, jointly increasing values of the lifetimes T_c and T_o are likely to occur. This is valid, regardless of which copula has been assumed to model the dependence. There are, of course, some copula specific differences in the joint density functions, as is natural to expect in view of (17). As can be seen from Fig. 8 and 9, the plots of the joint density of T_c and T_o are similar for the Frank and Placket cases, and are somewhat different to the density plots in case of the Gaussian copula given in Fig. 7. The Gaussian copula seems to preserve the peak at the early infant mortality which is inherent from the

(empirical) crude survival functions. Another, obvious characteristic of the joint density function for all three copulas is that it has two modes, more strongly expressed in the case of Frank and Placket copulas. Clearly, the choice of copula would be contingent on the availability and access to data, providing information about the interaction between diseases, and medical experts opinion (cf. Section 4.1 in Kaishev et al. (2007)).

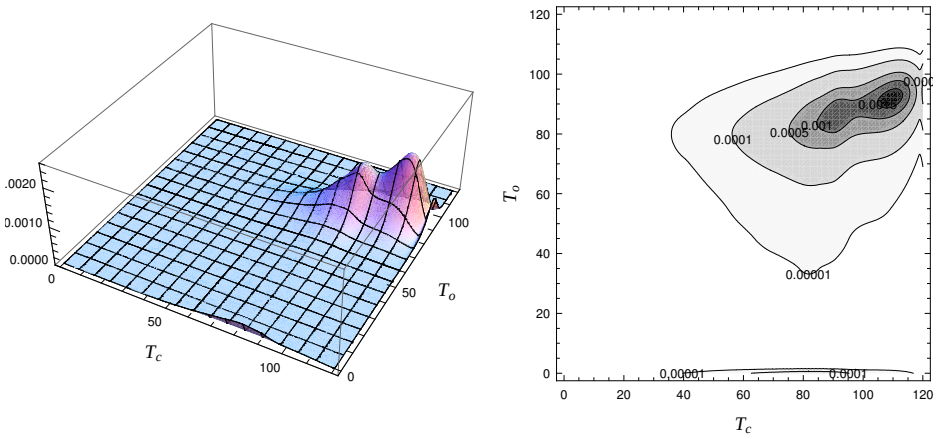

Figure 8: A 3D plot and a contour plot of the joint density of T_c and T_o, expressed through the Frank copula, for Kendall's $\tau = 0.35$.

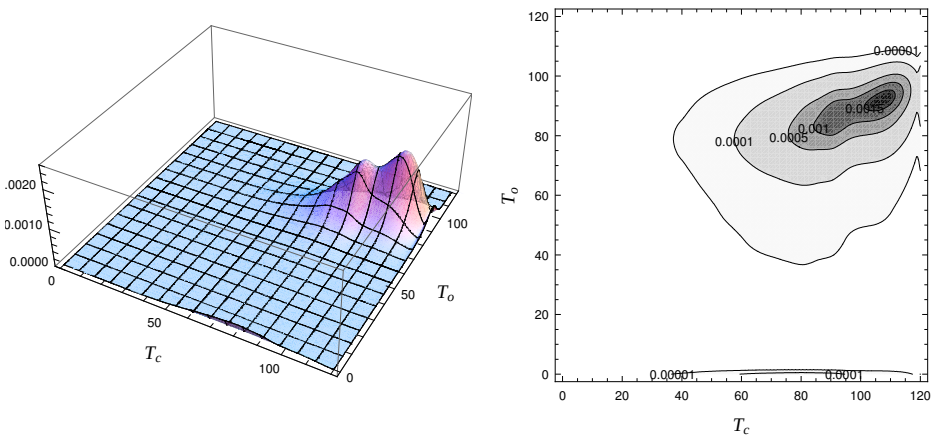

Figure 9: A 3D plot and a contour plot of the joint density of T_c and T_o, expressed through the Plackett copula, for Kendall's $\tau = 0.35$.

4.2 Multiple causes of death $(m = 4)$

We now illustrate the extension of the proposed methodology to the multivariate case by considering four competing causes of death, cancer (c), (ICD-10 codes C00 − D48), heart diseases (h), (ICD-10 codes I00 − I99), respiratory diseases (r), (ICD-10 codes J00 − J99), and other causes (o), grouped together. As in the bi-variate case, we have constructed a four decrement life table using England and Wales cause specific female mortality data for year 2007, published by the ONS (2008). For more details on how the four decrement life table was obtained see the Appendix. The interpolated crude survival functions $S^{(c)}(t)$, $S^{(h)}(t)$, $S^{(r)}(t)$, $S^{(o)}(t)$, $0 \leq t \leq 120$ and their derivatives are given in Fig. 10.

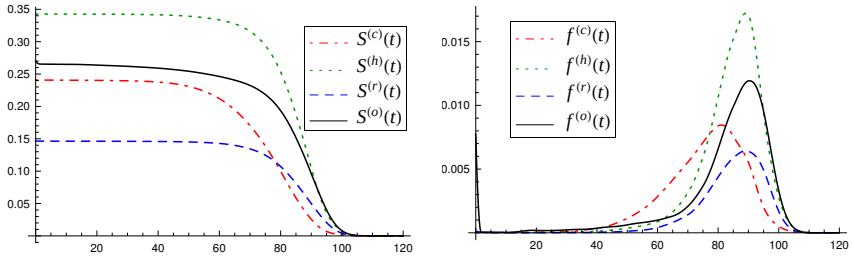

Figure 10: The crude survival functions (left panel) and their densities (right panel).

For illustrative purposes we have used the multivariate Frank copula to model purely positive dependence between the lifetimes T_c, T_h, T_r and T_o, which, as noted in the bi-variate case, is the meaningful range of dependence under the elimination definition of cause removal. The four net survival functions obtained as a solution to system (5), and their densities for the multivariate Frank copula with parameter $\theta = 3.46$, are presented in Fig. 11.

In the left panels of Fig. 12 and Fig. 14 we give the overall survival functions with each one of the three possible diseases individually removed, $j \in \{h, c, r\}$, under the ignoring and the eliminating definitions of removal, respectively, and compare them to the overall survival function with no disease removed, $S(t)$. As can be seen, the improvement in survival is more significant under the eliminating definition than under the ignoring one. This is confirmed also by comparing the corresponding gains in the life expectancy, summarized in the first three rows of Table 4. It is logical to

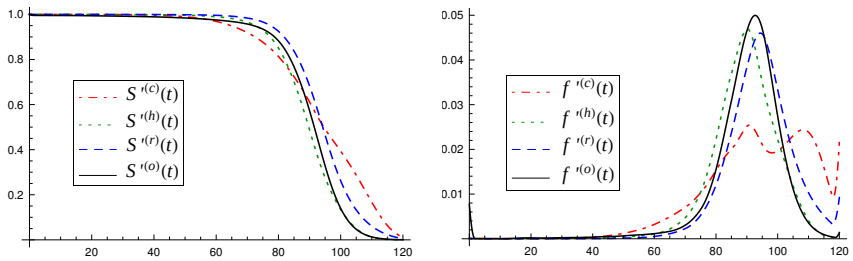

Figure 11: The net survival functions (left panel) and their densities (right panel).

obtain more significant gains in life expectancy under the *elimination* definition, since in that case the lifetime for the eliminated cause is pushed to the limiting age, which also causes other (positively) correlated lifetimes to increase. As a result, gain in life expectancy, due to the collective improvement of lifetimes, is much more significant. In contrast, under the *ignoring* definition the lifetime is simply omitted from the set of competing lifetimes, which leads to a much less expressed association with the remaining lifetimes and therefore, less gain in life expectancy.

Another interesting conclusion, drawn from Table 4, is that maximum gain in $\overset{\circ}{e}{}^{(-j)}_{65}$ is achieved when heart disease is removed, i.e. $j \equiv \{h\}$, and this is true under both the ignoring and eliminating definitions. Under the ignoring definition, the maximum gain in $\overset{\circ}{e}{}^{(-j)}_{0}$ is achieved if cancer is ignored, i.e., $j = c$, contrary to the eliminating definition where the maximum gain is attained if heart disease is eliminated. However, ignoring cancer or heart disease produces very similar gains in the life expectancy at birth, $\overset{\circ}{e}{}^{(-j)}_{0}$, as can be seen from Table 4.

As it is also illustrated in Fig. 12 and Fig. 14, under both definitions the most significant improvement in survival for the age range $40 \leq t \leq 85$ is achieved if cancer is removed, whereas for $85 \leq t \leq 120$ the best improvement in survival is due to removal of heart disease. As expected, improvement in survivorship due to the removal of respiratory disease is not as significant.

In the left panels of Fig. 13 and Fig. 15 we give the overall survival functions with all possible pairs of diseases removed, i.e., $j \equiv \{c, h\}$, $j \equiv \{c, r\}$, $j \equiv \{h, r\}$, under the ignoring and the eliminating definitions of removal, respectively and again contrast them to the overall survival functions with

no disease removed, $S(t)$. As in the case of removing only one disease at a time, if a pair of diseases is removed, improvement in survival is more significant under the eliminating definition, compared to the ignoring one. This is confirmed also comparing the corresponding gains in the actuarial functions, summarized in the second three rows of Table 4. Under the ignoring definition, the maximum gain in $\overset{\circ}{e}{}_0^{(-j)}$ and $\overset{\circ}{e}{}_{65}^{(-j)}$ is achieved if the pair, $j = ch$ is removed, whereas under the eliminating definition the maximum in $\overset{\circ}{e}{}_0^{(-j)}$ and $\overset{\circ}{e}{}_{65}^{(-j)}$ is attained if $j = hr$. However, as can be seen from Table 4, under the eliminating definition, the gains in $\overset{\circ}{e}{}_0^{(-j)}$ and $\overset{\circ}{e}{}_{65}^{(-j)}$ are very similar in the case of $j \equiv \{c, h\}$ and $j \equiv \{h, r\}$, so one may argue that under both definitions, the removal of cancer and heart ($j \equiv \{c, h\}$) brings about (most) significant gains in the life expectancy figures summarized in Table 4.

It is worth noting also another way in which the two definitions are different. Comparing the gains obtained if one cause is removed to the gains resulting from the removal of two causes, one can see that the gains nearly double under the ignoring definition while they are nearly the same under the eliminating definition. This is natural to expect since one and the same level of positive correlation between the lifetime random variables, T_c, T_h, T_r and T_o, has a different interpretation and numerical effect on the functions $\overset{\circ}{e}{}_0^{(-j)}$ and $\overset{\circ}{e}{}_{65}^{(-j)}$, under the two alternative definitions. And finally we note that regardless of the definition, the maximum gain is achieved when all three diseases are removed and this is illustrated by the last row of Table 4.

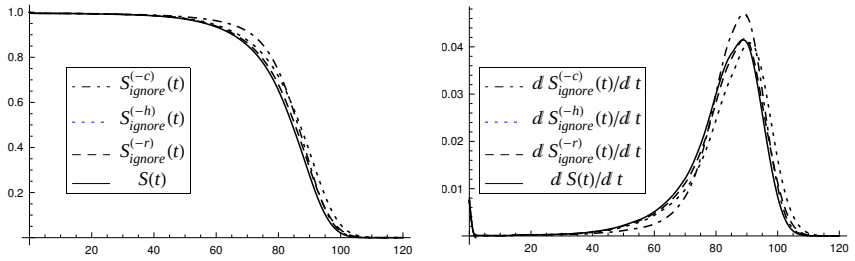

Figure 12: The overall survival functions with no disease ignored and with only one disease ignored (left panel) and their densities (right panel).

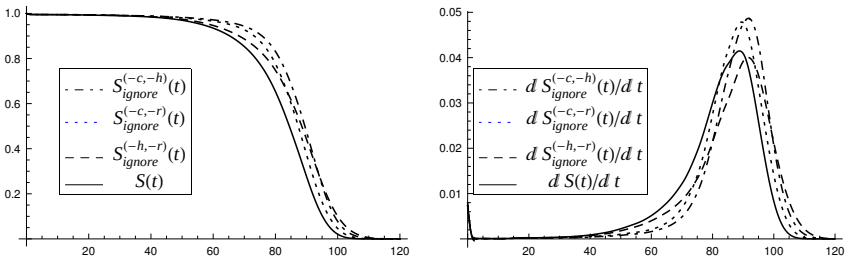

Figure 13: The overall survival functions with no disease ignored and with only two diseases ignored (left panel) and their densities (right panel).

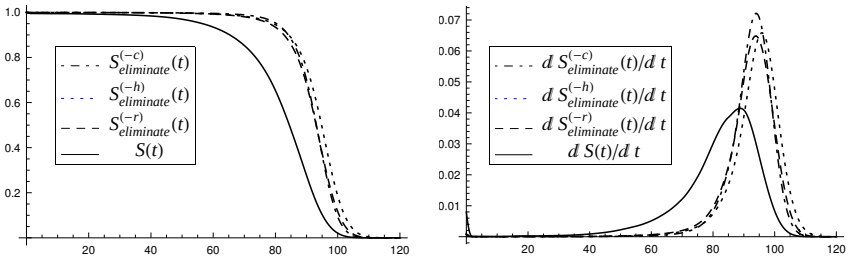

Figure 14: The overall survival functions with no disease eliminated and with only one disease eliminated (left panel) and their densities (right panel).

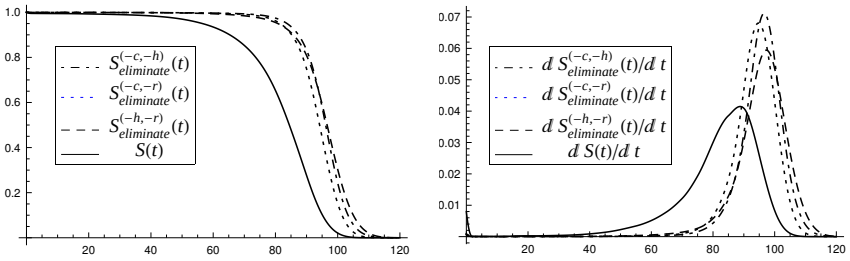

Figure 15: The overall survival functions with no disease eliminated and with only two diseases eliminated (left panel) and their densities (right panel).

Table 4: Multivariate Frank copula results.

τ	$\overset{\circ}{e}_0^{(-c)}$ [gain]		$\overset{\circ}{e}_{65}^{(-c)}$ [gain]	
	Ignore	Eliminate	Ignore	Eliminate
$j \equiv c$	84.15 [2.49]	92.25 [10.59]	21.28 [1.27]	27.55 [7.53]
$j \equiv h$	84.06 [2.40]	93.63 [11.97]	22.19 [2.17]	29.10 [9.08]
$j \equiv r$	82.56 [0.89]	91.95 [10.29]	20.80 [0.78]	27.44 [7.42]
$j \equiv c, h$	87.48 [5.82]	95.47 [13.81]	24.23 [4.22]	30.71 [10.69]
$j \equiv c, r$	85.35 [3.69]	93.45 [11.79]	22.31 [2.29]	28.73 [8.71]
$j \equiv h, r$	85.64 [3.98]	95.59 [13.93]	23.68 [3.66]	31.05 [11.03]
$j \equiv c, h, r$	89.91 [8.25]	98.05 [16.39]	26.49 [6.47]	33.26 [13.24]

5 Concluding remarks

In this paper, we have demonstrated how copula functions can be applied in modelling dependence between lifetime random variables in the context of competing risks. We have implemented the multivariate copula dependent competing risks model to study the impact of removing one or more causes of death on England & Wales 2007 cause specific mortality. In particular, we have focused at comparing and contrasting two alternative definitions of cause removal, namely *ignoring* and *eliminating* a cause, and their effect on the overall survival and the life expectancy at birth and age 65. For this purpose, we have provided expressions for the overall survival functions in terms of the specified copula (density) and the net (marginal) survival functions.

We have shown that there are substantial differences in the overall survival functions, given one or more risks are removed, under the two definitions which is also reflected in the values of the life expectancy. An important conclusion derived from this work is that the elimination definition is more appropriate for biostatistical, medical, demographic or actuarial applications, since it suffices to consider only positive dependence among the competing lifetimes and the model results are more intuitive and easily interpretable.

The methodology and results may be applied in: managing longevity risk; setting target levels for mortality rates that will assist with scenario testing and sensitivity analyses in the presence of dependence between causes of death; population forecasting and planning; life insurance business where the financial impact of mortality improvements on life insurance and annuities products may be investigated.

The question of how to estimate the (pairwise) correlations between causes of deaths via their associated lifetimes, requires further research in close collaboration with the medical profession. In this regard, promising directions of research may be to look at estimation, based on the so called Expectation-maximization algorithms, and also quantitative methods for modelling expert's opinion.

Acknowledgements

The authors would like to acknowledge the financial support received from the Institute of Actuaries and to thank the Actuarial Profession for the opportunity to present earlier versions of this work at the seminars "Mortality and Longevity - Making Financial Sense of the Highly Uncertain" held in London and Manchester in 2009. The authors would also like to thank the participants of these seminars for thought-provoking questions and useful comments.

References

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. (1997). Actuarial Mathematics. Itasca, III: Society of Actuaries.

Bryant, J. and Dignam, J.J. (2004). Semiparametric Models for Cumulative Incidence Functions. *Biometrics*, 60, 182-190.

Buettner, T. (2004). Approaches and experiences in projecting mortality patterns for oldest-old. *North American Actuarial Journal*, 6 (3), 14–29.

Carriere, J. (1994). Dependent Decrement Theory. *Transactions, Society of Actuaries*, v. XLVI, 45-65.

Carriere, J. (1995). Removing Cancer when it is Correlated with other Causes of Death. *Biometrical Journal*, 3, 339-350.

Chen, Y. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *J. R. Statist. Soc. B*, 72, Part 2, 235-251.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula methods in finance*. John Wiley and Sons Ltd.

Crowder, M. (2001). Classical Competing Risks, Chapman and Hall/CRC, London.

Craiu, R. V. and Reiser, B. (2006). Inference for the dependent competing risks model with masked causes of failure, *Lifetime Data Anal.*, 12, 21-33.

Elandt-Johnson, R.C. (1976). Conditional Failure Time Distributions under Competing Risk Theory with Dependent Failure Times and Proportional Hazard Rates. *Scandinavian Actuarial Journal*, 37-51.

Elandt-Johnson, R.C. and Johnson, N.L. (1980). Survival models and data analysis. Wiley, New York.

Escarela, G. and Carriere, J. (2003). Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research*, 12 (4), 333-349.

Fukumoto, K. (2005). Survival Analysis of Systematically Dependent Competing Risks:An Application to the U.S. Congressional Careers. paper presented at the poster session at the 22nd Annual Summer Meeting of the Society for Political Methodology, Tallahassee, FL, USA, July 21-23, 2005.

Gaynor, J.J., Feuer, E.J., Tan, C.C., Wu, D.H., Little, C.R., Straus, D.J., Clarkson, B.D., Brennan, M.F. (1993). On the use of cause-specific failure time and conditional failure probabilities: examples from clinical oncology data. *J. Am. Stat. Assoc.*, 88, 400-409.

Gooley, T.A., Leisenring, W., Crowley, J., Storer, B. (1999). Estimation of failure probabilities in the presence of competing risks: new representation of old estimators. *Stat Med*, 18, 695-706.

Honoré, B. E. and Lleras-Muney, A. (2006). Bounds in competing risks models and the war on cancer. *Econometrica*, 74, 4 , 1675-1698.

Hooker, P.F. and Longley-Cook, L.H. (1957). Life and other Contingencies. Vol. II. Cambridge, Great Britain: University Press.

Joe, H. (1997). Multivariate Models and Dependent Concepts. Chapman & Hall, London.

Jeong, J. and Fine, J.P. (2007). Parametric regression on cumulative incidence function. *Biostatistics*, 8, 2, 184-196.

Kaishev, V.K., Dimitrova, D.S., Haberman, S. (2007). Modelling the joint distribution of competing risks survival times using copula functions. *Insurance: Mathematics and Economics*, 41, 339-361.

Kalbfleisch, J.D. and Prentice, R.L. (2002). The statistical analysis of failure time data. 2nd ed. John Wiley & Sons.

Kaput, J., Swartz, D., Paisley, E., Mangian H., Daniel, W. L., Visek, W.J. (1994). Diet-Disease Interactions at the Molecular Level: An Experimental Paradigm. *The Journal of Nutrition*, 124, 1296S-1305S.

Lawless, J.F. (2003). Statistical models and methods for lifetime data. 2nd ed. John Wiley & Sons.

Lindqvist, H. B. (2007). Competing risks In: *Encyclopedia of Statistics in Quality and Reliability*, Ruggeri, F., Kenett, R. and Faltin, F. W. (eds). John Wiley & Sons Ltd, Chichester, UK, pp 335-341.

Lin, D.Y. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Stat Med*, 16, 901-910.

Lindqvist, H. B. and Skogsrud, G. (2009). Modeling of Dependent Competing Risks by First Passage Times of Wiener Processes. *IIE Transactions*, vol 41, 72-80.

Lo, S.M.S. and Wilke, R.A. (2010). A copula model for dependent competing risks. *Appl. Statist.*, 59, Part 2, pp 359-376.

Lobo, I. (2008). Epistasis: Gene interaction and the phenotypic expression of complex diseases like Alzheimer's. *Nature Education* 1, 1.

Moeschberger, M. L. and Klein, J. P. (1995). Statistical methods for dependent competing risks. *Lifetime Data Anal.*, 1, (2), 195-204.

Nelsen, R. (2006). An Introduction to Copulas. 2nd edition, Springer, New York.

National Center for Health Statistics, (1999). U.S. decennial life tables for 1989−−91, vol. 1, No 4. United States life tables eliminating certain causes of death. Hyattsville, Maryland.

Pintilie, M. (2006). Competing Risks: A practical Perspective, John Wiley and Sons, Chichester.

Office for National Statistics, (2008). Mortality Statistics: Deaths Registered in 2007, Review of the National Statistician on deaths in England and Wales 2007, National Statistics Publication, Series DR 07.

Salinas-Torres, V. H., de Brança Pereira, C. A. and Tiwardi, R.C. (2002). Bayesian Nonparametric Estimation in a Series or a Competing Risk Model. *Journal of Nonparametric Statistics*, 14, 449-458.

Solari, A., Salmaso, L., El Barmi, H. and Pesarin, F. (2008). Conditional tests in a competing risks model, *Lifetime Data Anal*, 14, 154-166.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Natn. Acad. Sci. USA*, 72, 20-22.

Valdez, E. (2001). Bivariate analysis of survivorship and persistency. *Insurance: Mathematics and Economics*, 29, 357-373.

Weir, M.R. (2005). Cardiovascular Disease in Patients With Chronic Kidney Disease: The CVD/CKD Interaction. *Medscape Cardiology*, 9, 2.

Yashin, A., Manton, K., and Stallard, E. (1986). Dependent competing risks: A stochastic process model. *Journal of Mathematical Biology*, 24, 119-140.

Zheng, M., and Klein (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula, *Biometrika*, 82 (1), 127-138.

Appendix

In order to illustrate the competing risks framework and compare the two cause removal definitions presented in the paper, we need a cause-specific mortality table. Unfortunately, such tables are not directly available for the UK population and so, here we describe how we have constructed a two and a four decrement UK female population data set (FP), using "Table 5. Death: underlying cause, sex and age-group, 2007: summary" from ONS (2008) and the England & Wales 2005-07 Interim Life Table as published by ONS. Table 5 from ONS (2008) contains number of deaths by cause of death and total number of deaths, relating to five year age groups, e.g. 5-9, 10-14, 15-19 and so on. The first and the last age spans for which data are given in the table are correspondingly 0-1 and 95+ and the causes of death are coded according to the International Classification of Diseases (ICD), 10th revision. So, from these data we have extracted the proportions in every age group of people dying from cancer (c), (ICD-10 codes C00-D48), from heart diseases (h), (ICD-10 codes I00-I99), from respiratory diseases (r), (ICD-10 codes J00-J99) and all other causes of death, (o), pooled together. Clearly, for the purpose of constructing the two decrement table we have combined the figures for heart and respiratory diseases and added them to the group of 'other' causes of death. The proportions obtained in this way were applied to the number of deaths, d_x, given in the England & Wales 2005-07 Interim Life Table and the resulting age-grouped, multiple decrement tables are given in Table 5 and Table 6.

Based on the crude data presented in Table 5 and Table 6, we easily obtain the observed values at ages $k = 1, 5, 10, \ldots, 95, 100$ of the crude survival functions $S^{(c)}(k)$ and $S^{(o)}(k)$, see Table 7, and $S^{(c)}(k)$, $S^{(h)}(k)$, $S^{(r)}(k)$ and $S^{(o)}(k)$, see Table 8. As mentioned in section 4, cubic spline functions were fitted to these crude survival data and an extrapolation has been performed over the 100-120 age range by setting $S^{(j)}(120) = 10^{-10}$, $j \in \{c, h, r, o\}$. It has to be noted that the spline functions have been fitted to $\log S^{(j)}(k)$ data and than transformed back to the original scale. The latter allows to avoid some unwanted wiggling of the spline curves when fitted directly to $S^{(j)}(k)$ data, as for example the fit becoming negative in the very old ages. In order to obtain the observed values of the crude survival functions presented in Tables 7 and 8, the following quantities were

calculated:
${}_\infty q_0^{(j)}$ - the multiple-decrement probability that a newborn will die from cause of death j, $j \in \{c, h, r, o\}$
${}_\infty d_0^{(j)}$ - the total number of deaths from cause of death j, $j \in \{c, h, r, o\}$, for all ages from 0 to ∞
The following formula were used to obtain the values of ${}_\infty q_0^{(c)} = 0.24$ and ${}_\infty q_0^{(o)} = 0.76$ in the two-decrement case, and ${}_\infty q_0^{(c)} = 0.24$, ${}_\infty q_0^{(h)} = 0.34$, ${}_\infty q_0^{(r)} = 0.15$ and ${}_\infty q_0^{(o)} = 0.27$ in the four-decrement case, based on the values given in Table 5 and Table 6:

$$ {}_\infty q_0^{(j)} = \frac{\sum_x d_x^{(j)}}{l_0} = \frac{{}_\infty d_0^{(j)}}{l_0}, j \in \{c, h, r, o\}. $$

The following formulae were used to calculate the values of ${}_k d_0^{(j)}$, $S_0^{(j)}(k)$ and $S(k)$ given in Table 7 and Table 8, based on the values given in Table 5 and Table 6:

$$ {}_k d_0^{(j)} = \sum_{x<k} d_x^{(j)} $$

$$ S^{(j)}(k) = {}_\infty q_0^{(j)} - \frac{{}_k d_0^{(j)}}{l_0} $$

$$ S(k) = \sum_j S^{(j)}(k) = \frac{l_k}{l_0} $$

$$ S^{(j)}(0) = {}_\infty q_0^{(j)} $$

$$ S(0) = \sum_j S^{(j)}(0) = \frac{l_0}{l_0} = 1, j \in \{c, h, r, o\}. $$

Table 5: England & Wales female general population two-decrement life table.

x	$d_x^{(c)}$	$d_x^{(o)}$	l_x
0-1	213	44097	10000000
1-4	1073	6807	9955690
5-9	1064	3226	9947810
10-14	1346	4274	9943520
15-19	1872	8908	9937900
20-24	2646	10384	9927120
25-29	3988	12542	9914090
30-34	6290	16450	9897560
35-39	13564	21136	9874820
40-44	23533	32087	9840120
45-49	43620	45210	9784500
50-54	73104	66816	9695670
55-59	114939	93951	9555750
60-64	174080	147690	9346860
65-69	242126	252514	9025090
70-74	315043	457087	8530450
75-79	390913	831207	7758320
80-84	410689	1331611	6536200
85-89	330792	1695808	4793900
90-94	187292	1546768	2767300
95-100	55287	768993	1033240
100+			208960

x - age span;

$d_x^{(j)}$ - the number of deaths due to cause of death j, $j \in \{c, o\}$, during the age interval x;

l_x - the number of living at the beginning of the age interval x;

Table 6: England & Wales female general population four-decrement life table.

x	$d_x^{(c)}$	$d_x^{(h)}$	$d_x^{(r)}$	$d_x^{(o)}$	l_x
0-1	213	670	1065	42362	10000000
1-4	1073	369	905	5533	9955690
5-9	1064	213	319	2695	9947810
10-14	1346	381	381	3513	9943520
15-19	1872	725	332	7851	9937900
20-24	2646	863	489	9032	9927120
25-29	3988	1735	761	10046	9914090
30-34	6290	2616	1030	12803	9897560
35-39	13564	4538	1455	15143	9874820
40-44	23533	8109	2571	21407	9840120
45-49	43620	13852	4502	26856	9784500
50-54	73104	22628	8185	36003	9695670
55-59	114939	34924	14672	44355	9555750
60-64	174080	60412	30146	57132	9346860
65-69	242126	116199	54753	81562	9025090
70-74	315043	219829	97070	140187	8530450
75-79	390913	407032	174312	249863	7758320
80-84	410689	655486	259676	416449	6536200
85-89	330792	827607	312966	555235	4793900
90-94	187292	692113	293420	561235	2767300
95-100	55287	284870	164274	319850	1033240
100+					208960

x - age span;

$d_x^{(j)}$ - the number of deaths due to cause of death j, $j \in \{c, h, r, o\}$, during the age interval x;

l_x - the number of living at the beginning of the age interval x;

Table 7: The crude survival functions $S^{(j)}(k)$, $j \equiv \{c, o\}$, obtained on the basis of the two decrement life table given in Table 5.

k	${}_kd_0^{(c)}$	${}_kd_0^{(o)}$	$S^{(c)}(k)$	$S^{(o)}(k)$	$S(k)$
0	–	–	0.2407	0.7593	1
1	213	44097	0.2407	0.7548	0.9956
5	1286	50904	0.2406	0.7542	0.9948
10	2350	54130	0.2405	0.7538	0.9944
15	3696	58404	0.2404	0.7534	0.9938
20	5568	67312	0.2402	0.7525	0.9927
25	8215	77695	0.2399	0.7515	0.9914
30	12202	90238	0.2395	0.7502	0.9898
35	18493	106687	0.2389	0.7486	0.9875
40	32057	127823	0.2375	0.7465	0.984
45	55591	159909	0.2352	0.7433	0.9785
50	99211	205119	0.2308	0.7387	0.9696
55	172315	271935	0.2235	0.7321	0.9556
60	287254	365886	0.212	0.7227	0.9347
65	461333	513577	0.1946	0.7079	0.9025
70	703460	766090	0.1704	0.6826	0.853
75	1018503	1223177	0.1389	0.6369	0.7758
80	1409416	2054384	0.0998	0.5538	0.6536
85	1820105	3385995	0.0587	0.4207	0.4794
90	2150897	5081803	0.0257	0.2511	0.2767
95	2338189	6628571	0.0069	0.0964	0.1033
100	2393476	7397564	0.0014	0.0195	0.0209

k - exact age in years;

${}_kd_0^{(j)}$ - the number of deaths due to cause of death j, $j \in \{c, o\}$, from age 0 to age k;

$S^{(j)}(k)$ - observed values at age k of the crude survival function for cause of death j, $j \in \{c, o\}$;

$S(k)$ - observed values at age k of the overall survival function;

Table 8: The crude survival functions $S^{(j)}(k)$, $j \equiv \{c, h, r, o\}$, obtained on the basis of the four decrement life table given in Table 6.

k	${}_kd_0^{(c)}$	${}_kd_0^{(h)}$	${}_kd_0^{(r)}$	${}_kd_0^{(o)}$	$S^{(c)}(k)$	$S^{(h)}(k)$	$S^{(r)}(k)$	$S^{(o)}(k)$	$S(k)$
0	–	–	–	–	0.2407	0.3427	0.1465	0.27	1
1	213	670	1065	42362	0.2407	0.3427	0.1464	0.2658	0.9956
5	1286	1038	1971	47895	0.2406	0.3426	0.1463	0.2652	0.9948
10	2350	1251	2290	50590	0.2405	0.3426	0.1463	0.265	0.9944
15	3696	1632	2670	54102	0.2404	0.3426	0.1462	0.2646	0.9938
20	5568	2356	3002	61953	0.2402	0.3425	0.1462	0.2638	0.9927
25	8215	3219	3491	70985	0.2399	0.3424	0.1461	0.2629	0.9914
30	12202	4954	4252	81031	0.2395	0.3422	0.1461	0.2619	0.9898
35	18493	7571	5282	93834	0.2389	0.342	0.146	0.2606	0.9875
40	32057	12109	6737	108977	0.2375	0.3415	0.1458	0.2591	0.984
45	55591	20218	9308	130384	0.2352	0.3407	0.1456	0.257	0.9785
50	99211	34070	13810	157239	0.2308	0.3393	0.1451	0.2543	0.9696
55	172315	56698	21996	193242	0.2235	0.3371	0.1443	0.2507	0.9556
60	287254	91621	36667	237598	0.212	0.3336	0.1428	0.2463	0.9347
65	461333	152033	66814	294730	0.1946	0.3275	0.1398	0.2405	0.9025
70	703460	268232	121567	376292	0.1704	0.3159	0.1343	0.2324	0.853
75	1018503	488061	218637	516480	0.1389	0.2939	0.1246	0.2184	0.7758
80	1409416	895093	392949	766342	0.0998	0.2532	0.1072	0.1934	0.6536
85	1820105	1550579	652625	1182791	0.0587	0.1877	0.0812	0.1517	0.4794
90	2150897	2378186	965591	1738026	0.0257	0.1049	0.0499	0.0962	0.2767
95	2338189	3070299	1259011	2299262	0.0069	0.0357	0.0206	0.0401	0.1033
100	2393476	3355169	1423284	2619111	0.0014	0.0072	0.0042	0.0081	0.0209

k - exact age in years;

${}_kd_0^{(j)}$ - the number of deaths due to cause of death j, $j \in \{c, h, r, o\}$, from age 0 to age k;

$S^{(j)}(k)$ - observed values at age k of the crude survival function for cause of death j, $j \in \{c, h, r, o\}$;

$S(k)$ - observed values at age k of the overall survival function;