Insurance: Mathematics and Economics 53 (2013) 464–477

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Dependent competing risks: Cause elimination and its impact on survival

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

Cass Business School, City University London, UK

HIGHLIGHTS

- We study the dependent competing risks model of human mortality.
- The dependence between lifetimes is modelled by a multivariate copula function.
- The effect on the overall survival of removing one or more causes of death is explored.
- Two alternative definitions of removal are considered (*ignoring* and *eliminating*).
- The eliminating definition is better suited for practical use and more intuitive.

ARTICLE INFO

Article history:
Received January 2012
Received in revised form
May 2013
Accepted 21 July 2013

Keywords:
Dependent competing risks model
Lifetimes
Failure times
Overall survival function
Copula functions
Cause elimination
Cause removal

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 the 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 the 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.

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 contributions in this strand of literature are the papers by Elandt-Johnson (1976), and also by Yashin et al. (1986) who consider the conditional independence of the times to death, given an assumed stochastic covariate process. Tsiatis (1975) shows that it is

* Correspondence to: Faculty of Actuarial Science and Insurance, Cass Business School, City University London, 106 Bunhill Row, EC1Y 8TZ London, UK. Tel.: +44 2070408470.

E-mail address: d.dimitrova@city.ac.uk (D.S. Dimitrova).

0167-6687/$ – see front matter
http://dx.doi.org/10.1016/j.insmatheco.2013.07.008

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 that 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 model 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 on 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, Anderson (1999). Several alternative four-dimensional copula models underlying the joint distribution of the lifetimes 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 the 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* definition 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 us 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 the 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