Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Competing risk model with bivariate random effects for clustered survival data

Xin Lai [a,b], Kelvin K.W. Yau [c,*], Liu Liu [d]

[a] School of Management, Xi'an Jiaotong University, China

[b] JC School of Public Health & Primary Care and Shenzhen Research Institute, Chinese University of Hong Kong, China

[c] Department of Management Sciences, City University of Hong Kong, Hong Kong

[d] School of Mathematics and V.C. & V.R. Key Lab, Sichuan Normal University, China

ARTICLE INFO

Article history:
Received 19 November 2015
Received in revised form 4 January 2017
Accepted 8 March 2017
Available online xxxx

Keywords:
Bivariate random effects
Competing risk
Cause-specific hazard model
GLMM

ABSTRACT

Competing risks are often observed in clinical trial studies. As exemplified in two data sets, the bone marrow transplantation study for leukaemia patients and the primary biliary cirrhosis study, patients could experience two competing events which may be correlated due to shared unobservable factors within the same cluster. With the presence of random hospital/cluster effects, a cause-specific hazard model with bivariate random effects is proposed to analyse clustered survival data with two competing events. This model extends earlier work by allowing random effects in two hazard function parts to follow a bivariate normal distribution, which gives a generalized model with a correlation parameter governing the relationship between two events due to the hospital/cluster effects. By adopting the GLMM formulation, random effects are incorporated in the model via the linear predictor terms. Estimation of parameters is achieved via an iterative algorithm. A simulation study is conducted to assess the performance of the estimators, under the proposed numerical estimation scheme. Application to the two sets of data illustrates the usefulness of the proposed model.

1. Introduction

Competing risks arise when individuals are exposed to a number of failure causes and the occurrence of one type of event removes other type of events from being observed; see for example in Kalbfleisch and Prentice (2002) and Pintilie (2006). Over the past decades, two classes of competing risk models have been intensively studied through Cox's proportional hazards models; one is to model the cumulative incidence function which focuses on the cumulative probability of the occurrence of a given event (Fine and Gray, 1999; Kim, 2007; Ha et al., 2016), and the other is referred to as the cause-specific hazard models which analyse the effects of covariates on a particular cause of failure (Prentice et al., 1978; Andersen and Borgan, 1985; Keiding et al., 2001; Lee et al., 2014).

In this paper, our interest is the modelling of cause-specific hazard for multi-centre competing risk data. Under multi-centre setting, individuals within a centre are often affected by some unobservable characteristics of the centre. Thus, the observed outcomes within the same centre may be correlated due to such a shared characteristic across individuals (Lai and Yau, 2010). In many previous studies, the dependence structure among outcomes was accommodated by incorporating

* Correspondence to: Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong. Fax: +852 3442 0189.
E-mail address: mskyau@cityu.edu.hk (K.K.W. Yau).

http://dx.doi.org/10.1016/j.csda.2017.03.011
0167-9473/

the random effects into the underlying statistical model; see Stiratelli et al. (1984) and Zeger et al. (1988). Lee et al. (2014) introduced the random effects into the cause-specific hazard model to describe the within-centre correlation, in which the proportional hazards assumption was used. From the perspective of censoring, the occurrence of competing risk could be taken as one of the censoring mechanisms (Li et al., 2008). Huang and Wolfe (2002) proposed an informative censoring frailty model to study the multi-cluster competing risks data, where the dependence between two competing events and the correlation within cluster were simultaneously modelled by normally distributed random effects. However, in these studies, the random effects between competing events are assumed either independent (Lee et al., 2014) or perfect linearly correlated (Huang and Wolfe, 2002).

In practice, the random effects between competing events may share some commonality because they originate from the same cluster. This motivates our study of a competing risk model with more general correlation in random effects between competing events. In a standard competing risk model with two causes of failure, our work of this paper introduces bivariate random effects into the cause-specific hazard model, where random effects between two competing events are assumed to follow a bivariate normal distribution. As special cases, it reduces to the settings in Lee et al. (2014) and Huang and Wolfe (2002) when the correlation parameter between two competing events is fixed as zero and ± 1 respectively. Through such extension, we can obtain an estimation of the correlation coefficient, which represents the association between two competing risks due to the cluster effect.

In Section 2, the cause-specific competing risk model with bivariate random effects is introduced. The estimation procedure for regression and variance component parameters is outlined in Section 3. In Section 4, a simulation study is conducted to evaluate the performance of the competing risk model with bivariate random effects. In Section 5, the bone marrow transplantation and the primary biliary cirrhosis (PBC) data sets, which motivate our study, are analysed to illustrate the applicability of proposed model. Further discussions are presented in Section 6.

2. Cause-specific hazard model with bivariate random effects

Suppose that the observed data with censoring are collected from M hospitals (or clusters). In each hospital, we assume that there are K distinct causes of event. Let T_{ij}^* denote the underlying time to the first event for patient j in hospital i and let $\varepsilon_{ij} \in (1, \ldots, K)$ be the corresponding cause of event. For the usual right censored data, we observe $T_{ij} = \min(T_{ij}^*, C_{ij})$, where C_{ij} is the censoring time. Given covariate X_{ij}, the cause-specific hazard function for cause k at time t is defined as

$$\lambda_k\left(t; X_{ij}\right) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T_{ij}^* < t + \Delta t, \varepsilon_{ij} = k | T_{ij}^* \ge t, X_{ij})}{\Delta t}. \tag{1}$$

We further define $\delta_{ij} = I(T_{ij}^* \le C_{ij})\varepsilon_{ij}$ with $I(\cdot)$ denote the indicator function, thus $\delta_{ij} \in (0, 1, \ldots, K)$. When the causes of event are known for all patients, the observed data consist of $(T_{ij}, \delta_{ij}, X_{ij},), j = 1, \ldots, n_i, i = 1, \ldots, M$ and $\sum_{i=1}^{M} n_i = N$.

For simplicity, this paper considers two causes of events ($k = 1$ or 2). In the standard cause-specific hazard analysis (Prentice et al., 1978), the hazard function for one cause (e.g. cause 1) is assumed to be independent of another cause (e.g. cause 2). In other words, the events from cause 2 are treated as independently censored data when analysing the effect of risk factors on the hazard function for cause 1, and vice versa. However, the events in competing risk scenario are very often correlated. For example, the graft-versus-host disease (GVHD) was found to be associated with relapse or treatment-related mortality in bone marrow transplant study (Lee et al., 2014). Furthermore, the dependence structure within the same cluster may also induce the correlation between two competing events (Katsahian et al., 2006). In this study, the random effects are employed to describe the correlation of hazard functions between two competing events. Specifically, we follow the Generalized Linear Mixed Model (GLMM) method (McGilchrist, 1994) to introduce random effects into both cause-specific proportional hazards functions (Cox, 1972)

$$\begin{aligned} \lambda_1\left(t; X_{ij}\right) &= \lambda_{01}(t)\exp(x_{ij}'\beta_1 + U_i) \\ \lambda_2\left(t; X_{ij}\right) &= \lambda_{02}(t)\exp(x_{ij}'\beta_2 + V_i) \end{aligned} \tag{2}$$

where x_{ij} is the vector covariate, β_k, $k = 1$ or 2, is the corresponding factor effect on hazard function for failure event k, and U_i and V_i, $i = 1, \ldots, M$ are random effects of hospital i on event 1 and event 2 respectively which are assumed to follow the bivariate normal distribution $N(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} \theta_1 & \rho\sqrt{\theta_1\theta_2} \\ \rho\sqrt{\theta_1\theta_2} & \theta_2 \end{pmatrix}.$$

The positive correlation ($\rho > 0$) implies that the patient with high hazard of experiencing one type of event is likely to have high chance in developing another type of failure event; while the negative correlation suggests that the increase in the hazard of one event may lower the risk of another event. When $\rho = -1$ or $+1$, the cause-specific competing risk model degenerates to the model in Huang and Wolfe (2002); when $\rho = 0$, it degenerates to the cause-specific model (Prentice et al., 1978) with independent random effects (Lee et al., 2014). Therefore, the proposed model can flexibly handle the correlation ranging from -1 to $+1$.

Please cite this article in press as: Lai, X., et al., Competing risk model with bivariate random effects for clustered survival data. Computational Statistics and Data Analysis (2017), http://dx.doi.org/10.1016/j.csda.2017.03.011