South African Statistical Journal
Vol. 53, No. 2, 73–86

A MIXTURE MODEL WITH APPLICATION TO DISCRETE COMPETING RISKS DATA

Bonginkosi D. Ndlovu
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa
e-mail: *bongi@dut.ac.za*

Sileshi F. Melesse
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Temesgen Zewotir
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

In this paper, we modify the continuous time mixture competing risks model (Larson and Dinse, 1985) to handle discrete competing risks data. The main result of the model is an alternate regression expression for the cumulative incidence function. The structure of the regression expression for the cumulative incidence function under this model, and the proportional hazards assumption for the conditional hazard rates with piece-wise constant baseline conditional hazards, combine to allow for another means to assess the covariate effects on the cumulative incidence function. This benefit comes at some computational costs because the parameters are estimated via an EM algorithm. The proposed model is applied to real data and it is found that it improves the exercise of evaluating the covariate effects on the cumulative incidence function compared to other discrete competing risks models.

Key words: Discrete time competing risks, Mixture competing risks model, Poisson regression model.

1. Introduction

Competing risks data arises in survival analysis experiments when subjects are at risk of failing from more than one cause of failure. The most popular approach has been to summarize the observed data on the pair $(T; D)$ via cause-specific hazards (Prentice et al., 1978). With J failure types, a typical format of the observed data is: $\boldsymbol{y}_i = (t_i, \boldsymbol{x}_i^T, D_i)^T$, $i = 1, \dots, n$, where t_i is a failure time when $D_i \in \{1, 2, \dots, J\}$ or a censoring time when $D_i = 0$, and $\boldsymbol{x}_i$ is a p-dimensional vector of covariates. With covariates, the cause-specific hazards are in most instances modelled via the Cox type regression (Cox, 1972), and the regression expression for the cumulative incidence function is then derived from the cause-specific hazards. These quantities, namely the cause-specific hazards and the cumulative incidence functions, have been the most quoted summary quantities for competing risks data, and

[1] Corresponding author.
MSC2010 subject classifications. 62-07.

the assessment of covariate effects on the competing risks process is conducted via these quantities. It is well documented in the literature, however, that the regression expression for the cumulative incidence function constructed with cause-specific hazards complicates the assessment of covariate effects. This has given impetus to the development of other regression models for the cumulative incidence function, such as the *transformation models* (Gray, 1988; Fine and Gray, 1999; Scheike, Zhang and Gerds, 2008; Klein and Andersen, 2005), where the cumulative incidence function is directly modelled on covariates.

The mixture competing risks model is one of the models under which an alternate regression model for the cumulative incidence function arises. This model, the subject of this article, was introduced in the continuous time domain, but we intend to modify it and present it as a discrete time competing risks model. The multinomial model (Ambrogi, Biganzoli and Boracchi, 2009; Tutz, 1995; Tutz and Schmid, 2016) is the most popular discrete time competing risks model. Here, the cause-specific hazards are estimated simultaneously by applying the multinomial distribution to the data in period-person format. An alternate approach to the multinomial model is to estimate the cause-specific hazards individually by fitting $J - 1$ binomial distributions; see, for example Lee, Feuer and Fine (2018) where, as in continuous time (Prentice et al., 1978), the cause-specific failure times are treated as events and the competing failure times are regarded as censoring times. The regression model for the cumulative incidence function under the mixture model, together with certain distributional assumptions for time to failure, combine to allow for an alternate means to assess the covariate effects on the cumulative incidence function. When the multinomial model or its alternative is assumed, the regression model for the cumulative incidence function is also derived from the cause-specific hazards, and therefore suffers similar shortcomings in relation to the assessment of covariate effects.

The mixture model was introduced to survival analysis in the context of competing risks models by Larson and Dinse (1985), and the more popular *mixture cure* model (see, for example, Kuk and Chen, 1992; Sy and Taylor, 2000; Peng and Dear, 2000; Peng, 2003) is an adaptation of this model. The mixture competing risk model assumes that the bivariate distribution of $(T; D)$ factorizes into two distributions, one for time to failure conditional on failure type and a marginal distribution for failure type, i.e. $P(T; D) = P(T|D)P(D)$. The conditional failure time distributions arise according to the failure types, and a given conditional distribution is summarized via what we term the *conditional hazard function*. The conditional hazard functions and the failure type probabilities replace the causes-specific-hazards as the summary quantities for observed competing risks data under the mixture model, in fact, the regression model for the cumulative incidence function is constructed from the regression models for these quantities in the place of the regression model for the cause-specific hazards. Towards the regression model for the conditional hazards, Larson and Dinse (1985) assumed *proportional hazards* with *piece-wise constant* conditional baseline hazards, i.e.,

$$\lambda_j(t; \boldsymbol{x}; \boldsymbol{\beta}_j) = \exp\left(\sum_{s=1}^{q} (\beta_{0js}) \mathbf{1}_s(t) + \boldsymbol{x}^T \boldsymbol{\beta}_{1j}\right), \qquad j = 1 \ldots J,$$

where $\mathbf{1}_s(t) = 1$ when $s = t$ and zero otherwise. Here, *follow up* is partitioned into q mutually exclusive and exhaustive intervals of the form: $[a_0, a_1), \ldots, [a_{q-1}, a_q)$. Let $\boldsymbol{\beta}_{0j}$ and $\boldsymbol{\beta}_{1j}$ represent the duration and the regression coefficients respectively, where $\boldsymbol{\beta}_{0j} = (\beta_{0j1}, \ldots, \beta_{0jq})^T$ and $\boldsymbol{\beta}_{1j} = (\beta_{1j1}, \ldots, \beta_{1jp})^T$ so that $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{0j}^T, \boldsymbol{\beta}_{1j}^T)^T$ becomes a vector of the parameters that describe the j^{th} conditional hazard throughout follow up. Note that this characterization of follow up corresponds to grouped survival times in discrete time with the distinction that failure/censoring times are known exactly in continuous time, but are unknown in discrete time. We also assume that the conditional baseline hazards are also piece-wise constant in this article.

The authors modelled the failure type probabilities on covariates via the multinomial distribution and the regression model is given by

$$\pi_j(\boldsymbol{x};\gamma) = \frac{\exp(\gamma_{0j} + \boldsymbol{x}^T\gamma_{1j})}{1 + \sum_{j=1}^{J-1}\exp(\gamma_{0j} + \boldsymbol{x}^T\gamma_{1j}}, \qquad j = 1, \ldots J-1,$$

where $\pi_J(\boldsymbol{x};\gamma) = 1 - \sum_{j=1}^{J-1}\pi_j(\boldsymbol{x};\gamma)$, $\gamma = (\gamma_1^T, \ldots, \gamma_{J-1}^T)^T$ and $\gamma_j = (\gamma_{0j}, \gamma_{1j}^T)^T$. Let all the parameters that describe the mixture model be represented by $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \gamma^T)^T$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_j^T)$.

The key result that arises from the factorization assumption under the mixture model is the regression model for the cumulative incidence function, and is given by

$$F_j(t|\boldsymbol{x};\gamma,\boldsymbol{\beta}_j) = \pi_j(\boldsymbol{x},\gamma)(1 - S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)), \qquad j = 1, \ldots J.$$

It is this form of the regression model for the cumulative incidence function, together with the proportional hazards assumption, albeit at the interval level, that combine to allow for an alternate means of assessing the covariate effects on the cumulative incidence function in continuous time.

Other authors have also studied the mixture competing risks model and the various forms of the model differ in terms of how the conditional baseline hazards are specified. The parametric formulation of the model, that is when the baseline conditional hazards are modelled with the usual lifetime distributions, has been studied by, amongst others, Lau, Cole, Moore and Gange (2008); Lau, Cole and Gange (2011) and the large sample properties of the model have been considered by Maller and Zhou (2002), with Ng, McLachlan, Yau and Lee (2004) considering the parametric model with clustering. Furthermore, Ng and McLachan (2003) and Escarela and Bowater (2008) have studied the semi-parametric formulation of the model. The model has also been extended to model competing risks data with immune subjects (Choi and Zhou, 2002; Zhiping, 2011).

It has been shown in univariate survival analysis models that the parameters which correspond to the hazard function can be estimated via a certain Poisson regression model when the baseline hazards are assumed to be piece-wise constant (Laird and Olivier, 1981; Holford, 1980). We also show that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \ldots, \boldsymbol{\beta}_j^T)^T$, the parameters for the conditional hazards, can also be estimated via the same Poisson regresssion model as pointed by one of the referees to the same article by Larson and Dinse (1985).

The remainder of this article is organized as follows: we begin by reviewing this equivalence between Poisson regression and the piece-wise constant baseline hazards, particularly the extension to the mixture model in Section 2, and discuss the adjustments to the model to enable it to handle discrete competing risks data. This is followed by an illustration in Section 3, and we conclude with a discussion in Section 4. The standard errors under the EM algorithm have to be adjusted to reflect the observed data standard errors, the workings are presented in the Appendix.

2. Estimation

The unknown parameters of the mixture model are estimated by maximizing the observed data likelihood function. In constructing an observed data likelihood function from the observed data $\boldsymbol{y} =$

$(\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \ldots, \boldsymbol{y}_n^T)^T$, a subject i that fails at time t_i from failure type j, contributes $\pi_j(\boldsymbol{x}_i;\gamma)f_j(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}_j)$ to the likelihood function whilst a censored one contributes $S(t_i|\boldsymbol{x}_i;\boldsymbol{\theta})$. The observed data log-likelihood function is then given by

$$\mathcal{L}_0(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}\big(\log \pi_j(\boldsymbol{x}_i;\gamma) + \log f_j(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}_j)\big) + d_{i0}\log S(t_i|\boldsymbol{x}_i;\boldsymbol{\theta}),$$

where $d_{ij} = I(D_i = j)$ and $d_{i0} = 1 - \sum_{j=1}^{J} d_{ij}$. We pose this as a missing information problem regarding the eventual failure status of censored subjects so as to apply the EM algorithm. We introduce a pseudo variable $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iJ})^T, i = 1, \ldots, n$, where v_{ij} is 1 or 0 according to whether a censored subject i eventually fails by cause j or not. Assuming that $\boldsymbol{v}_i$ is observed, the expression for the complete data log-likelihood can then be written as

$$L_c(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} g_{ij}\log \pi_j(\boldsymbol{x}_i;\gamma) + d_{ij}\log \lambda_j(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}_j)\big) + g_{ij}\log S_j(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}_j),$$

where $g_{ij} = d_{ij} + v_{ij}d_{i0}$. Since g_{ij} is linear in the log-likelihood of the complete data, the E-step entails replacing g_{ij} with $\underline{g}_{ij} = d_{ij} + \underline{v}_{ij}d_{i0}$, which is its expectation conditional on $\boldsymbol{y}$ and $\boldsymbol{\theta}^0$, where $\boldsymbol{\theta}^0$ is MLE of $\boldsymbol{\theta}$ at the previous M-Step. The conditional expectation $\underline{v}_{ij}$ of v_{ij} is given by

$$\underline{v}_{ij} = \mathrm{E}(v_{ij}|\boldsymbol{\theta}^0;\boldsymbol{y}) = \frac{\pi_j(\boldsymbol{x}_i;\gamma^0)S_j(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}_j^0)}{\sum_{l=1}^{J}\pi_l(\boldsymbol{x}_i;\gamma^0)S_l(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}_l^0)}.$$

Introducing the Q notation, the conditional expectation of the log-likelihood for the complete data can be written as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0) = \sum_{i=1}^{n}\sum_{j=1}^{J} \underline{g}_{ij}\log \pi_j(\boldsymbol{x}_i;\gamma) + d_{ij}\log \lambda_j(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}_j) + \underline{g}_{ij}\log S_j(t_i|\boldsymbol{x}_i;\boldsymbol{\beta}_j).$$

Let Δ_{is} represent the "exposure" or the time spent "alive" in the interval $[a_s, a_{s-1})$ by subject i, and define it as $\Delta_{is} = \min(s, a_s) - a_{s-1}$. Thus, a subject i that fails or is censored at time s during this interval has $\Delta_{is} = s - a_{s-1}$ as its exposure, or has $\Delta_{is} = a_s - a_{s-1}$ if it survives the interval. With these definitions, the conditional survival function can then be written as $S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j) = \exp(-\Lambda_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)) = \exp-(\sum_{s=1}^{t}\Delta_{is}\lambda_j(s|\boldsymbol{x};\boldsymbol{\beta}_j))$. Furthermore, if we define $d_{ijs} = 0$ for $s \le t_i - 1$ and $d_{ijt_i} = d_{ij}$, as well as $\underline{g}_{ijs} = \underline{g}_{ij}$ for $s = 1, \ldots, t_i$, the Q function can be written as

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0) &= \sum_{i=1}^{n}\sum_{j=1}^{J} \underline{g}_{ij}\log \pi_j(\boldsymbol{x}_i;\gamma) + \sum_{j=1}^{J}\left\{\sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{ijs}\log \lambda_j(s|\boldsymbol{x}_i;\boldsymbol{\beta}_j) - \underline{g}_{ijs}\Delta_{is}\lambda_j(s|\boldsymbol{x}_i;\boldsymbol{\beta}_j)\right\} \\ &= Q_0(\gamma|\gamma^0) + \sum_{j=1}^{J} Q_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0). \end{aligned}$$

It can easily be seen that $Q_0(\gamma|\gamma^0)$ is a kernel of a multinomial log-likelihood function with $(\underline{g}_{i1}, \ldots \underline{g}_{iJ}) \sim \mathcal{M}(1, \pi_1(\boldsymbol{x}_i;\gamma), \ldots \pi_J(\boldsymbol{x}_i;\gamma))$ for subject i. Furthermore, in likelihood, $Q_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)$ is equivalent to

$$\acute{Q}_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0) = \sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{ijs}\log(\underline{g}_{ijs}\Delta_{is}) + d_{ijs}\log\lambda_j(s|\boldsymbol{x}_i;\boldsymbol{\beta}_j) - \underline{g}_{ijs}\Delta_{is}\lambda_j(s|\boldsymbol{x}_i;\boldsymbol{\beta}_j),$$

where $d_{ijs} \sim \mathcal{P}(\underline{g}_{ijs}\Delta_{is}\lambda_j(s|\boldsymbol{x}_i;\boldsymbol{\beta}_j))$, because the difference is $d_{ijs}\log(\underline{g}_{ijs}\Delta_{is})$, a constant term. Therefore, the unknown parameters of a mixture model that correspond to the conditional hazards, with piece-wise constant baseline conditional hazards, can be estimated via a Poisson regression model: $d_{ijs} \sim \mathcal{P}(\underline{g}_{ijs}\Delta_{is}\lambda_j(s|\boldsymbol{x}_i;\boldsymbol{\beta}_j))$ with $\log(\underline{g}_{ijs}\Delta_{is})$ as an offset. In continuous time, even though the survival times are grouped into intervals, the exposure for each subject can be computed exactly because failure/censored times are known, but this is not the case when survival times are grouped into intervals in discrete time. In the absence of information regarding failure times, we assume they occur halfway through the interval (Vermunt, 1997). It is customary to assume that censoring occurs at the end of the interval. If the failure time for subject i is t_i, meaning that $a_{t-1} \leq t_i < a_t$, its total exposure is $\sum_{s=1}^{t_i-1}(a_s - a_{s-1}) + \frac{1}{2}(a_t - a_{t-1})$, otherwise, if t_i is a censoring time, the total exposure is $\underline{v}_{ij}\sum_{s=1}^{t_i}(a_s - a_{s-1})$. This is the required adjustment to enable the model to handle discrete data.

The M-step entails maximizing J Poisson likelihoods $\tilde{Q}_j(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)$, $j = 1,\ldots J$, individually within the GLM framework and a multinomial likelihood $Q_0(\gamma|\gamma^0)$. Because some statistical packages cannot handle a multinomial distribution with fractional responses, an alternative is to fit $J-1$ binomial distributions of the form $g_{ij} \sim \mathcal{B}(1,\pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma}_j))$ with the most prevalent failure type as the reference category to minimize the standard errors (Becg and Gray, 1984). To perform the M-step, the failure time data is first re-arranged into a period-person format; see, for example, Allison (1982) and Singer and Willet (1993). The `discSurv` R package (Welchowski and Schmid, 2019) can be used to facilitate the conversion of data into a period-person format. Under this format, subject $i \in \{1,\ldots n\}$ contributes d_{ijs}, $s = 1,\ldots t_i$, as the response variable with the corresponding exposure given as $\underline{g}_{ij}\Delta_{is}$ and an accompanying vector of covariate $\boldsymbol{x}_i$ repeated t_i times.

Let $\grave{\boldsymbol{y}}_j$, $j = 1,\ldots,J$, be failure time data according to cause j in the period-person format, let $\grave{\boldsymbol{y}}_0$ be the censored subjects also in the period-person format, and define $\tilde{\boldsymbol{y}}_j = (\grave{\boldsymbol{y}}_j^T, \grave{\boldsymbol{y}}_0^T)^T$. The failure type data $\boldsymbol{y}$ is the original data in the original format $\boldsymbol{y} = (\boldsymbol{y}_1^T,\ldots \boldsymbol{y}_n^T)^T$. With these definitions, the EM algorithm steps can then be summarized as follows:

1) Update the failure type data with $\boldsymbol{d}_i = \underline{\boldsymbol{g}}_i^{(r)} = (\underline{g}_{i1},\ldots,\underline{g}_{i(J-1)})^T$, $i = 1,\ldots,n$ as the response vector and the failure time data with $\underline{g}_{ij}^{(r)}\boldsymbol{\Delta}_i$ as the exposure vector in the $(r)^{\text{th}}$ iteration of the E-Step, where $\boldsymbol{\Delta}_i = (\Delta_{i1},\ldots,\Delta_{it_i})^T$.

2) Fit J Poisson distributions to $\tilde{\boldsymbol{y}}_j$ $(j = 1, 2\ldots, J)$ seperately to determine $\boldsymbol{\beta}_j^{(r+1)}$, $j = 1,\ldots,J$, and a multinomial distribution to the failure type data to determine $\boldsymbol{\gamma}^{(r+1)}$ in the corresponding M-Step, then proceed to Step 1.

To initialize the algorithm:

1) In the 0^{th} iteration of the E-Step, let $\boldsymbol{d}_i = (d_{i1},\ldots,d_{iJ-1})$, $i = 1,\ldots,n$ be the response vector for failure type data, and $\boldsymbol{\Delta}_i = (\Delta_{i1},\ldots,\Delta_{it_i})^T$, $i = 1,\ldots,n$ be the exposure vector for the failure time data. In the corresponding M-Step, fit J Poisson distributions to the failure time

data seperately to determine $\beta_j^{(1)}$, $j = 1, \ldots, J$, and a multinomial distribution to the failure type data to determine $\gamma^{(1)}$, then proceed to Step 1.

3. Illustration

We apply the proposed model to the unemployment data, originally analyzed by McCall (1996), given as '*UnempDur*' in `Ecdat` (Croissant, 2015) `R` package. The data focusses on the length of unemployment spell (two-week intervals) until transition to either full-time or part-time employment for unemployed individuals that were fully employed in their lost jobs. The covariates in the data set are Age, Unemployment Insurance, Displacement rate, Replacement rate, Wage rate and Tenure at the lost job.

Out of a sample of 3343 subjects, 676 are excluded for lack of complete information. Eventually out of a sample size of 2667, about 40% of the subjects exit to full-time employment, 13% to part-time employment, and 47% are censored.

The time in two-week interval is $t \in \{1, 2, 3, \ldots, 28\}$. We consider $t \in \{1, 2, 3, \ldots, 19, 20\}$ by collapsing the event/censoring time $t \geq 20$ into one interval $[20, 29)$ because there are relatively few events in this interval. Furthermore, some of the later intervals have no events, this should be avoided as it brings about instability to the model. With the exception of the "unemployment benefits" (UI) variable, all the variables are continuous. For the purposes of computing exposure, an interval length is taken to be 1 so that a failure time and censoring time have 0.5 and 1 exposure respectively, and this also applies to the interval $[20, 29)$.

This data was analysed by McCall (1996) to test the proposition that increasing the Disregard for the Unemployment Insurance (UI) recipients will encourage these individuals to seek part-time employment instead of full-time employment. Disregard is defined as the maximum amount that a UI recepient can earn from part-time employment with no reduction in benefits. One of the models considered for analysis in McCall (1996) was the ordinary (Prentice and Gloecker, 1978) competing risks model, where cause-specific hazards are estimated by treating the cause-specific failures as events and the competing events as censored. The author chose to connect the cause-specific hazards to the covariates via the complementary log link function (Prentice and Gloecker, 1978), that is,

$$h_j(t) = 1 - \exp(-\exp(\alpha_{0jt} + \boldsymbol{x}^T \boldsymbol{\alpha}_{1j})),$$

for $j = 1, 2$ and $t = 1, \ldots, 19$. We refer to this model as the ordinary competing risks model. The covariates that were considered were Unemployment Insurance (UI), Disregard Rate (DR), Replacement Rate (RR) and the interaction terms UI $\times$ DR as well as UI $\times$ RR. The Disregard Rate is the weekly Disregard amount divided by the weekly earnings in the lost job and the Replacement Rate is the weekly benefit divided by weekly earnings in the lost job. We also fitted our proposed model using the same covariate set. For convenience, the continuous covariates, that is DR and RR, were centered at 0.107 and 0.452, their respective average values, with UI recepients set as the reference category. We found that the interaction terms were insignificant. We then re-fitted the models without the interaction terms and the results are displayed in Tables 1 and 2. We plotted the cumulative incidence functions from the proposed model and the ordinary model in Figure 1.

It is evident that increasing the Disregard Rate by 50% lowers the probability of exiting the unemployment state to full-time employment significantly when one examines both plots from the

Table 1. Maximum likelihood estimates (with standard errors) for the mixture model (Poisson model) and the ordinary competing risks model (* denotes $P < 0.05$).

	Mixture Model (Poisson Model)				**Ordinary Model**	
Coefficient	$\hat{\beta}_1$	$\hat{S}_{10}$	$\hat{\beta}_2$	$\hat{S}_{20}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
T1	−2.561 (0.074)*	0.926	−2.361 (0.124)*	0.910	−2.718 (0.072)*	−3.946 (0.128)*
T2	−2.771 (0.085)*	0.869	−2.776 (0.149)*	0.855	−2.996 (0.084)*	−4.245 (0.151)*
T3	−2.916 (0.098)*	0.824	−3.031 (0.176)*	0.815	−3.187 (0.098)*	−4.429 (0.176)*
T4	−3.446 (0.138)*	0.798	−3.425 (0.226)*	0.789	−3.742 (0.137)*	−4.777 (0.226)*
T5	−2.636 (0.103)*	0.743	−2.772 (0.179)*	0.741	−2.962 (0.103)*	−4.122 (0.179)*
T6	−3.591 (0.179)*	0.722	−3.839 (0.321)*	0.725	−3.930 (0.179)*	−5.139 (0.321)*
T7	−2.419 (0.112)*	0.660	−2.831 (0.212)*	0.684	−2.798 (0.112)*	−4.133 (0.211)*
T8	−3.909 (0.259)*	0.648	−3.623 (0.338)*	0.666	−4.312 (0.259)*	−4.859 (0.338)*
T9	−3.007 (0.176)*	0.617	−3.788 (0.382)*	0.651	−3.422 (0.177)*	−5.021 (0.382)*
T10	−5.261 (0.578)*	0.613	−4.490 (0.580)*	0.643	−5.677 (0.577)*	−5.705 (0.579)*
T11	−2.984 (0.198)*	0.583	−3.904 (0.451)*	0.631	−3.413 (0.199)*	−5.118 (0.451)*
T12	−4.135 (0.379)*	0.574	−4.269 (0.580)*	0.622	−4.571 (0.379)*	−5.459 (0.580)*
T13	−2.742 (0.202)*	0.538	−3.197 (0.358)*	0.597	−3.198 (0.203)*	−4.397 (0.358)*
T14	2.306 (0.184)*	0.487	−3.315 (0.412)*	0.576	−2.802 (0.186)*	−4.483 (0.412)*
T15	−2.427 (0.231)*	0.446	−3.823 (0.581)*	0.563	−3.001 (0.233)*	−4.906 (0.579)*
T16	−2.828 (0.317)*	0.420	−3.650 (0.581)*	0.549	−3.437 (0.318)*	−4.693 (0.580)*
T17	−2.799 (0.354)*	0.395	−4.556 (1.002)*	0.543	−3.431 (0.354)*	−5.584 (1.002)*
T18	−2.660 (0.378)*	0.369	−3.750 (0.709)*	0.530	−3.401 (0.379)*	−4.722 (0.711)*
T19	−1.149 (0.213)*	0.269	−2.108 (0.339)*	0.469	−2.083 (0.219)*	−3.056 (0.339)*
UI	1.524 (0.064)*		0.568 (0.111)*		1.069 (0.063)*	1.153 (0.113)*
DR	−0.595 (0.494)		−1.809 (0.835)*		−1.823 (0.499)*	−0.379 (0.793)
RR	−0.071 (0.292)		−0.009 (0.570)		−0.759 (0.291)*	0.908 (0.507)

Table 2. Maximum likelihood estimates (with standard errors) for the mixture model (logistic model) (* denotes $P < 0.05$).

	Mixture Model (Logistic Model)	
Coefficient	$\hat{\gamma}$	
Constant	1.094	(0.059)*
UI	−0.526	(0.087)*
DR	−2.941	(0.654)*
RR	−2.148	(0.467)*

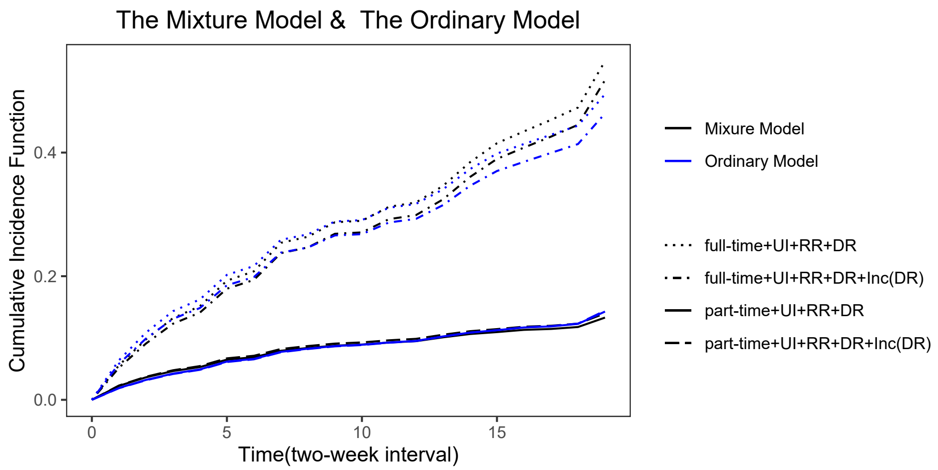

Figure 1. The cumulative incidence function of exit to full time and part time employment for the UI recipients with the effect of increasing DR via the mixture model and the ordinary model.

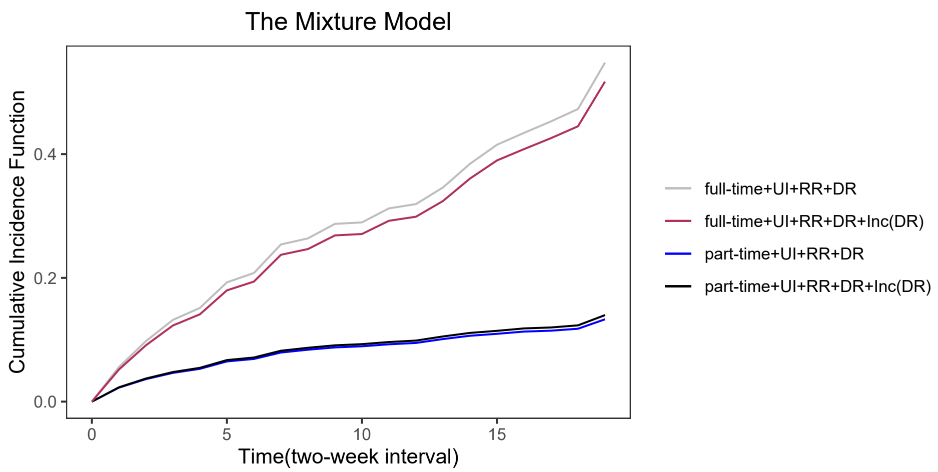

Figure 2. The cumulative incidence function of exit to full time and part time employment for the UI recipients with the effect of increasing DR via the mixture model.

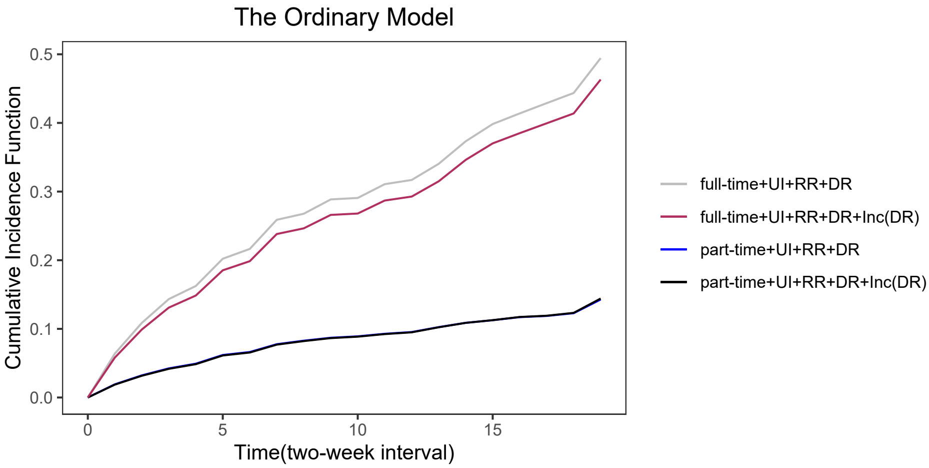

Figure 3. The cumulative incidence function of exit to full time and part time employment for the UI recipients with the effect of increasing DR via the ordinary model.

proposed model (Figure 3) and the ordinary model (Figure 1), but there seems to be no noticeable difference in the probability to exit to part-time before and after the increase in the Disregard Rate if we examine Figure 3. There is an increase in the probability of exiting the unemployment state to part-time employment, albeit marginal, if one considers Figure 1. The conclusions that are reached by examining the plot from the proposed model agree with McCall's (1996) own conclusions, despite the fact that we have excluded the interaction terms, that is, increasing the Disregard Rate tends to encourage unemployed individuals away from searching for full-time employment towards looking for part-time employment prospects. The fact that the ordinary model suggests that increasing the Disregard has no effect on UI recipients with regard to part-time employment, i.e., that the findings disagree with McCall (1996), may be attributable to discrepancies in certain aspects of the data sets used by McCall (1996) and ourselves. Putting aside the differences between our findings and McCall (1996), the proposed model compares favourably with the ordinary discrete time competing risks model and it should, therefore, also compare favourably with the multinomial model.

We now illustrate the ease with which the cumulative incidence function estimates can be computed via the proposed model. Suppose we wished to investigate the effect of increasing DR by 50% for an average unemployed subject, i.e., a UI recipient at average values for DR and RR in relation to exit to full-time employment. The expression for the survival function is given by

$$S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j) = \{S_{0j}(t)\}^{\exp(\boldsymbol{x}^T\boldsymbol{\beta}_j)}, \qquad \text{for } t \in [a_{t-1}, a_t),$$

where $S_{0j}(t) = \exp\{-\sum_{s=1}^{t} \Delta_s \exp(\beta_{0js})\}$ is the baseline survival function for cause j at time t. Their estimates are listed in Table 1 for convenience. For illustrative purposes, the proportion of exits to full-time employment after 5 months ($T = 10$) before the increase in DR can be computed as follows:

$$F_1(10) = \frac{1}{1 + \exp(-1.094)} \times (1 - 0.613) = 0.289$$

to

$$F_1(10) = \left(\frac{1}{1 + \exp\left(-(1.094 + 0.054 \times (-2.941))\right)}\right) \times (1 - 0.613^{\exp(0.054 \times -0.595)}) = 0.271$$

after the increase in DR. The proportion of exits to part-time employment increases from

$$F_1(10) = \left(1 - \frac{1}{1 + \exp\left(-1.094\right)}\right) \times (1 - 0.643)) = 0.089$$

to

$$F_1(10) = \left(1 - \frac{1}{1 + \exp\left(-(1.094 + 0.054 \times (-2.94))\right)}\right) \times (1 - 0.643^{\exp(0.054 \times -1.809)}) = 0.093.$$

4. Conclusion

We have proposed a discrete mixture competing risks model by adapting its continuous time equivalent into a discrete time model. In continuous time, the conditional hazards are assumed to follow the proportional hazards assumption with piece-wise constant conditional baseline hazards. This assumption allows for the conditional hazard parameters to be estimated within a GLM framework through a certain Poisson regression model. This fact, together with the structure of the regression model for the cumulative incidence function that emerges from the mixture model, allows for an alternate means to assess the covariate effects on the cumulative incidence function. We have shown that, with a slight modification, the proposed model can be applied to discrete competing risks data and thus serve as an alternative to the regression model for the cumulative incidence function that arises under the multinomial model.

Appendix

We rely on the multivariate delta method to determine the variance of the cumulative incidence estimates, which is given by

$$V(\hat{F}_j(t|\boldsymbol{x};\boldsymbol{\theta})) = \left(\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \gamma}, \frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_j}\right)^T \begin{bmatrix} \mathrm{V}(\gamma) & \mathbf{0} \\ \mathbf{0} & \mathrm{V}(\boldsymbol{\beta}_j) \end{bmatrix} \left(\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \gamma}, \frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_j}\right)\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}.$$

The partial derivatives are given as

$$\begin{aligned}
\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \gamma_{1kb}} &= -\pi_j(\boldsymbol{x};\gamma)\pi_k(\boldsymbol{x};\gamma)Q_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)x_b, \\
\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \gamma_{1ja}} &= \pi_j(\boldsymbol{x};\gamma)(1 - \pi_j(\boldsymbol{x};\gamma))Q_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)x_a, \\
\frac{\partial F_j(t|\boldsymbol{x};\boldsymbol{\theta})}{\partial \beta_{0js}} &= \pi_j(\boldsymbol{x};\gamma)S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)\Delta_s\lambda_j(t|\boldsymbol{x};\boldsymbol{\beta}_j), \\
\frac{\partial F_j(t)}{\partial \boldsymbol{\beta}_{1j}} &= \pi_j(\boldsymbol{x};\gamma)S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)\sum_{s=1}^{t}\Delta_s\lambda_j(t|\boldsymbol{x};\boldsymbol{\beta}_j),
\end{aligned}$$

where $Q_j(t|\boldsymbol{x};\boldsymbol{\beta}_j) = 1 - S_j(t|\boldsymbol{x};\boldsymbol{\beta}_j)$. The covariance matrices produced by statistical packages are based on the complete data i.e $\mathrm{V}(\boldsymbol{\gamma})$ and $\mathrm{V}(\boldsymbol{\beta}_j)$ are complete data covariance matrices. These must be adjusted to reflect observed data covariances. Towards that end, the methods by Oakes (1999) and Louis (1982) are some of the approaches to adjust the complete data covariance matrices. We apply the Oakes (1999) approach and it is given by

$$\mathcal{I}_y(\boldsymbol{\theta}^0) = -\frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} + \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{0T}},$$

where $\boldsymbol{\theta}^0$ is MLE of $\boldsymbol{\theta}$ at the convergence of EM algorithm. The first term of the above equation is the complete data information and second term is the missing information that we have to compute. We begin with $\boldsymbol{\beta}_j$, the vector of cause j failure time parameters, and assume that the subjects have been re-indexed so that the first k subjects are uncensored and the remaining $n-k$ are censored. For convenience, let $S_{ij} = S_j(s|\boldsymbol{x};\boldsymbol{\beta}_j)$, $\pi_{ij} = \pi_j(\boldsymbol{x}_i,\boldsymbol{\gamma})$, and $\lambda_{ijs} = \lambda_j(s|\boldsymbol{x};\boldsymbol{\beta}_j)$. The complete data Q score functions can be written as

$$\begin{aligned}
\frac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{0js}^0} &= \sum_{i=1}^{k} d_{ijs} - \Delta_{is}\lambda_{ijs}^0 + \sum_{i=k+1}^{n} d_{ijs} - \underline{v}_{ij}\Delta_{is}\lambda_{ijs}^0,\\
\frac{\partial \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{1ja}^0} &= \sum_{i=1}^{k}\sum_{s=1}^{t_i}(d_{ijs} - \Delta_{is}\lambda_{ijs}^0)x_{ia} + \sum_{i=k+1}^{n}\sum_{s=1}^{t_i}(d_{ijs} - \underline{v}_{ij}\Delta_{is}\lambda_{ijs}^0)x_{ia}.
\end{aligned}$$

Using the chain rule, the partial derivatives of the pseudo-variable $\underline{v}_{ij}$ are given by

$$\begin{aligned}
\frac{\partial\underline{v}_{ij}}{\partial\beta_{0js}} &= \frac{\partial\underline{v}_{ij}}{\partial S_{ij}}\frac{\partial S_{ij}}{\partial\beta_{0js}} = \frac{\big(\sum_{l=1}^{J}\pi_{ij}S_{ij}\big)\pi_{ij} - \pi_{ij}S_{ij}(\pi_{ij})}{\big(\sum_{j=1}^{J}\pi_{ij}S_{ij}\big)^2}\times -S_{ij}\Delta_{is}\lambda_{ijs} = -\underline{v}_{ij}(1-\underline{v}_{ij})\Delta_{is}\lambda_{ijs},\\
\frac{\partial\underline{v}_{ij}}{\partial\beta_{1ja}} &= \frac{\partial\underline{v}_{ij}}{\partial S_{ij}}\frac{\partial S_{ij}}{\partial\beta_{ija}} = -\frac{\big(\sum_{l=1}^{J}\pi_{il}S_{il}\big)\pi_{ij} - \pi_{ij}S_{ij}\pi_{ij}}{(\sum_{l=1}^{J}\pi_{il}S_{il})^2}S_{ij}x_{ia}\sum_{s=1}^{t_i}\Delta_{is}\lambda_{ijs}\\
&= -\underline{v}_{ij}(1-\underline{v}_{ij})x_{ia}\sum_{s=1}^{t_i}\Delta_{is}\lambda_{ijs}.
\end{aligned}$$

It then follows that

$$\frac{\partial^2\acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{0js}\partial\beta_{0js}^0} = \sum_{i=k+1:s\le t_i}^{n} -\Delta_s\lambda_{ijs}^0\frac{\partial\underline{v}_{ij}}{\partial\beta_{0js}} = \sum_{i=k+1:s\le t_i}^{n}\Delta_{is}\lambda_{ijs}^0\underline{v}_{ij}(1-\underline{v}_{ij})\Delta_{is}\lambda_{ijs},$$

since $\frac{\partial\underline{v}_{ij}}{\partial\beta_{0js}} = 0$ when $s > t_i$. Therefore the missing information corresponding to β_{0js} is

$$\left.\frac{\partial^2\acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{0js}\partial\beta_{0js}^0}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \sum_{i=k+1:s\le t_i}^{n}(\Delta_{is}\lambda_{ijs}^0)^2\underline{v}_{ij}^0(1-\underline{v}_{ij}^0), \quad s = 1,\ldots,q.$$

Regarding the regression coefficients, we have

$$\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{1jb}\partial\beta_{1ja}^0} = \sum_{i=k+1}^{n}\sum_{s=1}^{t_i}(-\Delta_{is}\lambda_{ijs}^0)x_{ia}\frac{\partial \underline{v}_{ij}}{\partial\beta_{1jb}} = \sum_{i=k+1}^{n}\sum_{s=1}^{t_i}(\Delta_{is}\lambda_{ijs}^0)x_{ia}x_{ib}\underline{v}_{ij}(1-\underline{v}_{ij})\sum_{s=1}^{t_i}\Delta_{is}\lambda_{ijs}$$
$$= \sum_{i=k+1}^{n}\underline{v}_{ij}(1-\underline{v}_{ij})x_{ia}x_{ib}\sum_{s=1}^{t_i}\Delta_{is}\lambda_{ijs}^0\sum_{s=1}^{t_i}\Delta_{is}\lambda_{ijs},$$

so that the missing information component regarding $\mathrm{Cov}(\beta_{1ja};\beta_{1jb})$ is

$$\left.\frac{\partial^2 \acute{Q}(\boldsymbol{\beta}_j|\boldsymbol{\beta}_j^0)}{\partial\beta_{1jb}\partial\beta_{1ja}^0}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \sum_{i=k+1}^{n}\underline{v}_{ij}^0(1-\underline{v}_{ij}^0)x_{ia}x_{ib}\left(\sum_{s=1}^{t_i}\Delta_{is}\lambda_{ijs}^0\right)^2.$$

We now focus on $\boldsymbol{\gamma}$. We begin by establishing the following results:

$$\frac{\partial\underline{v}_{ij}}{\partial\gamma_{1ja}} = \frac{\left(\sum_{l=1}^{J-1}\pi_{il}S_{il}+\pi_J S_{iJ}\right)\pi_{ij}(1-\pi_{ij})S_{ij}x_{ia} - \pi_{ij}S_{ij}\left(\pi_{ij}(1-\pi_{ij})S_{ij} - \pi_{ij}\sum_{l\neq j}^{J}\pi_{il}S_{il}\right)x_{ia}}{\left(\sum_{l=1}^{J-1}\pi_{il}S_{il}+\pi_J S_{iJ}\right)^2}$$
$$= \frac{\pi_{ij}S_{ij}\left\{\sum_{l\neq j}^{J}\pi_{il}S_{il}(1-\pi_{ij}) + \pi_{ij}\sum_{l\neq j}^{J}\pi_{il}S_{il}\right\}x_{ia}}{\left(\sum_{l=1}^{J-1}\pi_{il}S_{il}+\pi_J S_{iJ}\right)^2} = \frac{\pi_{ij}S_{ij}\times\sum_{l\neq j}^{J}\pi_{il}S_{il}}{\left(\sum_{l=1}^{J-1}\pi_{il}S_{il}+\pi_J S_{iJ}\right)^2}x_{ia}$$
$$= \underline{v}_{ij}(1-\underline{v}_{ij})x_{ia},$$
$$\frac{\partial\underline{v}_{ij}}{\partial\gamma_{1kb}} = \frac{\left(-\sum_{l=1}^{J-1}\pi_{il}S_{il}+\pi_{iJ}S_{iJ}\right)\pi_{ik}\pi_{ij}S_{ij}x_{ib} - \left(\pi_{ik}(1-\pi_{ik})S_{ik} - \pi_{ik}\sum_{l\neq k}^{J}\pi_{il}S_{il}\right)\pi_{ij}S_{ij}x_{ib}}{\left(\sum_{l=1}^{J-1}\pi_{il}S_{il}+\pi_{iJ}S_{iJ}\right)^2}$$
$$= \frac{\pi_{ij}\pi_{ik}S_{ij}x_{ib}\left\{-\sum_{l=1}^{j}\pi_{il}S_{il} - S_{ik} + \sum_{l=1}^{J}\pi_{il}S_{il}\right\}}{\left(\sum_{l=1}^{J-1}\pi_{il}S_{il}+\pi_{iJ}S_{iJ}\right)^2} = -\frac{\pi_{ij}S_{ij}}{\sum_{l=1}^{J}\pi_{il}S_{il}}\frac{\pi_{ik}S_{ik}}{\sum_{l=1}^{J}\pi_{il}S_{il}}x_{ib}$$
$$= -\underline{v}_{ij}\underline{v}_{ik}x_{ib}.$$

We again re-index the subjects as before, and write the "complete" data score function as follows

$$\frac{\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1ja}} = \sum_{i=1}^{k}(d_{ij}-\pi_{ij}^0)x_{ia} + \sum_{i=k+1}^{n}(\underline{v}_{ij}-\pi_{ij}^0)x_{ia}.$$

Then,

$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1ja}\partial\gamma_{1ja}^0} = \sum_{i=k+1}^{n}\frac{\partial\underline{v}_{ij}}{\partial\gamma_{1ja}}x_{ia} = \sum_{i=k+1}^{n}\underline{v}_{ij}(1-\underline{v}_{ij})x_{ia}^2,$$
$$\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1kb}\partial\gamma_{1ja}^0} = \sum_{i=k+1}^{n}\frac{\partial\underline{v}_{ij}}{\partial\gamma_{1jb}}x_{ia} = -\sum_{i=k+1}^{n}\underline{v}_{ij}\underline{v}_{ik}x_{ia}x_{ib}.$$

Thus, the missing information components with respect to $\mathrm{Var}(\gamma_{1ja})$ and $\mathrm{Cov}(\gamma_{1kb};\gamma_{1ja})$ are given by

$$\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial^2\gamma_{1ja}\partial\gamma_{1ja}^0}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \sum_{i=k+1}^{n}\underline{v}_{ij}^0(1-\underline{v}_{ij}^0)x_{ia}^2,$$
$$\left.\frac{\partial^2 Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^0)}{\partial\gamma_{1kb}\partial\gamma_{1ja}^0}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = -\sum_{i=k+1}^{n}\underline{v}_{ij}^0\underline{v}_{ik}^0x_{ia}x_{ib}.$$

References

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, **13**, 61–98.

Ambrogi, F., Biganzoli, E., and Boracchi, P. (2009). Estimating crude cumulative incidences through multinomial logit regression on discrete cause-specific hazards. *Computational Statistics & Data Analysis*, **53**, 2767–2779.

Becg, C. B. and Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika*, **71**, 11–18.

Choi, K. and Zhou, X. (2002). Large sample properties of mixture models with covariates for competing risks. *Journal of Multivariate Analysis*, **82**, 331–366.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, **34**, 187–202.

Croissant, Y. (2015). Ecdat: Data sets for econometrics. R package version 0.2-9.

Escarela, G. and Bowater, R. J. (2008). Fitting a semi-parametric mixture model for competing risks in survival data. *Communications in Statistics – Theory and Methods*, **37**, 277–293.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496–509.

Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, **16**, 1141–1154.

Holford, T. R. (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics*, **36**, 299–305.

Klein, J. P. and Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, **61**, 223–229.

Kuk, A. Y. C. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79**, 531–541.

Laird, N. and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, **76**, 231–240.

Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society: Series C*, **34**, 201–211.

Lau, B., Cole, S. R., and Gange, S. J. (2011). Parametric mixture models to evaluate and summarize hazard ratios in the presence of competing risks with time-dependent hazards and delayed entry. *Statistics in Medicine*, **30**, 654–665.

Lau, B., Cole, S. R., Moore, R. D., and Gange, S. J. (2008). Evaluating competing adverse and beneficial outcomes using a mixture model. *Statistics in Medicine*, **27**, 4313–4327.

Lee, M., Feuer, E. J., and Fine, J. P. (2018). On the analysis of discrete time competing risks data. *Biometrics*, **74**, 1468–1481.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B*, **44**, 226–233.

Maller, R. A. and Zhou, X. (2002). Analysis of parametric models for competing risks. *Statistica Sinica*, **12**, 725–750.

McCall, B. P. (1996). Unemployment insurance rules, joblessness, and part-time work. *Econometrica*, **64**, 647–682.

Ng, S., McLachlan, G., Yau, K. W., and Lee, A. (2004). Modelling the distribution of ischaemic stroke-specific survival time using an EM-based mixture approach with random effects adjustment. *Statistics in Medicine*, **23**, 2729–2744.

Ng, S. K. and McLachan, G. J. (2003). An EM-based semi-parametric mixture model approach to the regression analysis of competing-risks data. *Statistics in Medicine*, **22**, 1097–1111.

Oakes, D. (1999). Direct calculation of the information matrix via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, **61**, 479–482.

Peng, Y. (2003). Fitting semiparametric cure models. *Computational Statistics & Data Analysis*, **41**, 481–490.

Peng, Y. and Dear, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, **56**, 237–243.

Prentice, R. L. and Gloecker, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, **34**, 57–67.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics,*, **34**, 541–554.

Scheike, T. H., Zhang, M.-J., and Gerds, M. (2008). Predicting cumulative incidence probabilty by direct binomial regression. *Biometrika*, **1**, 205–220.

Singer, J. D. and Willet, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, **18**, 155–195.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazard cure models. *Biometrics*, **56**, 227–236.

Tutz, G. (1995). Competing risks models in discrete time with nominal or ordinal categories of response. *Quality and Quantity*, **29**, 405–420.

Tutz, G. and Schmid, M. (2016). *Modeling Discrete Time-to-Event Data*. Springer, Cham, Switzerland.

Vermunt, J. K. (1997). *Log-Linear Models for Event Histories*. Sage Publications, Thousand Oaks, CA, USA.

Welchowski, T. and Schmid, M. (2019). *discSurv: Discrete Time Survival Analysis*. R package version 1.4.0.

Zhiping, T. (2011). *On Competing Risks Data with Covariates and Long-Term Survivors*. Ph.D. thesis, The Hong Kong Polytechnic University.

Manuscript received 2018-03-31, revised 2019-07-11, accepted 2019-07-29.