STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:1097–1111 (DOI: 10.1002/sim.1371)

An EM-based semi-parametric mixture model approach to the regression analysis of competing-risks data

S. K. Ng*,† and G. J. McLachlan

Department of Mathematics, University of Queensland, Brisbane, Q4072, Australia

SUMMARY

We consider a mixture model approach to the regression analysis of competing-risks data. Attention is focused on inference concerning the effects of factors on both the probability of occurrence and the hazard rate conditional on each of the failure types. These two quantities are specified in the mixture model using the logistic model and the proportional hazards model, respectively. We propose a semi-parametric mixture method to estimate the logistic and regression coefficients jointly, whereby the component-baseline hazard functions are completely unspecified. Estimation is based on maximum likelihood on the basis of the full likelihood, implemented via an expectation-conditional maximization (ECM) algorithm. Simulation studies are performed to compare the performance of the proposed semi-parametric method with a fully parametric mixture approach. The results show that when the component-baseline hazard is monotonic increasing, the semi-parametric and fully parametric mixture approaches are comparable for mildly and moderately censored samples. When the component-baseline hazard is not monotonic increasing, the semi-parametric method consistently provides less biased estimates than a fully parametric approach and is comparable in efficiency in the estimation of the parameters for all levels of censoring. The methods are illustrated using a real data set of prostate cancer patients treated with different dosages of the drug diethylstilbestrol. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: competing risks; ECM algorithm; mixture models; prostate cancer data; regression analysis; semi-parametric approach

1. INTRODUCTION

Competing-risks problems arise naturally in a number of scientific fields, particularly in survival analysis. With covariates, the traditional approach based on statistical models for the observable failure/censoring time T is to represent the competing-risk failure rates by the cause-specific hazard functions via Cox's [1] proportional hazards assumption [2]; see also reference [3], Chapter 7, and reference [4]. Some examples of applying this approach can be found in references [5–7]. Recently, Lunn and McNeil [8] proposed an augmented data approach to analyse competing-risks data using readily available standard programs for fitting

* Correspondence to: Angus S. K. Ng, Department of Mathematics, University of Queensland, Brisbane, Q4072, Australia.
† E-mail: skn@maths.uq.edu.au

Received September 2001
Accepted June 2002

Cox's proportional hazards regression model with censored observations. A comparison of this augmented data approach with Kaplan–Meier methods and the cause-specific hazard approach to estimate the cumulative incidence functions in the competing-risks analysis can be found in reference [9].

An alternative analysis of competing-risk data postulates a mixture model that expresses the failure time distribution in terms of the marginal distribution of failure type and the conditional distribution of time to failure, given the type of failure. Within this mixture model framework, it is assumed that an individual will fail from a particular risk, chosen by a stochastic mechanism at the outset, characterized by the marginal distribution of each failure type. Suppose that there are g distinct causes of failure and the observed failure-time data is

$$\boldsymbol{y} = (t_1, \boldsymbol{x}_1^{\mathrm{T}}, D_1, \ldots, t_n, \boldsymbol{x}_n^{\mathrm{T}}, D_n)^{\mathrm{T}} \tag{1}$$

where the superscript T denotes vector transpose, t_j is the failure time or censoring time for the jth individual, $\boldsymbol{x}_j$ is a vector of covariates associated with the jth individual, and $D_j = i$ indicates that the jth individual fails due to the ith type of failure and $D_j = 0$ represents a censored observation. The survival function of T is modelled as

$$S(t; \boldsymbol{x}) = \sum_{i=1}^{g} \pi_i(\boldsymbol{x}) S_i(t; \boldsymbol{x}) \tag{2}$$

where $S_i(t; \boldsymbol{x})$ denotes the conditional survival function given failure is due to the ith cause, and $\pi_i(\boldsymbol{x})$ $(i = 1, \ldots, g)$ is the probability of failure from the ith cause; the $\pi_i(\boldsymbol{x})$ sum to one. In the context of competing-risk analysis, it means that each individual will fail from one of the g failure types, while the relative risk is characterized by the marginal distribution $\pi_i(\boldsymbol{x})$ given the characteristics of the individual $\boldsymbol{x}$. The model (2) thus allows the possibility of some useful interpretations on how the factors $\boldsymbol{x}$ influence the incidence of each cause and how they affect the failure time among individuals who failed from each cause. Larson and Dinse [10] were among the first to use model (2) to handle competing-risks problems. They assumed that the component-hazard functions, $h_i(t; \boldsymbol{x})$ $(i = 1, \ldots, g)$, follow a proportional hazards model, that is

$$h_i(t; \boldsymbol{x}) = h_{0i}(t) \exp\{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{\gamma}_i\} \quad (i = 1, \ldots, g) \tag{3}$$

where $\boldsymbol{\gamma}_i$ is a vector of regression coefficients and where the baseline hazard functions, $h_{0i}(t)$, are taken to be piecewise constant. An alternative to the specification of the baseline hazard functions is to adopt some common lifetime distributions for them. For example, Gordon [11] adopted the Gompertz distribution to specify the conditional survival functions in the context of estimating the 'cure' rate of breast cancer after a treatment therapy. However, in practice, it is difficult to verify the distributional assumptions adopted in the mixture models and the inference can be very sensitive to the choice of distributions. Common lifetime distributions may fail to adequately interpret data for which the failure rate is not monotonic increasing or decreasing [12]. One typical example in medical research is the existence of short-term failures. The occurrence of failures soon after the treatment indicates that a non-monotone bathtub-shaped hazard function with three phases (a decreasing hazard, following by a constant hazard, and finally by an increasing hazard) can possibly provide a more realistic model than the monotone hazard function; see for example reference [13] and the review [14]. Glasser [15] showed that a gamma mixture with a common scale parameter can have bathtub-shaped hazard function. For the analysis of competing-risks data, it implies that a larger number of components is required which results in a much more complicated model and hence may induce difficulties in the estimation. Alternatively, a flexible rich class of baseline hazards can be adopted. For example, Larson and Dinse [10] specified the baseline hazard function to be piecewise constant. Gelfand *et al.* [12] proposed a continuous baseline hazard in the form of the summation of an arbitrary number of parametric hazards such as the Weibull. However, the division of the baseline hazards into intervals or the determination of the number of parametric hazards for each component is somewhat arbitrary.

A non-parametric specification for the baseline hazard functions may be used to relax the parametric constraints. In particular, Efron [16] and Oakes [17] showed that, for the estimation of the regression coefficients, parametric specification of the baseline hazard functions will usually not improve much on the partial likelihood model, with the baseline hazard completely unspecified. Kuk [18] considered a semi-parametric generalization of the parametric mixture model of reference [10]. A marginal likelihood approach is adopted to estimate the regression parameters γ_i, whereby the baseline hazard functions $h_{0i}(t)$ in (3) are eliminated as nuisance parameters during the analysis. His method relies on the Monte Carlo approximation of the marginal likelihood [19] and the validity of assigning equal probability to each realization of the failure type for the set of censored observations.

In this paper, we propose an ECM-based semi-parametric mixture method that does not require Monte Carlo approximation. Moreover, estimation is based on maximum likelihood of the full likelihood. In Section 2 we present the semi-parametric mixture model, where parameters can be estimated by an extension of the EM algorithm that Meng and Rubin [20] termed the expectation-conditional maximization (ECM) algorithm. Some simulations performed to compare the proposed semi-parametric method with a fully parametric approach are reported in Section 3, and the analysis of a real data set is given in Section 4.

2. SEMI-PARAMETRIC MIXTURE MODEL AND ECM ALGORITHM

In the mixture model framework, a population may be split into g mutually exclusive components corresponding to each type of failure. The mixing proportions are assumed to have the logistic form

$$\pi_i(\boldsymbol{x};\boldsymbol{\alpha}) = \exp(a_i + \boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{x}) \bigg/ \left(1 + \sum_{l=1}^{g-1} \exp(a_l + \boldsymbol{b}_l^{\mathrm{T}}\boldsymbol{x})\right) \quad (i = 1,\ldots,g-1) \tag{4}$$

where $\boldsymbol{\alpha}_i = (a_i, \boldsymbol{b}_i^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^{\mathrm{T}},\ldots,\boldsymbol{\alpha}_{g-1}^{\mathrm{T}})^{\mathrm{T}}$ contains the logistic coefficients [21], and $\pi_g(\boldsymbol{x};\boldsymbol{\alpha}) = 1 - \sum_{i=1}^{g-1} \pi_i(\boldsymbol{x};\boldsymbol{\alpha})$. Let $\boldsymbol{\Psi} = (\boldsymbol{\alpha}^{\mathrm{T}}, \gamma_1^{\mathrm{T}},\ldots,\gamma_g^{\mathrm{T}})^{\mathrm{T}}$ be the vector containing the unknown parameters for the logistic and regression coefficients. On the basis of the observed data $\boldsymbol{y}$ given by (1), the log-likelihood function for $\boldsymbol{\Psi}$ under the mixture model (2) is given by

$$\log L(\boldsymbol{\Psi}) = \sum_{j=1}^{n} \left[\sum_{i=1}^{g} I(D_j = i) \log\{\pi_i(\boldsymbol{x}_j;\boldsymbol{\alpha}) f_i(t_j;\boldsymbol{x}_j,\gamma_i)\} \right. \\ \left. + I(D_j = 0) \log S(t_j;\boldsymbol{x}_j,\boldsymbol{\Psi}) \right] \tag{5}$$

where $I(A)$ is the indicator function for event A and $f_i(.)$ is the probability density function for the ith component. It is noted that the full likelihood (5) is used here because a partial likelihood approach as described in reference [3], pp. 170, fails to eliminate the baseline survival function (see equation (10)) from the likelihood formed under the mixture model (2). The maximum likelihood estimate of $\boldsymbol{\Psi}$ is obtained via the EM algorithm of Dempster *et al.* [22]. Further discussion of the EM algorithm in its application to mixture models in a general context may be found in references [23, 24].

In order to pose the problem as an incomplete-data one, an unobservable random vector $\boldsymbol{z}$ of zero-one indicator variables is introduced for each censored observation t_j, where $\boldsymbol{z}_j=(z_{1j},\ldots,z_{gj})^{\mathrm{T}}$, and where $z_{ij}=1$ or 0 according as the jth individual would have failed from cause i or not $(i=1,\ldots,g)$. The actual failure time for those censored observations was not introduced as an incomplete variable in the complete-data framework, as it did not simplify the calculations.

The complete-data log-likelihood is then given by

$$\log L_{\mathrm{c}}(\boldsymbol{\Psi})=\sum_{j=1}^{n}\left[\sum_{i=1}^{g} I(D_j=i)\log\{\pi_i(\boldsymbol{x}_j;\boldsymbol{\alpha})f_i(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_i)\}\right. \\ \left. +\sum_{i=1}^{g} I(D_j=0)z_{ij}\log \pi_i(\boldsymbol{x}_j;\boldsymbol{\alpha})S_i(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_i)\right] \tag{6}$$

It follows on application of the EM algorithm in the aforementioned framework that on the $(k+1)$th iteration of the E-step, we calculate the Q-function, which is the expectation of the complete-data log-likelihood conditional on the current estimate of the parameter and the observed data

$$Q(\boldsymbol{\Psi};\boldsymbol{\Psi}^{(k)})=\sum_{j=1}^{n}\left[\sum_{i=1}^{g} I(D_j=i)\log\{\pi_i(\boldsymbol{x}_j;\boldsymbol{\alpha})f_i(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_i)\}\right. \\ \left. +\sum_{i=1}^{g} I(D_j=0)\tau_{ij}^{(k)}\log \pi_i(\boldsymbol{x}_j;\boldsymbol{\alpha})S_i(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_i)\right] \tag{7}$$

where $\boldsymbol{\Psi}^{(k)}$ is the estimate of $\boldsymbol{\Psi}$ after the kth iteration and

$$\tau_{ij}^{(k)}=E(z_{ij}|\boldsymbol{y},\boldsymbol{\Psi}^{(k)}) \\ =\pi_i(\boldsymbol{x}_j;\boldsymbol{\alpha}^{(k)})S_i(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_i^{(k)})\Big/\sum_{l=1}^{g}\pi_l(\boldsymbol{x}_j;\boldsymbol{\alpha}^{(k)})S_l(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_l^{(k)}) \tag{8}$$

is the posterior probability that the jth individual with censored survival time t_j would have failed due to cause i $(i=1,\ldots,g)$.

The M-step provides the updated estimate $\boldsymbol{\Psi}^{(k+1)}$ that maximizes the Q-function with respect to $\boldsymbol{\Psi}$. It can be seen that the Q-function in (7) can be decomposed into

$$\begin{aligned}
&\sum_{j=1}^{n}\sum_{i=1}^{g}[I(D_j=i)\log\pi_i(\boldsymbol{x}_j,\boldsymbol{\alpha})+I(D_j=0)\tau_{ij}^{(k)}\log\pi_i(\boldsymbol{x}_j,\boldsymbol{\alpha})]\\
&\quad+\sum_{j=1}^{n}[I(D_j=1)\log f_1(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_1)+I(D_j=0)\tau_{1j}^{(k)}\log S_1(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_1)]\\
&\quad+\qquad\vdots\\
&\quad+\sum_{j=1}^{n}[I(D_j=g)\log f_g(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_g)+I(D_j=0)\tau_{gj}^{(k)}\log S_g(t_j;\boldsymbol{x}_j,\boldsymbol{\gamma}_g)]\\
&\quad=Q_0+Q_1+\cdots+Q_g
\end{aligned}\tag{9}$$

with respect to the unknown parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_g$, respectively. It implies that the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_g$ can be updated separately by maximizing Q_0 and $Q_1,\ldots,Q_g$, respectively. On differentiation of Q_0 with respect to $\boldsymbol{\alpha}_i$ $(i=1,\ldots,g-1)$, it follows that $\boldsymbol{\alpha}_i^{(k+1)}$ satisfies the equation

$$\sum_{j=1}^{n}[I(D_j=i)+I(D_j=0)\tau_{ij}^{(k)}-\pi_i(\boldsymbol{x}_j,\boldsymbol{\alpha})]\boldsymbol{x}_j=\mathbf{0}$$

For the maximization of Q_i with respect to γ_i $(i=1,\ldots,g)$, the proportional hazards assumption (3) is adopted for the component-hazard functions, $h_i(t;\boldsymbol{x})$ $(i=1,\ldots,g)$.

The ith component-survival function is given by

$$S_i(t;\boldsymbol{x})=S_{0i}(t)^{\exp\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\gamma}_i\}}\tag{10}$$

where $S_{0i}(t)$ is the baseline survival function. On letting $H_{0i}(t)$ denote the cumulative hazard function for the ith component $(i=1,\ldots,g)$, we have that

$$S_{0i}(t)=\exp\left\{-\int_0^t h_{0i}(u)\,\mathrm{d}u\right\}=\exp\{-H_{0i}(t)\}$$

From (9), it follows that, for $i=1,\ldots,g$

$$\begin{aligned}
Q_i=\sum_{j=1}^{n}[&-\{I(D_j=i)+I(D_j=0)\tau_{ij}^{(k)}\}H_{0i}(t_j)\exp(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\gamma}_i)\\
&+I(D_j=i)\{\log h_{0i}(t_j)+\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\gamma}_i\}]
\end{aligned}\tag{11}$$

On the M-step, it follows from (11) that we need to maximize Q_i with respect to $\boldsymbol{\gamma}_i$ and the function $H_{0i}(t)$. This maximization is implemented using a conditional approach, and the resulting algorithm can be viewed as an expectation-conditional maximization (ECM) algorithm [20]. With the application of the ECM algorithm here, the M-step is replaced by two conditional maximization (CM) steps. The first involves the calculation of $H_{0i}^{(k+1)}(t)$

by maximization of (11) with γ_i fixed at $\gamma_i^{(k)}$. The second CM step calculates $\gamma_i^{(k+1)}$ by maximization of (11) with $H_{0i}(t)$ fixed at $H_{0i}^{(k+1)}(t)$.

We now rearrange the failure time observations in increasing order and denote the m_i distinct failure times due to the ith cause by $t_{(i1)} < \cdots < t_{(im_i)}$ for $i = 1, \ldots, g$. By assuming a step function for $h_{0i}(t)$ with discontinuities at each observed failure time due to the ith cause and considering censored observations as censored at the preceding uncensored failure time [25, 26], it can be shown that, for fixed γ_i $(i = 1, \ldots, g)$, (11) is maximized with respect to $H_{0i}(t)$ at

$$H_{0i}^{(k+1)}(t_{(im)}) = \sum_{j=1}^{m} \left(\frac{d_{ij}}{\sum_{r \in R(t_{(ij)})} [I(D_r = i) + I(D_r = 0)\tau_{ir}^{(k)}] \exp(\boldsymbol{x}_r^{\mathrm{T}} \boldsymbol{\gamma}_i)} \right) \tag{12}$$

for $m = 1, \ldots, m_i$, where d_{ij} is the number of failures due to cause i at time $t_{(ij)}$ and $R(t_{(ij)})$ is the risk set at time $t_{(ij)}$. From (12), the updated estimates for the baseline survival functions are given by

$$S_{0i}^{(k+1)}(t_{(im)}) = \exp\{-H_{0i}^{(k+1)}(t_{(im)})\} \quad (i = 1, \ldots, g;\ m = 1, \ldots, m_i)$$

which can be substituted into (7) for the implementation of the next E-step. Alternatively, a discrete model assuming a step function for $S_{0i}(t)$ (reference [3], pp. 85) may be adopted to obtain the estimates of the baseline survival functions that maximize (11) for fixed γ_i. However, unlike (12), the equation must be solved iteratively when there are ties in the data. If there are relatively few ties in the data, they may be broken by adding randomly some infinitesimally small values to the tied data.

The solution to the second CM-step, however, does not exist in closed form. On differentiation of the Q-function with respect to $\boldsymbol{\gamma}_i$ for fixed $H_{0i}^{(k+1)}(t)$, it follows that $\boldsymbol{\gamma}_i^{(k+1)}$ satisfies the equation

$$\sum_{j=1}^{n} [I(D_j = i) - \{I(D_j = i) + I(D_j = 0)\tau_{ij}^{(k)}\} H_{0i}^{(k+1)}(t_j) \exp(\boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{\gamma}_i)] \boldsymbol{x}_j = \boldsymbol{0} \tag{13}$$

The ECM algorithm preserves the appealing convergence properties of the EM algorithm. It thus has reliable global convergence in that it monotonely increases the likelihood after each iteration, no matter what starting value is used. As the likelihood function usually has multiple maxima with mixture models, the ECM algorithm should be applied from different initial values to obtain the global maximum, which is usually taken to be the largest of the local maxima obtained. A detailed account of the convergence properties of the EM (ECM) algorithm can be found in references [27, 28]. It is noted that if $\boldsymbol{\gamma}_i^{(k+1)}$ is obtained by simply increasing the Q-function with respect to $\boldsymbol{\gamma}_i$ rather than a global maximization (13), then the algorithm is referred to as a generalized EM (GEM) algorithm [22].

In an application of mixture models for cure rate estimation, where only one type of failure is being observed, Sy and Taylor [29] and Peng and Dear [30] adopted the profile likelihood approach to estimate the regression parameters and the non-parametric baseline survival function. The estimation via the ECM algorithm is found to be more stable compared to the profile likelihood approach.

From (8) and (13), it can be seen that the baseline component-survival functions are not completely eliminated in the estimation process with our proposed method. Because the baseline component-survival functions are estimated on the basis of the current information (12), it means that, in comparison to Kuk's [18] semi-parametric approach, our model allows the non-parametric maximum likelihood estimates of the baseline survival functions to be used in the estimation of the logistic and regression parameters.

The standard errors of estimates of the parameters can be computed by applying the non-parametric bootstrap approach of Efron [31] with the resampling scheme modified for the competing-risks problems. Let n_i $(i=1,\ldots,g)$ be the number of failures due to ith cause, and let n_{g+1} be the number of censored observations. The bootstrap data are obtained by sampling separately from each of the $(g+1)$ sets, corresponding to cause i failures $(i=1,\ldots,g)$ and the censored observations, with the sizes of these bootstrap subsamples taken equal to n_i $(i=1,\ldots,g)$ and n_{g+1}, respectively [32, 33]. The standard errors of the estimates can be approximated by the sample standard deviations of the corresponding bootstrap estimates based on B independent bootstrap samples (say, $B=100$).

With the mixture approach, the cumulative incidence function and the conditional probability [34] can be obtained as follows. Let $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\gamma}}_i$, and $\hat{S}_{0i}(t)$ $(i=1,\ldots,g)$ be the maximum likelihood estimates of $\boldsymbol{\alpha},\boldsymbol{\gamma}_i$ $(i=1,\ldots,g)$, and the baseline survival functions, respectively, the estimated cumulative incidence function for the ith type of failure is given by

$$\pi_i(\boldsymbol{x};\hat{\boldsymbol{\alpha}})\{1-\hat{S}_{0i}(t)^{\exp(\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{\gamma}}_i)}\} \tag{14}$$

Similarly, the conditional probability for the ith type of failure within a specified time t given that failure due to other types does not occur during this period is estimated by

$$\frac{\pi_i(\boldsymbol{x};\hat{\boldsymbol{\alpha}})\{1-\hat{S}_{0i}(t)^{\exp(\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{\gamma}}_i)}\}}{\pi_i(\boldsymbol{x};\hat{\boldsymbol{\alpha}})+\sum_{l\neq i}^{g}\pi_l(\boldsymbol{x};\hat{\boldsymbol{\alpha}})\hat{S}_{0l}(t)^{\exp(\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{\gamma}}_l)}} \tag{15}$$

3. SIMULATION EXPERIMENTS

In this section we present the results of two simulation experiments for comparing the proposed semi-parametric method with a fully parametric approach. In both simulations we considered the sample size $n=1000$ and two distinct causes of failure $(g=2)$. The covariate x was a continuous variable, which was generated independently from the N$(0,1)$ distribution. In the first simulation study, we assume both the component-hazard functions $h_i(t;x)$ $(i=1,2)$ are exponential distributions with proportional hazards

$$h_i(t;x)=\lambda_i\exp(\gamma_i x) \tag{16}$$

The true parameter values were $(\lambda_1,\gamma_1,\lambda_2,\gamma_2)=(0.5,-0.5,1.0,-1.0)$. For the parameters in the logistic model (4), we used $a=-1.0$ and $b=0.5$. Given that an entity belongs to the first component, a sample failure time due to cause 1 was generated according to $h_1(t;x,\lambda_1,\gamma_1)$ using the inverse transform method. Similarly, for an entity belonging to the second component, a sample failure time due to cause 2 was generated according to $h_2(t;x,\lambda_2,\gamma_2)$. For each entity, the censoring time was generated from a uniform distribution U(c_1,c_2), where c_1 and c_2 are some constants. If the jth failure time were greater than the jth censoring time, it was taken to be censored at this censoring time. In the study, we considered three different sets of values for c_1 and c_2 so that comparison under different levels of censoring could be investigated. For each simulation set, we generated 500 independent samples and fitted the

Table I. Average bias, MSE, and the relative efficiency of estimates from the parametric mixture model and the proposed semi-parametric mixture method (simulation study 1).

Censoring distribution	Average per cent censored	Parameter	Parametric method		Semi-parametric method		Relative efficiency
			Average bias	MSE	Average bias	MSE	
U(2.0, 9.0)	9.1	a	−0.0008	0.0065	−0.0030	0.0067	0.98
		b	0.0029	0.0099	0.0011	0.0105	0.95
		γ_1	0.0051	0.0074	0.0049	0.0073	1.02
		γ_2	−0.0019	0.0026	−0.0042	0.0030	0.89
U(0.5, 5.0)	22.8	a	−0.0112	0.0118	−0.0196	0.0128	0.92
		b	−0.0007	0.0147	−0.0064	0.0149	0.98
		γ_1	−0.0010	0.0096	0.0006	0.0101	0.95
		γ_2	−0.0011	0.0032	−0.0038	0.0035	0.91
U(0.5, 1.8)	40.7	a	−0.0377	0.0611	−0.0645	0.0847	0.72
		b	−0.0162	0.0405	−0.0232	0.0469	0.87
		γ_1	0.0153	0.0263	0.0214	0.0289	0.91
		γ_2	−0.0014	0.0069	−0.0050	0.0067	1.03

simulated data using the proposed semi-parametric method. The number of replications was so chosen that the variation due to the Monte Carlo simulations contributes only a small fraction of the total variance of estimates. With the parametric modelling approach, we fitted a mixture of exponential distributions to the simulated data. The average bias, the mean square error (MSE), and the efficiency of the semi-parametric method relative to the fully parametric approach for each parameter are reported in Table I.

From Table I, it can be seen that the fully parametric approach and the proposed semi-parametric method are comparable to each other for mildly and moderately censored samples. For heavily censored samples, the parametric approach provides less biased estimates and is more efficient for the logistic parameters. The semi-parametric approach is generally less efficient, which is to be expected because the true model is a two-component mixture of exponential distributions.

For the second simulation study, the exponential distribution of the second component of mixture is replaced by a distribution with baseline hazard function specified as

$$h_{02}(t) = \lambda_{21}\beta_{21}t^{\beta_{21}-1} + \lambda_{22}\beta_{22}t^{\beta_{22}-1} \tag{17}$$

with $(\lambda_{21}, \beta_{21}, \lambda_{22}, \beta_{22}) = (1.5, 0.5, 0.01, 2.5)$. It is noted that, with these parameter values, the first term on the right-hand side of (17) corresponds to a decreasing Weibull hazard, whereas the second term corresponds to an increasing Weibull hazard. Therefore, the baseline hazard function $h_{02}(t)$ is bathtub-shaped [12, 33]. With the parametric modelling approach, we fitted a mixture of exponential and Weibull distributions to the simulated data. The effect of misspecification of the bathtub-shaped baseline hazard by a Weibull distribution can be obtained. The results are presented in Table II.

From Table II, it can be seen that the proposed semi-parametric approach provides consistently less biased estimates and is comparable in efficiency for all levels of censoring. In particular, for heavily censored samples, the parametric approach gives relatively large biased estimates for a and large MSE for the estimate of γ_2. The coefficient γ_2 corresponds to the

Table II. Average bias, MSE, and the relative efficiency of estimates from the parametric mixture model and the proposed semi-parametric mixture method (simulation study 2).

Censoring distribution	Average per cent censored	Parameter	Parametric method		Semi-parametric method		Relative efficiency
			Average bias	MSE	Average bias	MSE	
U(2.0, 10.0)	9.4	a	−0.0273	0.0065	0.0023	0.0062	1.05
		b	−0.0322	0.0082	0.0065	0.0087	0.95
		γ_1	0.0426	0.0075	−0.0018	0.0062	1.20
		γ_2	0.0042	0.0023	−0.0024	0.0025	0.95
U(0.5, 4.5)	22.9	a	−0.0484	0.0161	−0.0108	0.0171	0.94
		b	−0.0361	0.0167	0.0012	0.0180	0.93
		γ_1	0.0275	0.0119	−0.0031	0.0119	1.00
		γ_2	−0.0096	0.0038	−0.0011	0.0041	0.95
U(0.5, 1.0)	40.4	a	−0.1748	0.2166	−0.1002	0.1438	1.51
		b	−0.0657	0.0665	−0.0514	0.0673	0.99
		γ_1	0.0520	0.0441	0.0515	0.0482	0.92
		γ_2	0.0838	0.1183	−0.0065	0.0078	15.2

second component of the mixture, which baseline hazard function $h_{02}(t)$ is not monotonic increasing.

4. ANALYSIS OF PROSTATE CANCER DATA

As an illustration of the proposed semi-parametric mixture method, we consider the survival times of 506 patients with prostate cancer who entered a clinical trial during 1967–1969 and who were randomly allocated to different levels of treatment with the drug diethylstilbestrol (DES). These data were considered by Byar and Green [35] and are published in reference [36]. Kay [37] analysed a subset of these data by considering eight risk factors, defined by eight categorical variables: drug treatment (RX: 0, 0.0 or 0.2 mg; 1, 1.0 or 5.0 mg); age group (AG: 0, <75 years; 1, 75 to 79 years; 2, $\geqslant 80$ years); weight index (WT: 0, $\geqslant 100$; 1, 80–99; 2, <80); performance rating (PF: 0, normal; 1, limitation of activity); history of cardiovascular disease (HX: 0, no; 1, yes); serum haemoglobin (HG: 0, $\geqslant 12$ g/100 ml; 1, 9–12 g/100 ml; 2, <9 g/100 ml); size of primary lesion (SZ: 0, $<30\,\text{cm}^2$; 1, $\geqslant 30\,\text{cm}^2$), and Gleason stage/grade category (SG: $0, \leqslant 10; 1, >10$). There were 483 patients with complete information on the covariates. Kay [37] considered three types of failure: (i) death due to cancer; (ii) death due to cardiovascular (CVD) disease; (iii) death due to other reasons. He fitted separate cause-specific Cox proportional hazards models to each type of failure. On the other hand, Cheng *et al.* [38] reanalysed the same data subset, but they classified the three causes of death as: (i) prostate cancer; (ii) CVD; (iii) other causes, as in reference [35].

We analysed the same data subset with the latter classification of three causes of death for comparison purposes. In addition, following the analyses of Kay [37] and Cheng *et al.* [38], the three levels of factors AG, WT and HG are treated as a linear contribution according to being 0, 1 or 2, though this is not an usual practice in regression analysis. There were 125 patients who died from prostate cancer, 139 patients who died due to CVD disease, and 80 patients who

Table III. Maximum likelihood estimates (with standard errors) for ECM-based semi-parametric mixture method (* denotes P-value < 0.05).

Coefficient	Logistic model		Components		
	Prostate cancer	CVD	Prostate cancer	CVD	Other
Constant	-1.32^*(0.58)	-0.77 (0.60)			
RX	0.03 (0.48)	0.88^*(0.45)	-0.60^*(0.29)	-0.17 (0.34)	0.23 (0.43)
AG	-0.54 (0.33)	-0.32 (0.34)	0.10 (0.30)	0.43^*(0.22)	0.48 (0.31)
WT	-0.39 (0.32)	-0.35 (0.36)	0.24 (0.28)	-0.01 (0.31)	0.53 (0.34)
PF	0.53 (0.64)	0.29 (0.59)	0.20 (0.35)	0.52 (0.35)	0.69 (0.81)
HX	0.45 (0.41)	1.50^*(0.50)	-0.01 (0.34)	0.48 (0.46)	1.00 (0.57)
HG	0.13 (0.45)	0.01 (0.45)	0.51^*(0.26)	-0.16 (0.47)	0.97^*(0.50)
SZ	0.41 (0.57)	-1.12 (0.62)	0.61^*(0.28)	1.18^*(0.53)	0.94 (0.55)
SG	3.01^*(0.63)	1.45^*(0.45)	0.88^*(0.45)	-0.27 (0.35)	1.18^*(0.56)

Table IV. Maximum likelihood estimates (with standard errors) for three-component Weibull mixture model (* denotes P-value < 0.05).

Coefficient	Logistic model		Components		
	Prostate cancer	CVD	Prostate cancer	CVD	Other
Constant	-1.22^*(0.52)	-0.57 (0.48)			
RX	-0.05 (0.45)	0.86^*(0.42)	-0.43 (0.24)	-0.17 (0.16)	0.09 (0.30)
AG	-0.57 (0.32)	-0.37 (0.33)	0.08 (0.22)	0.33^*(0.16)	0.29 (0.21)
WT	-0.40 (0.34)	-0.45 (0.33)	0.15 (0.17)	0.04 (0.17)	0.34 (0.21)
PF	0.58 (0.66)	0.43 (0.60)	0.22 (0.31)	0.28 (0.22)	0.80 (0.54)
HX	0.35 (0.43)	1.44^*(0.39)	0.04 (0.31)	0.44 (0.26)	0.69 (0.35)
HG	0.13 (0.38)	0.01 (0.39)	0.35 (0.20)	-0.08 (0.24)	0.81^*(0.35)
SZ	0.45 (0.61)	-0.95 (0.67)	0.56^*(0.20)	0.76^*(0.31)	0.68 (0.44)
SG	2.90^*(0.50)	1.07^*(0.45)	0.65 (0.37)	-0.01 (0.23)	0.61 (0.38)

died from other causes. The remaining 139 survival times were all censored; the proportion of censored observation is 28.8 per cent. The proposed semi-parametric three-component mixture approach is adopted and the result is presented in Table III. For comparison, we also fitted a parametric Weibull mixture model and Kuk's semi-parametric model [18]. The results are presented in Tables IV and V, respectively. Standard errors of the maximum likelihood estimates are obtained by the non-parametric bootstrap approach with $B = 100$ replications, as described in Section 2. For Kuk's semi-parametric approach, the Monte Carlo approximation of the marginal likelihood is based on $r = 1000$ replications. Standard errors of the estimates are obtained by inverting the matrix of second derivatives of the marginal log-likelihood based on $r = 100\,000$ replications.

Based on the cause-specific hazard approach, Cheng *et al.* [38] found that treatment with high dose DES significantly reduced the risk of prostate cancer while increasing the risk of CVD. In addition, the SZ and SG variables are highly significant for the death time due to prostate cancer. Besides RX, other significant risk factors for CVD are AG and HX, while AG, WT, HG and SG are all related to the failure from other causes. However, a drawback of the cause-specific hazard approach is that the competing causes of failure are not jointly estimated; that is, a separate model is fitted for each failure cause, treating other fail-

Table V. Semi-parametric estimates (with standard errors) based on Kuk's mixture approach (* denoted P-value < 0.05).

Coefficient	Logistic model		Components		
	Prostate cancer	CVD	Prostate cancer	CVD	Other
Constant	−0.54 (0.46)	−0.38 (0.41)			
RX	−0.05 (0.29)	0.57*(0.23)	−0.62*(0.20)	0.03 (0.15)	−0.01 (0.23)
AG	−0.51 (0.27)	−0.25 (0.19)	0.10 (0.18)	0.38 (0.21)	0.53*(0.18)
WT	−0.34 (0.26)	−0.27 (0.25)	0.18 (0.16)	−0.08 (0.17)	0.57*(0.21)
PF	0.59 (0.52)	0.51 (0.50)	0.27 (0.30)	0.38 (0.29)	1.10*(0.53)
HX	0.11 (0.31)	1.09*(0.30)	−0.11 (0.25)	0.73*(0.22)	0.57*(0.27)
HG	0.05 (0.30)	−0.15 (0.29)	0.48*(0.22)	−0.03 (0.21)	0.85*(0.30)
SZ	0.41 (0.51)	−0.87 (0.57)	0.72*(0.24)	0.85*(0.42)	1.04 (0.54)
SG	2.16*(0.31)	0.89*(0.31)	1.42*(0.30)	−0.07 (0.21)	0.48 (0.47)

ure causes as censored. Subsequently, estimation of the unconditional (marginal) probability $\pi_i(\boldsymbol{x})$ or cumulative incidence function can only be accomplished by combining estimates of each failure cause. A factor that has strong influence on the cause-specific hazard function may have no effect on the unconditional probability or cumulative incidence function; see for example reference [39] on the analysis of prostate cancer data. Thus, a direct comparison of parameter estimates corresponding to the various failure types is complicated under the cause-specific hazard approach [8, 9, 39, 40]. Recently, Lunn and McNeil [8] proposed two methods for joint estimation of parameter in models for competing risks in survival analysis, based on an adaptation of Cox's proportional hazards regression model [1] and the independent risks assumption. Advantages of this approach are that it exploits the use of statistical tests with existing statistical software and provides hazard ratios comparing competing events between covariate values. On the other hand, with the mixture approach, we simultaneously estimate the logistic coefficients $\boldsymbol{\alpha}$ and the regression coefficients γ_i $(i = 1, \ldots, g)$. Thus, useful interpretations on how the risk factors influence the incidence of each cause of death and how they affect the time to death among patients dying from each cause are obtained.

With the analysis of the prostate cancer data using the proposed ECM-based semi-parametric method, it follows from Table III that higher DES dosage not only reduces the probability of death due to prostate cancer, but also prolongs the time to death given that death is due to prostate cancer. On the other hand, higher DES dosage increases the probability of death due to CVD, but DES dosage does not have a significant effect on time to death due to CVD. Patients with a history of cardiovascular disease $(\text{HX} = 1)$ have a higher probability of death due to CVD, compared to those patients without such a history. Patients with high-grade tumours $(\text{SZ} = 1$ and $\text{SG} = 1)$ have higher probability of death due to prostate cancer and also shorter time to death given that death is due to prostate cancer. These results are consistent with the observations made by Byar and Green [35] that patients with high-grade tumours were at greater risk of prostate cancer death, whereas patients with a history of cardiovascular disease, with low-grade tumours, and treated with a high dose of DES were at greater risk of dying from CVD. In Figure 1 we illustrate the effect of DES dosage on the cumulative incidence function of death due to prostate cancer for patients with high-grade tumours $(\text{SZ} = 1$, $\text{SG} = 1$, and the other variables set to zero). For young $(\text{AG} = 0)$ patients with high-grade tumours but no history of CVD, it can be seen from Figure 1 that the cumulative incidence of death

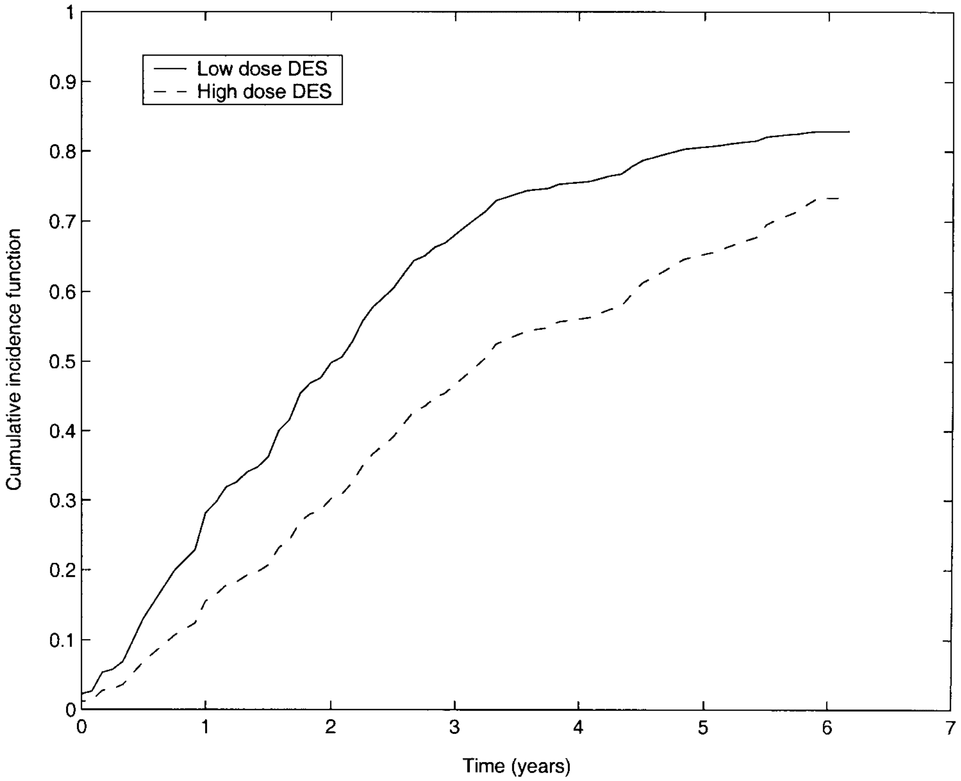

Figure 1. Cumulative incidence function of death due to prostate cancer for young $(\mathrm{AG}=0)$ patients with weight index $\geqslant 100$, normal PF, no history of CVD, haemoglobin $\geqslant 12\,\mathrm{g}/100\,\mathrm{ml}$, and high-grade tumours $(\mathrm{SZ}=1, \mathrm{SG}=1)$.

due to prostate cancer tends to increase much more rapidly in the low-dose DES group than that in the high-dose group. It can be seen from Table III that, besides RX, SZ and SG, the HG variable is also significant for the conditional hazard rate of prostate cancer death. Other significant risk factors for death due to CVD, besides RX and HX, are AG, SZ and SG.

With the fully parametric approach, it can be seen from Table IV that higher DES dosage reduces the probability of death due to prostate cancer, but DES dosage has only a marginally significant effect on time to death due to prostate cancer. In contrast to the result obtained by the semi-parametric approach (Table III), it can be seen that HG is not an important factor on the time to death due to prostate cancer. In addition, with the fully parametric approach, a larger value of SG increases the probability of prostate cancer death, but SG does not have a significant effect on time to death due to prostate cancer. Its effect on the time to death due to other causes is also not significant, which is different from the result obtained by the semi-parametric approach.

From Tables III and V, it can be seen that the results obtained by Kuk's [18] and our semi-parametric approaches lead to similar conclusions on the effect of DES dosage on death due to prostate cancer and CVD. However, with Kuk's approach [18], age of patients has only a marginally significant effect on time to death given death is due to CVD. Moreover, HX not only increases the probability of death due to CVD, but also lowers the time to death due to CVD. Factors that have a significant effect on time to death due to other causes are also not the same.

5. DISCUSSION

We have proposed an ECM-based semi-parametric mixture method for the regression analysis of competing-risks data. In contrast to Kuk's approach [18], the proposed method does not require Monte Carlo approximation. Estimation is undertaken by maximum likelihood via the ECM algorithm, and the process allows the non-parametric maximum likelihood estimates of the baseline survival functions to be used in the estimation of the parameters. As described in Section 2, the proposed estimation procedure via the ECM algorithm is more stable than with the profile likelihood approach and the likelihood is monotonic increasing after each iteration. We also performed some simulation studies using a profile likelihood approach instead of the ECM algorithm. It was found that the MSEs of the estimates obtained by the former approach were relatively larger. Similarly, when it was applied to the real prostate cancer data, convergence was not as stable as with the ECM algorithm, and it took a longer time to converge.

Comparison of the ECM-based semi-parametric mixture method with a fully parametric mixture approach is presented in Section 3. In summary, when the true model is an exponential mixture, the fully parametric approach and the semi-parametric method are comparable under mild and moderate censoring. When one of the components has a bathtub-shaped hazard, the semi-parametric approach consistently provides less biased estimates and is comparable in efficiency in the estimation of the parameters for all levels of censoring. In particular, for heavily censored samples, the parametric approach gives relatively large MSE for the estimate of the regression parameter γ_2 which corresponds to the component with bathtub-shaped baseline hazard function.

An attractive feature of the mixture model approach for analysing competing risks data is that it does not have to make assumptions about the independence of the competing risks [33]. In addition, there is interest in practice as to how factors influence the probability of occurrence and how they relate to the failure rates of each type separately. A factor that is important for the probability of occurrence may not be important for the failure risk and vice versa. The mixture model (2) considers the influence of factors on both the probability of occurrence and the hazard rate conditional on each of the failure types using the logistic model and the proportional hazards model, respectively. In particular, the probability of occurrence for the ith cause is estimated based on the information on the uncensored observations and the posterior probabilities (8) of failure for the censored observations. The mixture model (2) allows us to determine the effect of factors on these two quantities simultaneously. However, a parametric or semi-parametric mixture model for competing-risks data should not be used indiscriminately. Such a model generally requires long-term follow-up and large samples. Otherwise, identifiability problems between the coefficient parameters in the logistic part and the component parts may occur. A simple and informative way of checking the proportional hazards assumption in the semi-parametric mixture model is provided by plotting $\log(-\log(1-\hat{F}_i(t)/\hat{\pi}_i))$ versus time for each level of the variable, where $\hat{F}_i(t)$ is the estimated cumulative incidence function, such as the Aalen–Johansen estimator [41], for the ith cause and $\hat{\pi}_i$ is the estimated final levelled cumulative incidence function. The latter requires large sample with long-term follow-up as described above. Approximately parallel lines should result to support the proportional hazards assumption for the conditional distributions.

The application of the semi-parametric mixture method to a real data set has been given in Section 4. The conclusions obtained from the semi-parametric and the fully parametric approaches have some different interpretations. The reason may be that there exist five causes of death other than prostate cancer death or death due to CVD: other cancer; respiratory disease; other specific non-cancer cause; unspecified non-cancer cause, and unknown cause. These causes of death are grouped into the third component 'other causes'. Thus, the survival or the hazard function of this group may itself be a mixture, which implies that it would be inadequate to model it by the common lifetime distribution adopted for the first (prostate cancer death) and second (death due to CVD) components.

ACKNOWLEDGEMENTS

The authors wish to thank the referees for helpful comments on the paper.

REFERENCES

1. Cox DR. Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society*, *Series B* 1972; **34**:187–220.
2. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics* 1978; **34**:541–554.
3. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. Wiley: New York, 1980.
4. Benichou J, Gail MH. Estimates of absolute cause-specific risk in cohort studies. *Biometrics* 1992; **46**:813–826.
5. Gaynor JJ, Feuer EJ, Tan CC *et al*. On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *Journal of the American Statistical Association* 1993; **88**:400–409.
6. Fusaro RE, Bacchetti P, Jewell NP. A competing risks analysis of presenting AIDS diagnoses trends. *Biometrics* 1996; **52**:211–225.
7. McGiffin DC, Galbraith AJ, O'Brien MF *et al.* An analysis of valve re-replacement after aortic valve replacement with biologic devices. *Journal of Thoracic and Cardiovascular Surgery* 1997; **113**:311–318.
8. Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995; **51**:524–532.
9. Tai B, Machin D, White I, Gebski V. Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. *Statistics in Medicine* 2001; **20**:661–684.
10. Larson MG, Dinse GE. A mixture model for the regression analysis of competing risks data. *Applied Statistics* 1985; **34**:201–211.
11. Gordon NH. Application of the theory of finite mixtures for the estimation of 'cure' rates of treated cancer patients. *Statistics in Medicine* 1990: **9**:397–407.
12. Gelfand AE, Ghosh SK, Christiansen C *et al*. Proportional hazards models: a latent competing risk approach. *Applied Statistics* 2000; **49**:385–397.
13. Blackstone EH, Naftel DC, Turner ME. The decomposition of time-varying hazard into phases, each incorporating a separate stream of concomitant information. *Journal of the American Statistical Association* 1986; **81**:615–624.
14. Rajarshi S, Rajarshi MB. Bathtub distributions: a review. *Communications in Statistics—Theory and Method* 1988; **17**:2597–2621.
15. Glasser M. Bathtub and related failure rate characterizations. *Journal of the American Statistical Association* 1980; **75**:667–672.
16. Efron B. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association* 1977; **72**:557–565.
17. Oakes D. Survival times: aspects of partial likelihood. *International Statistical Review* 1981; **49**:235–264.
18. Kuk AYC. A semiparametric mixture model for the analysis of competing risks data. *Australian Journal of Statistics* 1992; **34**:169–180.
19. Kuk AYC, Cheng YW. The Monte Carlo Newton-Raphson algorithm. *Journal of Statistical Computation and Simulation* 1997; **59**:233–250.

20. Meng XL, Rubin DB. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 1993; **80**:267–278.
21. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 1982; **38**:1041–1046.
22. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *Journal of the Royal Statistical Society*, *Series B* 1977; **39**:1–38.
23. McLachlan GJ, Basford KE. *Mixture Models*: *Inference and Applications of Clustering*. Marcel Dekker: New York, 1988.
24. McLachlan GJ, Peel D. *Finite Mixture Models*. Wiley: New York, 2000.
25. Breslow NE. Contribution to the discussion of DR Cox. *Journal of the Royal Statistical Society*, *Series B* 1972; **34**:216–217.
26. Breslow NE. Covariance analysis of censored survival data. *Biometrics* 1974; **30**:89–99.
27. McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions.* Wiley: New York, 1997.
28. Meng XL. On the rate of convergence of the ECM algorithm. *Annals of Statistics* 1994; **22**:326–339.
29. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics* 2000; **56**:227–236.
30. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics* 2000; **56**:237–243.
31. Efron B. Bootstrap methods: another look at the jackknife. *Annals of Statistics* 1979; **7**:1–26.
32. Golbeck AL. Bootstrapping current life table estimators. In *Bootstrapping and Related Techniques*, Jöckel KH, Rothe G, Sendler W (eds). Springer-Verlag: Heidelberg, 1992; 197–201.
33. Ng SK, McLachlan GJ, McGiffin DC, O'Brien MF. Constrained mixture models in competing risks problems. *Environmetrics* 1999; **10**:753–767.
34. Pepe MS. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association* 1991; **86**:770–778.
35. Byar DP, Green SB. The choice of treatment for cancer patients based on covariate information: application to prostate cancer. *Bulletin du Cancer* 1980; **67**:477–490.
36. Andrews DF, Herzberg AM. *Data*: *a Collection of Problems from Many Fields for the Student and Research Worker*. Springer: New York, 1985; 261–274.
37. Kay R. Treatment effects in competing-risks analysis of prostate cancer data. *Biometrics* 1986; **42**:203–211.
38. Cheng SC, Fine JP, Wei LJ. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics* 1998; **54**:219–228.
39. Fine JP. Analysing competing risks data with transformation models. *Journal of the Royal Statistical Society*, *Series B* 1999; **61**:817–830.
40. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**:496–509.
41. Aalen OO, Johansen S. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* 1978; **5**:141–150.