Lifetime Data Anal (2006) 12:21–33
DOI 10.1007/s10985-005-7218-3

Inference for the dependent competing risks model with masked causes of failure

Radu V. Craiu · Benjamin Reiser

Received: 12 January 2005 / Accepted: 30 November 2005

Abstract The competing risks model is useful in settings in which individuals/units may die/fail for different reasons. The cause specific hazard rates are taken to be piecewise constant functions. A complication arises when some of the failures are masked within a group of possible causes. Traditionally, statistical inference is performed under the assumption that the failure causes act independently on each item. In this paper we propose an EM-based approach which allows for dependent competing risks and produces estimators for the sub-distribution functions. We also discuss identifiability of parameters if none of the masked items have their cause of failure clarified in a second stage analysis (e.g. autopsy). The procedures proposed are illustrated with two datasets.

Keywords Dependent competing risks · Masked cause · Missing data · Piecewise constant hazard · Second stage data

1. Introduction

In survival data studies it is often the case that the individuals or items under study can experience any one of J possible types or causes of failure; failure for each item being due to only one failure cause. Crowder (2001) provides a recent review of this competing risks problem for which one needs to estimate the failure rates for each cause. While items that do not fail during the experiment have no failure time/cause associated with them, it is also possible that some of the items that fail during this period have a cause of failure that is only known to belong to a certain subset of all possible failures, in other words, their cause of failure is *group masked*. In this case we say that the actual failure cause is masked by the restricted group (Sen et al. 2001).

R. V. Craiu (✉)
Department of Statistics, University of Toronto, 100 St. George Street,
Toronto, Ontario, M5S 3G3, Canada
e-mail: craiu@utstat.toronto.edu

B. Reiser
Department of Statistics, University of Haifa, Haifa, 31905, Israel
e-mail: reiser@stat.haifa.ac.il

Sometimes, one may be able to conduct a second-stage analysis, such as autopsy, in which the true cause can be uniquely determined for a sample of the masked items (Flehinger et al. 1998).

For the competing risks model with masked causes of failure, some authors have derived semiparametric and nonparametric inference procedures for the case with two failure causes and no second-stage analysis, which often occurs in carcinogenicity bioassays: Dinse (1986) proposed nonparametric maximum likelihood estimators of prevalence and mortality; Goetghebeur and Ryan (1990) derived a modified log-rank test for comparing the survival of populations, which they later extended to proportional hazards regression (Goetghebeur and Ryan 1995); Racine-Poon and Hoel (1984) considered inference for this model when a probability of death from each missing cause is provided by a pathologist; and Kodell and Chen (1987) tackled the problem via the EM algorithm. In the case of a general number of failure causes and availability of second-stage analysis data, Flehinger et al. (1998, 2002) propose maximum likelihood estimation under a model with proportional cause-specific hazards (Flehinger et al. 1998) and a model with completely parametric cause-specific hazards (Flehinger et al. 2002). Craiu and Duchesne (2004a, b) use a semiparametric model with piecewise constant hazard functions which presents robust properties and can be adapted to most situations in which some second stage data is available.

The issue of masking has been widely pursued by the reliability community. Sen et al. (2001) and Flehinger et al. (2002) provide reviews of this literature. Almost all of the research concerned with masking makes the strong symmetry assumption that the probability of masking does not depend on the true failure cause. Lin and Guess (1994) and Guttman et al. (1995) discuss how misleading this assumption can be. In addition independence of failure causes is typically assumed.

It is often the case that a good balance between flexibility and accuracy on one side and computational feasibility on the other side is achieved by the piecewise constant cause-specific hazard functions. In addition, these models allow for likelihood based methods for estimation and testing to be used (Craiu and Duchesne 2004a; Lawless 2003). It is well known that the independence between the competing risks cannot be tested using the failure times and the failure causes even if these are completely known for all items in the study. Kalbfleisch and Prentice (2002) emphasize that in many examples (e.g., life time data) dependence between the competing risks is reasonable although a parametric model for the dependency is hard to specify. In this paper, we show that models with piecewise constant hazards can be used for estimation in a general situation in which the competing risks do not act independently. If the competing risks are not independent it is known that the marginal survival functions corresponding to one cause are not identifiable, so instead we develop estimates for the sub-distribution (cumulative incidence) functions. The use of sub-distribution functions for competing risks has been stressed by many authors (e.g. Gaynor et al. 1993; Gooley et al. 1999; Kalbfleisch and Prentice 2002; Lawless 2003).

The paper is organized as follows. In Section 2 we describe the competing risks model with masked information on the cause of death/failure and the EM algorithm used for estimation of cause specific hazard rate. The estimation of sub-distribution functions is presented in Section 3. Section 4 discusses the special case where no second stage data is available. The paper concludes in Section 5 with a discussion of two different data examples.

2. Data and models

Consider n systems, each consisting of J separate modules (risks or causes) and suppose that the failure of any of the modules results in the death/failure of the system. Often one is

only able to narrow down the cause of system failure to a group of possible causes, that is the failure is group masked. Such occurrences are often met in engineering and medical applications. It may also be the case that after the first stage of the experiment, a subsample of items with group masked failure cause is selected and sent to a second stage analysis in which the exact cause of failure is determined for the selected items.

Suppose we observe n systems for a period of time of length τ. In the case of masked data, for each system i (system is henceforth called item), there are three possible occurrences, at least in the first stage of the experiment: i fails because of cause j_i at time t_i; i fails because of a cause that is not known precisely, but is known to belong to a group of failure causes $g_i \subset \{1,\ldots,J\}$; or i had still not failed by time t_i, the time we ceased observing the item i. Therefore, some of the uncensored items will have a masking group instead of a failure cause, and all the items have a failure time or a censoring time. We assume that there are G proper masking groups (i.e. they contain more than one cause) and we denote M=G+J. The observation for item i is then $(t_i, \gamma_{ig_1}, \ldots, \gamma_{ig_M}, \delta_{i1}, \ldots, \delta_{iJ})$, where γ_{ig} is the indicator that item i's failure cause was masked to group g at the first stage; if the failure cause is known to be j at the first stage, then we say that it is masked to g=$\{j\}$. Also, δ_{ij} is the indicator that item i's actual failure cause is j. If an item is right-censored, then all the indicators δ_{ij}, j=1,...,J, are 0. If an item is masked in stage one and is not taken to a second stage analysis then all δ_{ij}, j=1,...,J, are unknown. We denote by OBS the observed data.

Let T denote the failure time and C denote the failure cause. The cause-specific hazard functions are defined as

$$\lambda_j(t) = \lim_{h \downarrow 0} \frac{\Pr(t<T \le t+h, C=j | T \ge t)}{h}, \quad j=1,\ldots,J. \tag{1}$$

The sub-distribution functions are

$$F_j(t) = \Pr(T \le t, C=j) = \int_0^t \lambda_j(u) S(u) du \; j=1,\ldots,J \tag{2}$$

and, in turn, yield to the cause-specific probability density functions $f_j(t) = \frac{d}{dt} F_j(t) = \lambda_j(t) S(t)$ where $S(t)$=Pr(T>t) is the item's survival function. While we allow dependence between the risks of failure, we assume that the failure occurs due to only one cause. The overall hazard rate is $\lambda(t) = \sum_{j=1}^{J} \lambda_j(t)$ and $S(t) = exp\{-\int_0^t \lambda(u) du\}$.

In this paper we use a model with piecewise constant cause-specific hazard functions, that is, for each cause j we divide the interval $[0,\tau]$ in K_j pieces and assume that

$$\lambda_j(t) = \sum_{k=1}^{K_j} \lambda_{jk} 1_{jk}(t), \tag{3}$$

where $0 = a_{j0} < a_{j1} < \cdots < a_{jK_j}$, and $1_{jk}(t)$ is the indicator that $t \in (a_{jk-1}, a_{jk}]$.

As a result of masking, in addition to the parameters λ_{jk}, one must consider the *masking probabilities*

$$P_{g|j} = \Pr(\text{cause masked to group } g \text{ at stage 1} | C=j), \quad j \in g. \tag{4}$$

One could extend (4) to include time-dependent masking probabilities, e.g. piecewise constant $P_{g|j}(t)$ (see Craiu and Duchesne 2004a). In this paper we assume

time-independent masking probabilities and denote by $\boldsymbol{\theta}$ the set of all the parameters $(\lambda_{11},\ldots,\lambda_{JK_J},P_{g_1|1},\ldots,P_{g_G|J})$.

Using $\boldsymbol{\theta}$, one can easily calculate quantities that are of interest to practitioners such as the diagnostic probabilities (Flehinger et al. 1998, 2002)

$$\begin{aligned}\pi_{j|g}(t) =& \Pr(\text{actually failed of cause } j|\text{failed at time } t \text{ and failure cause masked in } g)\\ =& \frac{\lambda_j(t)P_{g|j}}{\sum_{l\in g}\lambda_l(t)P_{g|l}} \text{ (using Bayes rule).}\end{aligned} \tag{5}$$

For such data, Craiu and Duchesne (2004a) have developed an EM-based analysis in which the δ_{ij}, for any masked item i and for all causes of failure j, are treated as missing data. The EM algorithm relies in this case on the observation that, without masking, inference for the competing risks model with piecewise constant hazards can be carried in closed form. Moreover, since for each masked item i and each cause j the expectation of δ_{ij} is linear in the observed data, the implementation of the algorithm is straightforward.

Conditional on j being the true cause of failure for a masked item i, the random vector $\{\gamma_{ig} : g \in \{g_1,\ldots,g_{G+J}\}, g \ni j\}$ has a multinomial distribution with total 1 and $\Pr(\gamma_{ig}=1|C=j)=P_{g|j}$. Along with equations (1)–(4) this results in the following log-likelihood function.

$$\begin{aligned}l_C(\boldsymbol{\theta}) =& \sum_{i=1}^{n}\sum_{j=1}^{J}\left\{\left[\delta_{ij}\ln\sum_{k=1}^{K_j}\lambda_{jk}1_{jk}(t_i)-\sum_{k=1}^{K_j}\lambda_{jk}\int_0^{t_i}1_{jk}(u)\,du\right]\right.\\ &\left.+\delta_{ij}\left[\left(1-\sum_{g\in\mathcal{G}_j^*}\gamma_{ig}\right)\ln\left(1-\sum_{g\in\mathcal{G}_j^*}P_{g|j}\right)+\sum_{g\in\mathcal{G}_j^*}\gamma_{ig}\ln P_{g|j}\right]\right\},\end{aligned} \tag{6}$$

where $\mathcal{G}_j^*$ is the set of all groups that contain cause j with the exception of group $\{j\}$.

For the EM algorithm the E-step consists in computing the expected value of the complete data log-likelihood (6) for a given value $\boldsymbol{\theta}'$ of the parameters given the observed data, OBS, i.e., compute

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}') =& E_{\boldsymbol{\theta}'}[l_C(\boldsymbol{\theta})|\text{OBS}]\\ =& \sum_{i=1}^{n}\sum_{j=1}^{J}\left\{\left[E_{\boldsymbol{\theta}'}[\delta_{ij}|\text{OBS}]\ln\sum_{k=1}^{K_j}\lambda_{jk}1_{jk}(t_i)-\sum_{k=1}^{K_j}\lambda_{jk}\int_0^{t_i}1_{jk}(u)\,du\right]\right.\\ &\left.+E_{\boldsymbol{\theta}'}[\delta_{ij}|\text{OBS}]\left[\left(1-\sum_{g\in\mathcal{G}_j^*}\gamma_{ig}\right)\ln\left(1-\sum_{g\in\mathcal{G}_j^*}P_{g|j}\right)+\sum_{g\in\mathcal{G}_j^*}\gamma_{ig}\ln P_{g|j}\right]\right\}\end{aligned} \tag{7}$$

Then the M-step consists in finding the value of $\boldsymbol{\theta}$ that maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}')$. In our context, this is done by running the following algorithm:

Initial step Set $\hat{\lambda}_{jk}^{(0)}=\sum_{i=1}^{n}1[\delta_{ij}\text{ observed and equal to }1]\big/e_{jk}, k=1,\ldots,K_j, j=1,\ldots,J$ and $\hat{P}_{g|j}^{(0)}=1/\#\mathcal{G}_j, j=1,\ldots,J, g=g_1,\ldots,g_M$, where $\#$ denotes cardinality, $e_{jk}=\sum_{i=1}^{n}\int_0^{t_i}1_{jk}(u)du$ denotes the total time lived by all items (exposure) in the interval $(a_{jk-1},a_{jk}]$.

E-step Using (5) compute $E_{\hat{\boldsymbol{\theta}}^{(l-1)}}[\delta_{ij}|\text{OBS}]$ as

$$E_{\boldsymbol{\theta}}[\delta_{ij}|\text{OBS}] = \begin{cases} 1, & \text{cause of failure of } i \text{ known to be } j. \\ 0, & \text{cause of failure of } i \text{ known not to be } j. \\ \hat{\pi}_{j|g_i}(t_i), & \text{cause of } i \text{ masked in } g_i \text{ and no stage 2 data for } i. \end{cases} \tag{8}$$

M-step Set

$$\hat{\lambda}_{jk}^{(l)} = \frac{\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}^{(l-1)}}[\delta_{ij}|\text{OBS}]\ 1_{jk}(t_i)}{e_{jk}} \quad \text{and} \quad \hat{P}_{g|j}^{(l)} = \frac{\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}^{(l-1)}}[\delta_{ij}|\text{OBS}]\ \gamma_{ig}}{\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}^{(l-1)}}[\delta_{ij}|\text{OBS}]} \tag{9}$$

and compute $\hat{\pi}_{j|g_i}(t_i)$ from (5) using (9).

Stopping rule When $\|\hat{\boldsymbol{\theta}}^{(l)} - \hat{\boldsymbol{\theta}}^{(l-1)}\| \leq \varepsilon$, for a pre-selected (small) value of the tolerance ε. In our applications, we used the squared Euclidean norm for $\|\cdot\|$ and a tolerance $\varepsilon = 10^{-8}$.

While the marginal distributions of the competing risks are not identifiable unless the risks are independent, this does not , however, affect the identifiability of the parameters of interest, $P_{g|j}$ and λ_{jk}. We show in the next section that the sub-distribution functions can be computed in closed form and their asymptotic variance can be estimated using the SEM algorithm (Meng and Rubin 1991).

The EM algorithm converges under mild assumptions regarding the choice of intervals $(a_{jk-1}, a_{jk}]$. In particular, one needs to ensure that for each cause j and for each interval I_{jk}, there is at least one item i which has failed during I_{jk} and has a failure cause masked in a group g that contains or is identically j.

3. Estimation

3.1 Sub-distribution functions

Besides the diagnostic probabilities $\pi_{j|g}$ of particular interest in applications are the sub-distribution functions $F_j(t)$ defined in (2) for all $1 \leq j \leq J$ and $t \geq 0$.

Following Lawless (2003) we obtain

$$F_j(t) = \int_0^t S(u)\lambda_j(u)du = \int_0^t \exp\left\{-\sum_{h=1}^{J}\Lambda_h(u)\right\}d\Lambda_j(u), \tag{10}$$

where $\Lambda_j(t) = \int_0^t \lambda_j(v)dv$ for all causes j. Note that

$$S(t) = \exp\left\{-\sum_{j=1}^{J}\Lambda_j(t)\right\} = \prod_{j=1}^{J} G_j(t)$$

where $G_j(t) = \exp\{-\Lambda_j(t)\}$. Although the $G_j(t)$ have the properties of survivor functions they are not related to any observable random variables. However, for the special case of independent competing risks they do represent the marginal (cause specific) survivor functions of the J risks.

For the piecewise constant model one can compute the estimates of the sub-distribution functions since

$$\Lambda_j(t) = \int_0^t \sum_{k=1}^{K_j} 1_{jk}(u)\lambda_{jk}du = \sum_{k=1}^{K_j} \lambda_{jk}\Delta_{jk}(t), \tag{11}$$

where

$$\Delta_{jk}(t) = \begin{cases} 0 & \text{if } t \le a_{jk-1}, \\ t - a_{jk-1} & \text{if } a_{jk-1} < t \le a_{jk}, \\ a_{jk} - a_{jk-1} & \text{if } a_{jk} < t. \end{cases}$$

If we assume that the intervals have the same endpoints, i.e. $K_j = K$ and $a_{jk} = a_k$ for all $1 \le j \le J$ and $1 \le k \le K$ and if we denote $m(t) = \max\{k : a_{k-1} \le t \le a_k\}$ we obtain

$$\begin{aligned} F_j(t) &= \int_0^t \exp\left\{-\sum_{h=1}^{J}\sum_{l=1}^{K}\lambda_{hl}\Delta_l(u)\right\}\sum_{k=1}^{K}1_k(u)\lambda_{jk}du \\ &= \sum_{k=1}^{K}\int_0^t \exp\left\{-\sum_{h=1}^{J}\sum_{l=1}^{K}\lambda_{hl}\Delta_l(u)\right\}1_k(u)\lambda_{jk}du \\ &= \sum_{k=1}^{m(t)}\int_{a_{k-1}}^{\min\{a_k,t\}} \exp\left\{-\sum_{h=1}^{J}\sum_{l=1}^{k-1}\lambda_{hl}(a_l - a_{l-1})\right\}\exp\left\{-\sum_{h=1}^{J}\lambda_{hk}(u - a_{k-1})\right\}\lambda_{jk}du \\ &= \sum_{k=1}^{m(t)}\frac{\lambda_{jk}}{\lambda_{\cdot k}}\exp\left\{-\sum_{l=1}^{k-1}\lambda_{\cdot l}(a_l - a_{l-1})\right\}[1 - \exp\{-\lambda_{\cdot k}(\min\{a_k, t\} - a_{k-1})\}] \end{aligned} \tag{12}$$

where $\lambda_{\cdot k} = \sum_{j=1}^{J}\lambda_{jk}$. Thus, an estimator $\hat{F}_j(t)$ can be obtained by replacing in (12) the λ_{jk} with their estimates $\hat{\lambda}_{jk}$. Similarly, $\Lambda_j(t)$ can be estimated using $\hat{\lambda}_{jk}$ in (11).

In the case in which the intervals have different end points across causes, the above result cannot be directly extended. Instead, one should first arrange the end points in increasing order, say $0 = b_1 < b_2 \cdots < b_R$ where $R = \sum_{j=1}^{J} K_j$ and calculate $F_j(t)$ separately for $t \in [b_h, b_{h+1})$ since a recursive expression can be found as $F_j(t) = F_j(b_h) + (F_j(t) - F_j(b_h))$. Specifically,

$$F_j(b_2) = \lambda_{j1}\frac{[1 - \exp(-b_2\sum_{j=1}^{J}\lambda_{j1})]}{\sum_{j=1}^{J}\lambda_{j1}},$$

$$\begin{aligned} F_j(b_h) &= F_j(b_{h-1}) \\ &+ \sum_{\{k:a_{jk} < b_h < a_{jk+1}\}}\lambda_{jk}\left[\exp\left(-b_h\sum_{l=1}^{J}\sum_{\{k:a_{lk} < b_h < a_{lk+1}\}}\lambda_{lk}\right) - \exp\left(-b_{h+1}\sum_{l=1}^{J}\sum_{\{k:a_{lk} \le b_h < a_{lk+1}\}}\lambda_{lk}\right)\right] \\ &\times \exp\left[-\sum_{l=1}^{J}\sum_{\{k:a_{lk+1} < b_h\}}a_{lk}(\lambda_{lk} - \lambda_{lk-1})\right] \Big/ \sum_{l=1}^{J}\sum_{\{k:a_{lk} < b_h < a_{lk+1}\}}\lambda_{lk} \end{aligned} \tag{13}$$

for any $2 < h \le R$ and any $1 \le j \le J$. In addition, for any $t \in [b_h, b_{h+1})$,

$$F_j(t)=F_j(b_h)$$
$$+\sum_{\{k:a_{jk}<b_h<a_{jk+1}\}}\lambda_{jk}\left[\exp\left(-b_h\sum_{l=1}^{J}\sum_{\{k:a_{lk}<b_h<a_{lk+1}\}}\lambda_{lk}\right)-\exp\left(-t\sum_{l=1}^{J}\sum_{\{k:a_{lk}<b_h<a_{lk+1}\}}\lambda_{lk}\right)\right]$$
$$\times\exp\left[-\sum_{l=1}^{J}\sum_{\{k:a_{lk+1}<b_h\}}a_{lk}(\lambda_{lk}-\lambda_{lk-1})\right]\Big/\sum_{l=1}^{J}\sum_{\{k:a_{lk}<b_h<a_{lk+1}\}}\lambda_{lk}. \quad (14)$$

Examples for cases with equal and unequal endpoints are analyzed in Section 5.

Lawless (2003) shows that the variance of $\hat{F}_j(t)$ is $\hat{w}^{(j)'}V\hat{w}^{(j)}$ where V is the asymptotic variance of θ obtainable directly via the supplemented EM algorithm (Meng and Rubin 1991) and the matrix $\hat{w}$ has the entries given by $\hat{w}_{hk}^{(j)}=\frac{\partial F_j}{\partial\lambda_{hk}}$ estimated at the MLE. Unlike the complete data, in the case of masking the variance-covariance matrix V is not diagonal due to the correlations between various estimators induced by the EM algorithm.

The complexity of the likelihood does not stop us from using the Newton–Raphson algorithm to find the maximum likelihood estimator. However, we have chosen to work with the EM algorithm due to the natural interpretation of the masking as a missing data mechanism. In addition, the calculations required to implement the M-step are straightforward while computing the Hessian of the observed log-likelihood, as required by the Newton–Raphson, can be quite involved, especially if the number of masking groups is large.

4. No second stage data

An important special case is encountered when none of the masked items are sent for a second stage analysis. In this case, the information on the masking probabilities can be obtained only via the hazard rate estimates. When only stage-one data is available, Flehinger et al. (1998) show that under the assumption of proportional hazards for the competing risks the resulting likelihood function is over-parameterized with the model parameters being unidentifiable. Their argument is presented for independent competing risks but can readily be extended to the dependent case. Let us denote by n_c the number of censored items, n_g the number of items whose failure is masked in group g at stage 1 and n_j the number of items that have been identified during stage 1 as having failed due to cause j for all proper masking groups g and all causes $1\le j\le J$. For each item i we denote its failure time $t_i^{(j)}$ if it failed because of cause j, or $t_i^{(g)}$ if its failure cause is masked in group g. If item i is right censored its censoring time is denoted $t_i^{(c)}$. If $S(t)$ is the overall survival function, define the sub-density function of time to failure due to cause j

$$g_j(t)=\lambda_j(t)S(t). \quad (15)$$

Then, following equation (4.5) from Flehinger et al. (2001) the likelihood function for stage-one data is

$$L(\theta)=\prod_{i=1}^{n_c}S(t_i^{(c)})\prod_{j=1}^{J}\prod_{i=1}^{n_j}P_jg_j(t_i^{(j)})\times\prod_{g}\prod_{i=1}^{n_g}\sum_{r\subset g}P_{g|r}g_r(t_i^{(g)}), \quad (16)$$

where $P_j = 1 - \sum_{g \in \mathcal{G}^*} P_{g|j}$. Define the overall probability density function of the failure time $f(t) = \sum_{j=1}^{J} g_j(t)$ and $\phi_{jk} = \lambda_{jk}/\lambda_{\cdot k}$. Then

$$f(t) = \sum_{j=1}^{J} g_j(t) = \sum_{j=1}^{J} \sum_{k=1}^{K} \lambda_{jk} 1_k(t) S(t). \tag{17}$$

Replace *S(t)* from (17) in (15) and after a series of simple manipulations we obtain that for each $1 \le k \le K$ $g_j(t) = \phi_{jk} f(t)$ for $t \in (a_{k-1}, a_k]$ or, if we define $\phi_j(t) = \sum_{k=1}^{K} \phi_{jk} 1_k(t), g_j(t) = \phi_j(t) f(t)$. Then the likelihood (16) becomes

$$L = \left\{ \prod_{i=1}^{n_c} S(t_i^{(c)}) \prod_{i=1}^{n-n_c} f(t_i) \right\} \times \left[\prod_{j=1}^{J} \prod_{i=1}^{n_j} P_j \phi_j(t_i^{(j)}) \prod_{g} \prod_{i=1}^{n_g} \sum_{r \subset g} P_{g|r} \phi_r(t_i^{(g)}) \right] \tag{18}$$

and thus can be expressed as $L(\theta) = L_1 \times L_2$ where L_1 is the usual likelihood function for a system with n failures which, in the case of the piecewise constant hazards, determines the maximum likelihood estimators for the overall hazard rate $\lambda(t) = \sum_{k=1}^{K} \lambda_{\cdot k} 1_k(t)$ (Lawless 2003). In order to examine in detail L_2 define $n_j^{(k)}$ to be the number of failures due to cause j in interval $I_k = (a_{k-1}, a_k]$ and $n_g^{(k)}$ the number of failures masked in group g during the interval I_k. Then

$$\begin{aligned} L_2 &= \prod_{k=1}^{K} \left[\prod_{j=1}^{J} \left(\prod_{i=1}^{n_j^{(k)}} P_j \phi_{jk} \right) \prod_{g} \prod_{i=1}^{n_g^{(k)}} \left(\sum_{r \in g} P_{g|r} \phi_{rk} \right) \right] \\ &= \prod_{k=1}^{K} \left[\prod_{j=1}^{J} (P_j \phi_{jk})^{n_j^{(k)}} \prod_{g} \left(\sum_{r \in g} P_{g|r} \phi_{rk} \right)^{n_g^{(k)}} \right] \end{aligned} \tag{19}$$

In the particular case of two causes (J=2) and two intervals (K=2) if g={1,2} the parameters are $P_j = 1 - P_{g|j}$ and $\phi_{1j} = 1 - \phi_{2j}$ for j=1,2. From (19) we get that L_2 can be expressed as the product of two trinomials

$$\begin{aligned} L_2 =& (P_1\phi_{11})^{n_1^{(1)}} (P_2\phi_{21})^{n_2^{(1)}} (P_{g|1}\phi_{11} + P_{g|2}\phi_{21})^{n_g^{(1)}} \\ & \times (P_1\phi_{12})^{n_1^{(2)}} (P_2\phi_{22})^{n_2^{(2)}} (P_{g|1}\phi_{12} + P_{g|2}\phi_{22})^{n_g^{(2)}} \end{aligned}$$

because $P_{g|1}\phi_{11} + P_{g|2}\phi_{21} = 1 - P_1\phi_{11} - P_2\phi_{21}$. From L_2 we obtain four estimating equations for the four unknown parameters. However, depending on the numbers of masked and unmasked items failing in each interval due to cause 1 or 2 the solution to the estimating equations may not exist or may not be completely identifiable. For instance, if $n_1^{(2)} = n_2^{(2)} = n_1^{(1)} = 30$, $n_2^{(1)} = 20, n_g^{(1)} = 10$, and $n_g^{(2)} = 60$ the solutions to the maximum likelihood equations are ϕ_{12}=−1/2, ϕ_{11}=−1, P_1=−1/2 and P_2=6, obviously outside the parameter space. If, on the other hand, we consider a case in which $n_1^{(1)} = n_2^{(2)} = 15, n_1^{(2)} = n_2^{(1)} = 30$ and $n_g^{(1)} = n_g^{(2)} = 15$ then the solutions are uniquely determined as $\phi_{11} = 1/3, \phi_{12} = 2/3, P_1 = P_2 = 0.75$.

In fact, if another interval is added, the number of equations will still match the number of unknowns but the parameters have additional constraints. In our experiments we have

generated data in which solving (19) can be done exactly and data in which the solution lies outside the parameter space. In the latter case we have noticed that the EM converges to points that are highly variable according to the starting values used in the algorithm. The erratic behavior is a sign of unidentifiability under observed data and identifiability under complete data, exactly as is the case here, and has been discussed by Dempster et al. (1977). The lack of identifiability can be avoided if we choose different end-points for the intervals used in different cause-specific hazards.

For theoretical results, it is necessary to make sure that the map $Q(\boldsymbol{\theta}'|\boldsymbol{\theta})$ is continuous and has a unique maximizer (Vaida 2005). In particular, we choose the intervals such that for each failure cause j and each interval 1_{jk}, there exists an i such that $j \in g_i$ and $1_{jk}(t_i)=1$. With this choice of the intervals the conditions of Wu (1983, Theorem 2) apply and imply that the limit points of any instance $\boldsymbol{\theta}^{(l)}$ of the EM algorithm are stationary points of the observed-data loglikelihood, l_{OBS} (obtained in equation (5) in Craiu and Duchesne 2004a), and $l_{\text{OBS}}(\boldsymbol{\theta}^{(l)})$ converges monotonically to $l_{\text{OBS}}(\boldsymbol{\theta}^*)$ for some stationary point $\boldsymbol{\theta}^*$. Without second stage data θ^* may be a local maximum even if there is an identifiable global maximum so multiple starting points are always recommended.

Alternative solutions to the model without second stage data involve the assumption of symmetry, in which the masking probability $P_{g|j}$ does not depend on the cause j. Dewanji and Sengupta (2003) use an EM-based approach for grouped masked data without second stage that can be used only with the symmetry assumption. They provide an alternative approach which replaces symmetry with the strong assumption that the diagnostic probabilities are known. We do not consider the latter method. In Section 5 we reanalyze one of their datasets and find no indication of symmetry. As a result, our hazard estimates differ significantly from theirs.

5. Examples

5.1 Hard-drive data

The first example we consider here consists of 172 failure times observed over a period of 4 years during which 10,000 computer hard-drives were monitored. There are three possible causes of failure: j=1 if failure is due ''electronic hard'', j=2 if the item fails because of ''head flyability'' and j=3 if ''head/disc magnetics'' break down. A detailed description of the detection of these failure causes can be found in Flehinger et al. (2001). In summary, the detection system makes it possible to have masked data with the following masking groups $g_1=\{1,3\}$ and $g_2=\{1,2,3\}$. Some of the masked items are subjected to a laboratory analysis in which the exact cause of failure is determined. However, the prohibitive costs of such a procedure do not allow for all masked items to be diagnosed.

We fit a model with piecewise constant hazards and assume the same end-points for the intervals. The four intervals are constructed using the model selection method formulated in Craiu and Lee (2005): [0.0.81), [0.81,1.58), [1.58,3.77), [3.77,4). Flehinger et al. (2001, 2002) analyze the data assuming that the risks are independent. However, for our analysis this assumption is not necessary.

In Table 1 we show the estimates for the masking probabilities and their standard errors produced with the SEM algorithm. The estimates for the cause specific hazard rates can be found in Craiu and Lee (2005). In Fig. 1 we compare the sub-distribution function $\hat{F}_j(t) = \Pr(T \le t, C = j)$, represented by the three lines for j=1,2,3 with $1 - \hat{G}_j(t)$ represented using the plotting symbols j=1,2,3 where $\hat{G}_j(t) = \exp\{-\hat{\Lambda}_j(t)\}$. This type of

Table 1 Masking probability estimates, Flehinger et al. (2002) hard-drive data

Estimates of the $P_{g\|j}$'s			
Flehinger et al.			
Masking group	j=1	j=2	j=3
g={1,3}	0.412	0	0.446
g={1,2,3}	0.310	0.469	0.436
Our estimates			
Masking group	j=1	j=2	j=3
g={1,3}	0.410 (0.0680)	0	0.443 (0.0353)
g={1,2,3}	0.305 (0.0658)	0.448 (0.0988)	0.442 (0.0356)

Numbers in parentheses are asymptotic standard errors computed with SEM

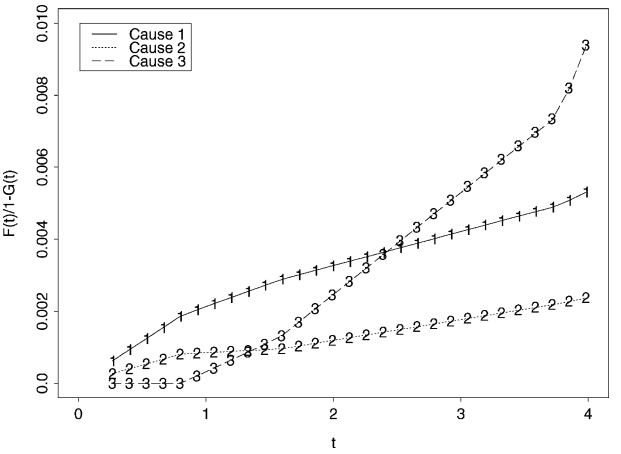

Fig. 1 Hard-drive data. The lines represent the sub-distributions functions $\hat{F}_j(t)$ for j=1,2,3. The functions $1 - \hat{G}_j(t)$ are shown for j=1,2,3 using j as a plotting symbol

graphical comparison is often seen in the literature (Gaynor et al. 1993; Gooley et al. 1999). One can see that the two graphs are right on top of each other and similar to the ones obtained by Flehinger et al. (2002) for each of the three causes. This indicates that the modeling assumptions made by Flehinger et al. (2002) are not unreasonable. For this particular dataset inference on cause specific survival seems to be robust regardless of whether one assumes piecewise constant cause specific hazards with independent or dependent causes or whether one assumes independent causes with Weibull hazards.

5.2 Carcinogenicity data

We consider here data provided by a large experiment with a total of 5000 rodents which was conducted by the British Industrial Biological Research Association (Peto et al. 1980) to investigate the carcinogenicity of different nitrosamines administered in drinking water. We follow Dewanji and Sengupta (2003) and consider here only the 192 observations corresponding to the control group. The data has been previously analyzed using a model for grouped observations and assuming symmetry of masking probabilities (Dewanji and Sengupta 2003). The data consists of the time to death (in days) and information regarding the cause of death. There are three possible causes of death as follows: j=1 if death without tumor, j=2 if death is due to some other causes but tumor is present (incidental), and j=3 if

Table 2 Masking probability estimates for the carcinogenicity data

Estimates of the $P_{g\mid j}$'s			
Masking group	j=1	j=2	j=3
g={1,2}	0.013 (0.00467)	0.046 (0.00001)	0
g={2,3}	0	0.000 (0.00314)	0.074 (0.00002)
g={1,2,3}	0.028 (0.00027)	0.000 (0.00043)	0.000 (0.00198)

Table 3 Cause-specific hazard estimates, carcinogenicity data

Estimates of the λ_{jk}'s			
	j=1	j=2	j=3
1st Interval	0.000331 (0.01207)	0.000016 (0.00002)	0.000053 (0.00049)
2nd Interval	0.002363 (0.01388)	0.000456 (0.05040)	0.000639 (0.00098)
3rd Interval	0.004711 (0.00001)	0.000863 (0.00013)	0.001748 (0.00466)
4th Interval	0.008695 (0.00030)	0.004246 (0.00005)	0.002453 (0.00001)

death is due to tumor (fatal). One can imagine that the presence of a tumor may increase the risk of death from other causes so the assumption of independence between the competing risks may be false. In addition, in nine cases, the cause of death is not exactly known and the possible masking groups are: g_1={1,2} corresponding to the situation when all we know is that the death is not due directly to tumor, g_2={2,3} corresponding to the *probably fatal* and *probably incidental* cases, and g_3={1,2,3} corresponds to the case in which nothing is known about the cause of death.

In Table 2 we present the estimates of the masking probabilities along with the SEM-derived standard errors. It can be seen that there is no indication of symmetry. We have considered four intervals for each hazard function as follows. The choice was made such that the intervals satisfy the requirement imposed for the convergence of the EM algorithm and we also tried to be close to the endpoints of the intervals chosen by Dewanji and Sengupta (2003).

Cause 1: [0,827), [827,926), [926,1001), [1001,1213)
Cause 2: [0,801), [801,951), [951,1051), [1051,1213)
Cause 3: [0,851), [851,976), [976,1101), [1101,1213)

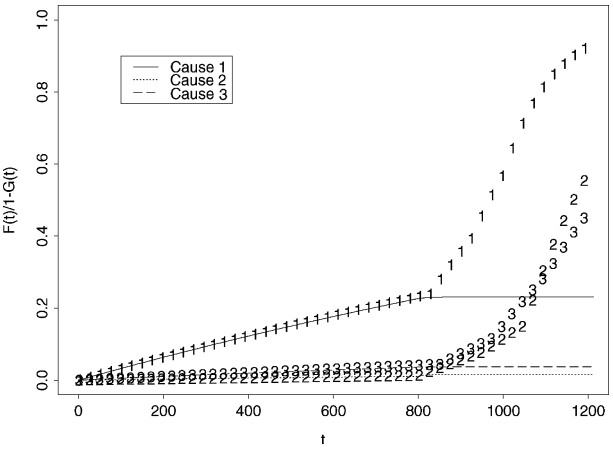

Fig. 2 Carcinogenicity data. Lines represent the sub-distribution functions $\hat{F}_j(t)$ for the three causes j=1,2,3. The functions $1-\hat{G}_j(t)$ are shown for j=1,2,3 using j as a plotting symbol

In Table 3 we show the estimates of the hazard rates for the four intervals considered and the standard errors. As in the previous example, in Fig. 2 we show $\hat{F}_j(t)$ (represented by lines) and $1 - \hat{G}_j(t)$ (using plotting symbol j) for j=1,2,3. One can see that in this case the difference is quite large indicating the importance of examining the sub-distribution functions.

Acknowledgements We would like to thank Anup Dewanji and Debasis Sengupta for making their data available to us. We also would like to thank the Associate Editor and one referee for a set of thorough and helpful suggestions. The first author's research was supported by an individual operating grant from the Natural Sciences and Engineering Research Council of Canada and an individual Connaught grant.

References

Craiu RV, Duchesne T (2004a) Inference based on the EM algorithm for the competing risk model with masked causes of failure. Biometrika 91:543–558

Craiu RV, Duchesne T (2004b) Using EM and data augmentation for the competing risks model. In: Gelman A, Meng XL (eds), Applied Bayesian modeling and causal inference from an incomplete-data perspective. Wiley

Craiu RV, Lee TCM (2005) Model selection for the competing risks model with and without masking. Technometrics 47:457–467

Crowder M (2001) Classical competing risks. Chapman & Hall

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc Ser B 39:1–22

Dewanji A, Sengupta D (2003) Estimation of competing risks with general missing pattern in failure types. Biometrics 59:1063–1070

Dinse GE (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data. J Am Stat Assoc 81:328–335

Flehinger BJ, Reiser B, Yashchin E (1998) Survival with competing risks and masked causes of failures. Biometrika 85:151–164

Flehinger BJ, Reiser B, Yashchin E (2001) Statistical analysis for masked data. In: Balakrishnan N, Rao CR (eds), Handbook of statistics, Vol. 20. Elsevier Science. pp 499–522

Flehinger BJ, Reiser B, Yashchin E (2002) Parametric modeling for survival with competing risks and masked failure causes. Lifetime Data Anal 8:177–203

Gaynor JJ, Feuer EJ, Tan CC, Wu DH, Little CR, Straus DJ, Clarkson BD, Brennan MF (1993) On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. J Am Stat Assoc 88:400–409

Goetghebeur E, Ryan L (1990) A modified log rank test for competing risks with missing failure types. Biometrika 77:151–164

Goetghebeur E, Ryan L (1995) Analysis of competing risks survival data when some failure types are missing. Biometrika 82:821–833

Gooley TA, Leisenring W, Crowley J, Storer B (1999) Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med 18:695–706

Guttman I, Lin DKJ, Reiser B, Usher JS (1995). Dependent masking and system life data analysis: Bayesian inference for two-component systems. Lifetime Data Anal 1:87–100

Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data. 2nd ed. John Wiley & Sons

Kodell RL, Chen JJ (1987) Handling cause of death in equivocal cases using the EM algorithm (with rejoinder). Commun Stat A 16:2565–2585

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

Lin DKJ, Guess FM (1994) System life data analysis with dependent partial knowledge on the exact cause of system failure. Microelectron Reliab 34:535–544

Meng XL, Rubin DB (1991) Using EM to obtain asymptotic variance: the SEM algorithm. J Am Stat Assoc 86:899–909

Peto R, Pike M, Day N, Gray R, Lee P, Parish S, Peto J, Richards S, Wahrendorf J (1980) Guidelines for simple sensitive significance tests for carcinogenic effects in long-term animal experiments. In: Long term and short term screening assays for carcinogens: a critical appraisal. International Agency for Research on Cancer, Lyon, pp 311–426

Racine-Poon AH, Hoel DG (1984) Nonparametric estimation of the survival function when cause of death is uncertain. Biometrics 40:1151–1158
Sen A, Basu S, Banerjee M (2001) Analysis of masked failure data under competing risks. In: Balakrishnan N, Rao CR (eds), Handbook of statistics, Vol. 20. Elsevier Science, pp 523–540
Vaida F (2005) Parameter convergence for the EM and MM algorithms. Stat Sinica 15:831–840
Wu CFJ (1983) On the convergence of the EM algorithm. Ann Stat 11:95–103