Lifetime Data Anal (2009) 15:41–58
DOI 10.1007/s10985-008-9091-3

The analysis of multivariate recurrent events with partially missing event types

Bingshu E. Chen · Richard J. Cook

Received: 2 April 2007 / Accepted: 25 June 2008 / Published online: 12 July 2008

Abstract In many clinical studies, subjects are at risk of experiencing more than one type of potentially recurrent event. In some situations, however, the occurrence of an event is observed, but the specific type is not determined. We consider the analysis of this type of incomplete data when the objectives are to summarize features of conditional intensity functions and associated treatment effects, and to study the association between different types of event. Here we describe a likelihood approach based on joint models for the multi-type recurrent events where parameter estimation is obtained from a Monte-Carlo EM algorithm. Simulation studies show that the proposed method gives unbiased estimators for regression coefficients and variance–covariance parameters, and the coverage probabilities of confidence intervals for regression coefficients are close to the nominal level. When the distribution of the frailty variable is misspecified, the method still provides estimators of the regression coefficients with good properties. The proposed method is applied to a motivating data set from an asthma study in which exacerbations were to be sub-typed by cellular analysis of sputum samples as eosinophilic or non-eosinophilic.

Keywords Incomplete data · Monte-Carlo EM algorithm · Multivariate recurrent events · Random effect model

B. E. Chen (✉)
Clinical Trials Group, National Cancer Institute of Canada, Queen's University,
10 Stuart Street, Kingston, ON, Canada K7L 3N6
e-mail: bechen@ctg.queensu.ca

R. J. Cook
Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West,
Waterloo, ON, Canada N2L 3G1
e-mail: rjcook@uwaterloo.ca

1 Introduction

In many studies of chronic diseases, subjects are at risk of multiple types of recurrent events. For example, in cardiovascular trials strokes may be classified according to location, in neurological studies migraines may be differentiated by severity, and in respiratory studies asthma exacerbations may be sub-typed according to a cellular analyses of sputum samples. Analysis of this kind of data can be based on intensity functions (Andersen et al. 1993), random effect models (Abu-Libdeh et al. 1990; Chen et al. 2005), or marginal models (Ng and Cook 1999; Cai and Schaubel 2004). In some situations, however, the occurrence of an event is observed but the specific type is not determined. For example, in a recent respiratory trial (Jayaram et al. 2006), asthma exacerbations were to be sub-typed by the analysis of sputum cell counts as either non-eosinophilic, characterized by variable airflow limitation, or eosinophilic, characterized by airway inflammation. Cellular analysis of sputum is not valid if patients take rescue medication for the treatment of the exacerbation before providing a sputum sample. For exacerbations where samples were provided after the use of rescue medication, the exacerbation type is therefore not determined.

This problem of missing event types with recurrent event data is analogous to the competing risk problem with masked cause of failure, which has been studied by several authors (Miyakawa 1984; Reiser et al. 1995; Lu and Tsiatis 2001; Flehinger et al. 2002; Craiu and Duchesne 2004). Schaubel and Cai (2006) consider multivariate recurrent event data with occasionally missing event types, and propose an approach based on the empirical expectation of estimating functions under proportional marginal regression models for rate functions of the different types of events.

Here we develop an alternative method for the analysis of multi-type point processes with missing event types based on multivariate random effects model. We use a likelihood-based method with the estimation carried out based on the Monte-Carlo EM algorithm to deal with the missing data problem. Because it is likelihood based it can accommodate event types that are missing at random (Molenberghs and Kenward 2007). A multivariate log normal random effect is adopted to accommodate possible positive or negative correlations between event types. In the E-step, a Gibbs' sampler (Geman and Geman 1984) is used to estimate the expectation of the complete log likelihood function conditional on the observed data. The random effect formulation determines how the information from the observed data is used to inform the imputations at the E-step; the assumption about the random effect distribution is therefore an alternative to the proportionality assumption of Cai and Schaubel (2004) and Schaubel and Cai (2006). The M-step can be carried out either using standard software (e.g. `Splus/R`) for a Poisson regression model or using Newton-Raphson algorithm with piecewise constant conditional baseline intensity functions (Cook and Lawless 2007). Upon convergence, the covariance matrix for the estimated parameters is obtained using the missing information principle (Louis 1982). This approach does not rely on any assumptions of proportional rate functions for the different types of events, and being likelihood-based, it accommodates data that may be missing at random.

The remainder of the paper is organized as follow. In Sect. 2, we give the notation and model specification for the random effects model. In Sect. 3, we develop a Monte-Carlo EM algorithm for obtaining the maximum likelihood estimate, and

provide the consistent estimator for the corresponding variance–covariance matrix. Simulation studies are reported in Sect. 4 to assess the finite sample properties of the proposed methods. The random effect model is applied to data from a recent asthma trial in Sect. 5, and concluding remarks are made in Sect. 6.

2 A multivariate random effects model

Consider a study involving m subjects in which each subject is at risk of $J \geq 2$ different types of recurrent events. Let $[0, \tau_i]$ denote the period of observation for subject i and let the function $Y_i(t) = I(t \leq \tau_i)$ indicate whether subject i is at risk for events at time $t-$, $i = 1, \ldots, m$. The right-continuous counting process $\{N_{ij}(s), s > 0\}$, records the cumulative number of events of type j for subject i where $dN_{ij}(t) = 1$ if subject i experiences an event of type j at t and is zero otherwise. One can then represent the full event data as $\{N_{ij}(s), s > 0, j = 1, \ldots, J\}$ for subject i, $i = 1, 2, \ldots, m$. Alternatively, if $dN_{i\cdot}(t) = \sum_{j=1}^{J} dN_{ij}(t)$, one can let $\{N_{i\cdot}(s), s > 0\}$ record the occurrence of *any* event and let $\delta_{ij}(t) = 1$ if $dN_{ij}(t) = 1$ and $\delta_{ij}(t) = 0$ otherwise. In this notation the event data may be represented through $\{dN_{i\cdot}(t), \boldsymbol{\delta}_i(t), t > 0\}$ where $\boldsymbol{\delta}_i(t) = [\delta_{i1}(t), \ldots, \delta_{iJ}(t)]'$. When the types of the events are sometimes missing, $\{N_{i\cdot}(s), s > 0\}$ is always observed over $[0, \tau_i]$, but $\boldsymbol{\delta}_i(t)$ may be incompletely observed over this interval. We let $\boldsymbol{x}_i$ denote a $p \times 1$ covariate vector for subject i, $i = 1, \ldots, m$.

One framework for developing multivariate models is to assume that conditional on a vector-valued random effect the events are generated by independent Poisson processes. We construct a multiplicative conditionally Poisson model (Andersen et al. 1993) for the recurrent event processes by specifying the conditional intensity functions as

$$\lambda_{ij}(t; v_{ij}, x_i) = v_{ij} Y_i(t) \lambda_{0j}(t) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_j), \quad \text{for } i = 1, \ldots, m, \text{ and } j = 1, \ldots, J, \tag{2.1}$$

where $\lambda_{0j}(t; v_{ij}, x_i)$ is the conditional baseline intensity function for events of type j, $j = 1, \ldots, J$, β_j is an associated $p \times 1$ vector of regression coefficients, and v_{ij} is a random effect. Semiparametric models can be adopted in which the functional form of $\lambda_{0j}(t)$ is left unspecified but piecewise constant functions (Lawless and Zhan 1998) represent a weakly parametric alternative that provides sufficient robustness for most applications. In this case, if we wish to have Q pieces in the piecewise constant conditional baseline intensity for events of type j, we let $0 = a_{j0} < a_{j1} < \cdots < a_{jQ_j} = \infty$ denote points on the positive real line such that $\lambda_{j0}(t) = \lambda_{jq}$ if $a_{j,q-1} < t \leq a_{jq}, q = 1, \ldots, Q_j$. We denote the parameter vector characterizing the conditional baseline intensities as $\boldsymbol{\rho}$, which in the piecewise constant model would be $\boldsymbol{\rho} = (\boldsymbol{\rho}_1', \ldots, \boldsymbol{\rho}_J')'$ where $\boldsymbol{\rho}_j = (\rho_{j1}, \ldots, \rho_{jQ_j})'$ with $\rho_{jq} = \log \lambda_{jq}, q = 1, \ldots, Q_j$. Often it is reasonable to use the same number and location of break-points for the different types of events and so let $Q_j = Q$ and $a_{jq} = a_q$, $j = 1, \ldots, J$.

We assume that the vectors of random effects $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iJ})'$, $i = 1, \ldots, m$, are independent and identically distributed and follow a multivariate log normal

distribution. Let $u_{ij} = \log(v_{ij})$, and we further assume that $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iJ})'$ has mean zero and covariance matrix Σ so that the joint p.d.f. of $\boldsymbol{u}$ is

$$g(\boldsymbol{u}_i; \Sigma) \propto \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{u}_i' \Sigma^{-1} \boldsymbol{u}_i\right). \tag{2.2}$$

Let $\mathcal{D}_i = \{t_{i1}, \cdots, t_{iK_i}\}$ be the collection of distinct event times for subject i and let $C_i(t)$ indicate the completeness of the data from subject i at $t \in \mathcal{D}_i$ so

$$C_i(t) = \begin{cases} 1 & \text{if the event type is known at time } t \\ 0 & \text{if the event type is missing at time } t. \end{cases}$$

Therefore, if $C_i(t_{ik}) = 1$ for some $t_{ik} \in D_i$, then the observed data vector at t_{ik} is $\{dN_{i\cdot}(t_{ik}), \boldsymbol{\delta}_i(t_{ik}), \boldsymbol{x}_i\}$; if $C_i(t_{ik}) = 0$, the observed data vector is $\{dN_{i\cdot}(t_{ik}), \cdot, \boldsymbol{x}_i\}$. We assume that the event types are missing at random, which means that

$$\begin{aligned} &P(C_i(t_{ik}), k = 1, \ldots, K_i | t_{ij}, \boldsymbol{\delta}_i(t_{ij}), j = 1, \ldots, K_i, \boldsymbol{x}_i) \\ &\quad = P(C_i(t_{ik}), k = 1, \ldots, K_i | t_{ij}, C_i(t_{ij}), \boldsymbol{\delta}_i(t_{ij}), j = 1, \ldots, K_i, \boldsymbol{x}_i) \end{aligned} \tag{2.3}$$

which states that the distribution of the availability of the event type at t_{ik} depends on the observable data at all event times, including observed types if available. One can sub-divide $\mathcal{D}_i$ into mutually exclusive subsets $\mathcal{D}_{ij}$, $j = 1, \ldots J$, which contain the times for those with known event type, and $\mathcal{M}_i$, which contain the times of the events with missing event types.

3 Maximization by Monte-Carlo EM

The observed data likelihood function includes a complex integration of a possibly high dimensional function. Numerical integration and numerical maximization could therefore be used to obtain the maximum likelihood estimate. As an alternative approach, we consider the EM algorithm (Dempster et al. 1977) based on the complete data comprising known event times and types ($\{dN_{i\cdot}(s), \boldsymbol{\delta}_i(s), 0 \le s \le \tau_i\}$) and random effects ($\boldsymbol{u}_i$). If the unknown parameters are defined by $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\rho}', \Sigma)'$, the complete data likelihood function is given by,

$$\begin{aligned} L_i^{(C)}(\boldsymbol{\theta}) = \prod_{j=1}^{J} \Bigg\{ &\prod_{t_{ik} \in \mathcal{D}_i} \left[\{\lambda_{0j}(t_{ik}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_j + u_{ij})\}^{\delta_{ij}(t_{ik})}\right] \\ &\times \exp\{-\Lambda_{0ij} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_j + u_{ij})\} \Bigg\} g(\boldsymbol{u}_i; \Sigma) \end{aligned}$$

where $\Lambda_{0ij} = \int_0^\infty Y_i(t)\lambda_{0j}(t)dt$ is the cumulative conditional baseline intensity function for type j events for the ith subject. The corresponding complete data log-likelihood function for subject i is given by

$$\ell_i^{(C)}(\boldsymbol{\theta}) = \sum_{j=1}^{J} \left\{ \sum_{t_{ik}} \left(\delta_{ij}(t_{ik}) \left[\log \lambda_{0j}(t_{ik}) + \boldsymbol{x}_i' \boldsymbol{\beta}_j + u_{ij} \right] \right) - \Lambda_{0ij} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_j + u_{ij}) \right\} + \log g(\boldsymbol{u}_i; \Sigma). \tag{3.1}$$

Here we describe the details of a Monte Carlo Expectation Maximization (MCEM) algorithm (Wei and Tanner 1990) to obtain the maximum likelihood estimates.
THE E-STEP: Let $\widehat{\boldsymbol{\theta}}^{(h)}$ denote the current estimate of $\boldsymbol{\theta}$. The E-step involves taking the expectation of the complete data log-likelihood function (3.1), conditional on the observed data and evaluated at $\widehat{\boldsymbol{\theta}}^{(h)}$. This conditional expectation can be partitioned into two parts, where one part contains parameters from the recurrent event processes and the other contains parameters of the random effect distribution. Thus maximizing the expected complete data log-likelihood is equivalent to maximizing

$$Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(h)}) = Q_1(\boldsymbol{\rho}, \boldsymbol{\beta}; \widehat{\boldsymbol{\theta}}^{(h)}) + Q_2(\Sigma; \widehat{\boldsymbol{\theta}}^{(h)})$$

where $Q_1(\boldsymbol{\rho}, \boldsymbol{\beta}; \widehat{\boldsymbol{\theta}}^{(h)})$ is given by

$$\sum_{i=1}^{m} \sum_{j=1}^{J} \left(\sum_{t_{ik}} \left[E\{\delta_{ij}(t_{ik}) | data; \widehat{\boldsymbol{\theta}}^{(h)}\} \{\log \lambda_{0j}(t_{ik}) + \boldsymbol{x}_i \boldsymbol{\beta}_j\} \right] - E\{e^{u_{ij}} | data; \widehat{\boldsymbol{\theta}}^{(h)}\} \Lambda_{0j} e^{x_{ij}' \boldsymbol{\beta}_j} \right),$$

and

$$Q_2(\Sigma; \widehat{\boldsymbol{\theta}}^{(h)}) = \sum_{i=1}^{m} E\{\log g(\boldsymbol{u}_i; \Sigma) | data; \widehat{\boldsymbol{\theta}}^{(h)}\}.$$

When $C_i(t_{ik}) = 1$, $\boldsymbol{\delta}_i(t_{ik})$ is observed, but otherwise we have $\delta_{i\cdot}(t_{ik}) = \sum_{j=1}^{J} \delta_{ij}(t_{ik}) = 1$, so the vector $\boldsymbol{\delta}_i(t_{ik})$ is a random variable. The conditional expectations of $\delta_{ij}(t_{ik})$ and $\boldsymbol{u}_i$ given the observed data are not analytically tractable. We therefore propose a Gibbs' sampling algorithm (Geman and Geman 1984) to obtain random samples of both $\boldsymbol{\delta}_i(t_{ik})$ and $\boldsymbol{u}_i$ from their respective conditional distributions evaluated at the current estimate of the parameter vector.

For the rth Gibbs' sample, the random samples $\boldsymbol{\delta}_i(t_{ik})$ and $\boldsymbol{u}_i$ are obtained by iterating as follows. Given an event occurred and the realized value $\boldsymbol{u}_i^{(r-1,h)}$ for the previous Gibbs' sample, the vector $\boldsymbol{\delta}_i(t)$ has a multinomial distribution. The conditional probability mass function (p.m.f.) is

$$\Pr\left\{dN_{ij}^{(r,h)}(t_{ik})=1\,\middle|\,C_i(t_{ik})=0, dN_{i\cdot}(t_{ik})=1, \boldsymbol{u}_i^{(r-1,h)}, \widehat{\boldsymbol{\theta}}^{(h)}\right\}$$
$$=\frac{\widehat{\lambda}_{0j}^{(h)}(t_k)\exp\left\{\boldsymbol{x}_i\widehat{\boldsymbol{\beta}}_j^{(h)}+u_{ij}^{(r-1,h)}\right\}}{\sum_{j=1}^{J}\widehat{\lambda}_{0j}^{(h)}(t_k)\exp\left\{\boldsymbol{x}_i\widehat{\boldsymbol{\beta}}_j^{(h)}+u_{ij}^{(r-1,h)}\right\}}, \tag{3.2}$$

and sampling from this multinomial p.m.f. is trivial.

With the missing event types imputed by the Gibbs' samples of $\left\{dN_{ij}^{(r,h)}(t)|\boldsymbol{u}_i^{(r-1,h)}, \widehat{\boldsymbol{\theta}}^{(h)}\right\}$, we let $N_{ij}^{(r,h)}=\int Y_i(t)dN_{ij}^{(r,h)}(t)$ be the total number of type j events for subject i, where $\left\{N_{ij}^{(r,h)}(t)\right\}_{j=1}^{J}$ is the complete recurrent event process. Given data from such a recurrent process, the conditional joint probability density function (j.p.d.f.) for $\boldsymbol{u}_i$ is proportional to

$$g^*(\boldsymbol{u}_i|\{N_{ij}^{(r,h)}(t)\}_{j=1}^{J}, \boldsymbol{x}_i; \widehat{\boldsymbol{\theta}}^{(h)})=\prod_{j=1}^{J}\exp\left\{N_{ij}^{(r,h)}u_{ij}-\widehat{\Lambda}_{0j}^{(h)}\exp(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_j+u_{ij})\right\}$$
$$g(\boldsymbol{u}_i;\widehat{\Sigma}^{(h)}). \tag{3.3}$$

This conditional j.p.d.f. is difficult to sample from directly, but a rejection sampling algorithm (Robert and Casella 2004) can readily be adopted. It is easy to generate samples from the multivariate normal distribution $g(\boldsymbol{u};\widehat{\Sigma}^{(h)})$, we use $G(\boldsymbol{u})=Cg(\boldsymbol{u};\widehat{\Sigma}^{(h)})$ as the envelope function, such that $g^*(\boldsymbol{u})\le G(\boldsymbol{u})$ for all $\boldsymbol{u}\in\mathcal{R}^J$. Furthermore, to maximize the acceptance probability of the rejection sampling procedure, the constant C should be as small as possible. By simple algebra, we have $C=\prod_{j=1}^{J}C_j$, where

$$C_j=\begin{cases}\exp\left[N_{ij}^{(r,h)}\left\{\log\left(\frac{N_{ij}^{(r,h)}}{\widehat{\Lambda}_{ij}^{(h)}}\right)-1\right\}\right] & \text{if } N_{ij}^{(r,h)}>0\\ 1 & \text{if } N_{ij}^{(r,h)}=0,\end{cases}$$

and $\widehat{\Lambda}_{ij}^{(h)}=\widehat{\Lambda}_{0ij}^{(h)}\exp(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_j)$. Our experience in the simulation studies suggests that choosing C in this way reduces the number of rejected samples from thousands to an average of eight to ten samples in all the settings we considered, also see remarks in Chen and Cook (2003).

The rejection sampling algorithm can be conducted in the following three steps until a sample of $u_i^{(r,h)}$ is obtained:

1. Sample $\boldsymbol{u}$ from the multivariate normal distribution $g(\boldsymbol{u};\widehat{\Sigma}^{(h)})$.

2. Sample s from the uniform distribution $U(0,1)$.
3. If $s \leq \dfrac{g^*(\boldsymbol{u})}{G(\boldsymbol{u})}$, let $u_i^{(r,h)} = \boldsymbol{u}$, else go to step 1.

After a burn-in of R_0 steps, an additional R steps of random points are sampled. Based on the law of large numbers, $E(\delta_{ij}(t)|data;\widehat{\boldsymbol{\theta}}^{(h)})$ can be empirically estimated by replacing the random variables with the generated Gibbs' samples,

$$\tilde{\delta}^{(h)}(t_{ik}) = \begin{cases} \dfrac{1}{R}\displaystyle\sum_{r=R_0+1}^{R_0+R} dN_{ij}^{(r,h)}(t_{ik}) & \text{if } C_i(t_{ik}) = 0 \\ \delta_i(t_{ik}) & \text{if } C_i(t_{ik}) = 1. \end{cases}$$

The conditional expectation of the random effect $E(v_{ij}|data;\widehat{\boldsymbol{\theta}}^{(h)}) = E(e^{u_{ij}}|data;\widehat{\boldsymbol{\theta}}^{(h)})$ can be estimated by

$$\tilde{v}_{ij}^{(h)} = \frac{1}{R}\sum_{r=R_0+1}^{R_0+R} \exp\left\{u_{ij}^{(r,h)}\right\},$$

at the hth iteration. Therefore, $Q_1(\boldsymbol{\rho}, \boldsymbol{\beta}; \widehat{\boldsymbol{\theta}}^{(h)})$ can be consistently estimated by

$$\tilde{Q}_1(\boldsymbol{\rho}, \boldsymbol{\beta}) = \sum_{i=1}^{m}\sum_{j=1}^{2}\left[\sum_{t_{ik}} \tilde{\delta}^{(h)}(t_{ik})\{\log \lambda_{0j}(t_{ik}) + \boldsymbol{x}_i\boldsymbol{\beta}_j\} - \Lambda_{0j}\exp(\boldsymbol{x}_i\boldsymbol{\beta}_j + \log \tilde{v}_{ij}^{(h)})\right], \tag{3.4}$$

and $Q_2(\Sigma; \widehat{\boldsymbol{\theta}}^{(h)})$ can be consistently estimated by

$$\begin{aligned}\tilde{Q}_2(\Sigma; \widehat{\boldsymbol{\theta}}^{(h)}) &= \frac{1}{R}\sum_{i=1}^{m}\sum_{r=R_0+1}^{R_0+R} \log g\left(\boldsymbol{u}_i^{(r,h)};\ \Sigma\right) \\ &= \frac{1}{R}\sum_{i=1}^{m}\sum_{r=R_0+1}^{R_0+R}\left\{\log|\Sigma|^{-1} - \frac{1}{2}[\boldsymbol{u}_i^{(r,h)}]'\Sigma^{-1}[\boldsymbol{u}_i^{(r,h)}]\right\}.\end{aligned}$$

To accelerate the EM algorithm, for the initial EM iterations we use a relatively small value of R_0 and R. When the parameter vector appears to be nearing the maximum likelihood estimate, we increase the number of burn-in samples (R_0) and the number of Gibbs' samples (R) to reduce simulation error.

THE M-STEP: In the M step of the EM algorithm, the functions $\tilde{Q}_1(\boldsymbol{\rho}, \boldsymbol{\beta}; \widehat{\boldsymbol{\theta}}^{(h)})$ and $\tilde{Q}_2(\Sigma; \widehat{\boldsymbol{\theta}}^{(h)})$ can be maximized separately. The first function, $\tilde{Q}_1(\boldsymbol{\rho}, \boldsymbol{\beta}; \widehat{\boldsymbol{\theta}}^{(h)})$, can be maximized using a flexible piecewise constant parametric model for the conditional baseline intensities (Lawless and Zhan 1998). Existing software for Poisson regression can be used for estimation with piecewise constant models (Cook and Lawless 2007)

if we let $\tilde{\delta}^{(h)}(t_{ik})$ be a case weight for the event at time t_{ik}, and let $\log(\tilde{v}_{ij}^{(h)})$ be an offset term.

For the second function $\tilde{Q}_2(\Sigma; \widehat{\boldsymbol{\theta}}^{(h)})$, one needs to maximize

$$\sum_{i=1}^{m} \sum_{r=R_0+1}^{R_0+R} \log |\Sigma|^{-\frac{1}{2}} - \frac{1}{2}[u_i^{(r,h)}]' \Sigma^{-1} [\boldsymbol{u}_i^{(r,h)}],$$

to find the estimate for the covariance matrix $\Sigma^{(h+1)}$. This is simply a log-likelihood function arising from a sample of *i.i.d.* observations $\left\{u_i^{(r,h)} : i = 1, \ldots, m; r = R_0 + 1, \ldots, R_0 + R\right\}$ from a mean zero multivariate normal distribution. The maximum likelihood estimate for the covariance matrix in this case is

$$\widehat{\Sigma}^{(h+1)} = \frac{1}{mR} \sum_{i=1}^{m} \sum_{r=R_0+1}^{R_0+R} [\boldsymbol{u}_i^{(r,h)}][\boldsymbol{u}_i^{(r,h)}]'.$$

By iterating between the E-step and M-step, the EM-algorithm will eventually converge to the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$, with iterations terminated when

$$\widehat{\epsilon}^{(h)} = \frac{||\widehat{\boldsymbol{\theta}}^{(h+1)} - \widehat{\boldsymbol{\theta}}^{(h)}||}{||\widehat{\boldsymbol{\theta}}^{(h)}||} \leq 10^{-4}\,. \tag{3.5}$$

This convergence criteria is used in the simulation studies and the data analysis.

Efron and Hinkley (1978) show that the observed information matrix is a more appropriate measure of information than the expected information, defined here as the expectation of the negative of the Hessian matrix of the observed data likelihood. We therefore estimate the asymptotic covariance matrix for the maximum likelihood estimates by applying the Missing Information Principle (Louis 1982) to get an estimate of the observed information matrix.

The observed information matrix for $\widehat{\boldsymbol{\theta}}$ is given by

$$I(\widehat{\boldsymbol{\theta}}) = I^{(c)}(\widehat{\boldsymbol{\theta}}) - I^{(m)}(\widehat{\boldsymbol{\theta}}),$$

where the complete data information matrix $I^{(c)}(\widehat{\boldsymbol{\theta}})$ can be consistently estimated by

$$\widehat{I}^{(c)}(\widehat{\boldsymbol{\theta}}) = -\sum_{i=1}^{m} \frac{1}{R} \sum_{r=R_0+1}^{R_0+R} \frac{\partial^2 Q_{ir}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},$$

and the missing data information matrix $I^{(m)}(\widehat{\boldsymbol{\theta}})$ can be consistently estimated by

$$\widehat{I}^{(m)}(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{m} \frac{1}{R} \sum_{r=R_0+1}^{R_0+R} \left[\frac{\partial Q_{ir}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\left[\frac{\partial Q_{ir}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right]\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} + \sum_{i\neq j}\left[\frac{1}{R}\sum_{r=R_0+1}^{R_0+R}\frac{\partial Q_{ir}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\left[\frac{1}{R}\sum_{r=R_0+1}^{R_0+R}\frac{\partial Q_{jr}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right]\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.$$

For inference about the treatment effect on all processes, we consider the following null hypothesis

$$H_0 : \beta_1 = \cdots = \beta_J = 0.$$

Global inferences can be made based on the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ and its corresponding asymptotic covariance matrix

$$\hat{\Phi} = \widehat{I}^{-1}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \hat{\Phi}_{\beta\beta} & \hat{\Phi}_{\beta\rho} & \hat{\Phi}_{\beta\Sigma} \\ \hat{\Phi}_{\rho\beta} & \hat{\Phi}_{\rho\rho} & \hat{\Phi}_{\rho\Sigma} \\ \hat{\Phi}_{\Sigma\beta} & \hat{\Phi}_{\Sigma\rho} & \hat{\Phi}_{\Sigma\Sigma} \end{pmatrix}.$$

Under the H_0, the test statistic $T = \hat{\boldsymbol{\beta}}' \hat{\Phi}_{\beta\beta}^{-1} \hat{\boldsymbol{\beta}}$ follows an asymptotic χ^2 distribution with J degrees of freedom. Other omnibus tests include the Simes' global test(Simes 1986), and those based on weighted linear combination of the $\hat{\beta}_j$ (Wei and Johnson 1985), which have been applied to multivariate recurrent event data in another context (Chen et al. 2005).

4 Simulation studies

To assess the finite sample properties of the proposed methods, we carried out simulation studies for a two sample problem in which $m = 100$ subjects were randomly allocated to one of two groups. The covariate x_i was therefore a scalar which was set to 1 if subject i was allocated in a treatment group and $x_i = 0$ otherwise. Under a mixed Poisson model we let $v_{ij} = \exp(u_{ij})$ where $\boldsymbol{u}_i = (u_{i1}, u_{i2})'$ followed a bivariate normal distribution with mean zero and covariance matrix Σ. Conditional on the frailty variable $\boldsymbol{v}_i = (v_{i1}, v_{i2})'$, the bivariate recurrent events were generated from intensity functions (2.1), using $J = 2$ and time homogeneous conditional baseline intensity functions $\lambda_{10}(t) = \lambda_{20}(t) = 2.0$. Each subject was observed continuously over the interval [0, 1]. One thousand replications were simulated for each combination of regression coefficients $(\beta_1,\ \beta_2)$ valued 0 and $\log 2$ (=0.69), frailty variable variance $\sigma_{11} = \sigma_{22} = 1.0$, and covariance $\sigma_{12} = 0$ and 0.2. After the recurrent event data were generated, a fraction MT of recurrent event types were randomly masked so that the specific event type was unknown in the data analysis. Here the percentages of missing event types were set to $MT = 0$, 10 and 20%. The EM algorithm of Sect. 2 was then used to fit the mixed Poisson model with piecewise constant conditional baseline intensity functions; to assess the validity of the procedure, a simple two piece model was used with common break points $a_{j0} = 0$, $a_{j1} = 0.5$, and $a_{j2} = 1.0$.

To accelerate the EM algorithm, in the $E-STEP$ we started with a relatively small number of burn-in runs, $R_0 = 10$, and Gibbs samples $R = 50$, and iterated until $\widehat{\boldsymbol{\theta}}^{(h)}$ was close to the maximum likelihood estimate (using the criteria $\widehat{\epsilon}^{(h)} \leq 10^{-1}$). The number of iterations depended on the sample size, the dimension of parameter vector, the starting value, and other factors. In our simulation studies, about 60 to 80 iterations were typically required for this step. Because the sample sizes R_0 and R were relatively small, this step was quite fast. Next, the number of burn-in samples was increased to $R_0 = 500$ and the number of Gibbs' samples was increased to $R =$ 5,000; this ensured that the variation from the Monte-Carlo simulation was relatively small. The MCEM algorithm was stopped when the convergence criteria (3.5) was satisfied. With this convergence criteria, an additional 20–30 iterations were typically needed in our simulation studies.

One thousand data sets were simulated and for each data set the MCEM algorithm was used for estimation. We investigated the performance of the estimator of the cumulative conditional intensity function $\Lambda_{01}(\tau)$ for event 1 at $\tau = 1.0$. The box plots in Fig. 1 show the empirical distribution of the estimator when the percentage of missing event types $MT = 20\%$ among the four different parameter configurations obtained by looking at combinations of $\exp(\beta_1) = \exp(\beta_2) = 1.0$ or 2.0 and $\sigma_{12} = 0$ or 0.2. The medians of $\widehat{\Lambda}_{01}(1.0)$ are all around 0 and 75% of the estimators are located within the interval of $[-0.13, 0.16]$ for all four settings that we considered.

The empirical biases, empirical coverage probabilities of the 95% confidence intervals, and the empirical rejection rates of Wald tests of the null hypotheses $H_{10}: \beta_1 = 0$

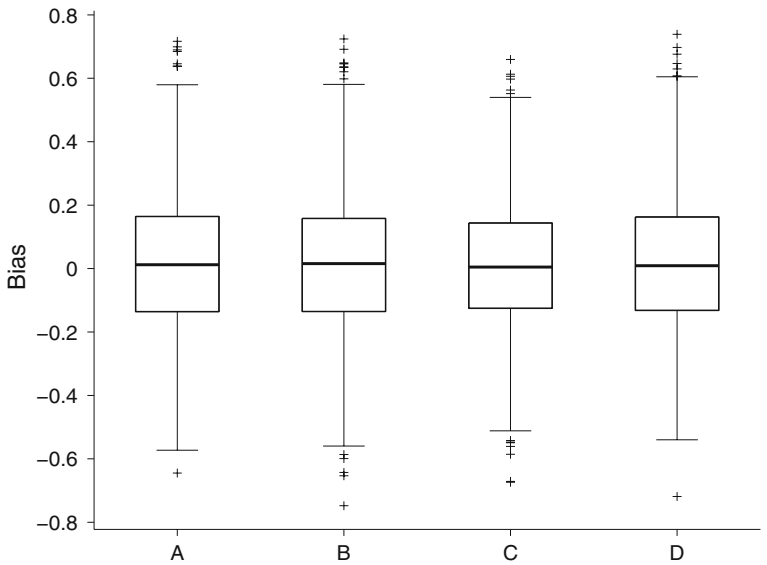

Fig. 1 Box plot of bias for the baseline cumulative conditional intensity function $\Lambda_{01}(\tau)$ at $\tau = 1.0$. Results are based on 1,000 simulation samples with $\exp(\beta_1) = \exp(\beta_2) = 1.0, \sigma_{12} = 0.0$ for plot A, $\exp(\beta_1) = \exp(\beta_2) = 2.0, \sigma_{12} = 0.0$ for plot B, $\exp(\beta_1) = \exp(\beta_2) = 1.0, \sigma_{12} = 0.2$ for plot C and $\exp(\beta_1) = \exp(\beta_2) = 2.0, \sigma_{12} = 0.2$ for plot D

Table 1 Empirical bias, coverage probabilities and rejection rates for the regression parameters

e^{β_1}	e^{β_2}	σ_{12}	MT[a]	$\hat{\beta}_1$			$\hat{\beta}_2$		
				BIAS[b]	CP[c]	REJ[d]	BIAS	CP	REJ
Mixed Poisson processes									
1.0	1.0	0.0	0	−0.007	95.1	4.9	−0.003	93.9	6.1
2.0	1.0	0.0	0	0.010	93.7	94.4	−0.001	94.6	5.4
2.0	2.0	0.0	0	−0.004	94.4	93.4	−0.007	93.8	92.7
1.0	1.0	0.2	0	0.001	94.6	5.4	0.001	94.1	5.9
2.0	1.0	0.2	0	0.009	94.9	94.0	0.001	94.5	5.5
2.0	2.0	0.2	0	0.004	94.7	93.7	0.002	94.8	93.9
1.0	1.0	0.0	10	0.008	95.1	4.9	−0.005	94.5	5.5
2.0	1.0	0.0	10	−0.004	95.2	91.5	0.009	93.8	6.2
2.0	2.0	0.0	10	−0.001	95.6	91.4	−0.002	95.3	92.6
1.0	1.0	0.2	10	−0.006	94.4	5.6	−0.001	94.5	5.5
2.0	1.0	0.2	10	0.002	94.7	85.4	−0.003	93.5	6.5
2.0	2.0	0.2	10	−0.006	95.4	83.1	0.014	95.3	82.8
1.0	1.0	0.0	20	0.011	95.1	4.9	−0.003	94.6	5.4
2.0	1.0	0.0	20	−0.008	95.4	88.4	0.010	94.1	5.9
2.0	2.0	0.0	20	0.029	95.2	92.3	0.006	96.4	91.4
1.0	1.0	0.2	20	0.016	94.0	6.0	0.005	95.0	5.0
2.0	1.0	0.2	20	0.008	94.6	77.5	0.001	94.6	5.4
2.0	2.0	0.2	20	−0.015	94.7	78.8	0.015	94.3	80.2
Bivariate binary frailty									
1.0	1.0	0.2	10	−0.002	94.3	5.7	−0.001	94.9	5.1
2.0	1.0	0.2	10	−0.001	95.6	94.8	−0.002	94.5	5.5
2.0	2.0	0.2	10	0.006	94.7	95.6	0.004	94.0	94.8
Additive gamma frailty									
1.0	1.0	0.2	10	0.007	95.1	4.9	−0.013	95.1	4.9
2.0	1.0	0.2	10	0.010	95.4	92.2	0.008	94.4	5.6
2.0	2.0	0.2	10	0.016	95.3	93.4	0.006	94.4	93.2

Results are based on 1,000 simulation samples. Subjects were observed in interval $[0, 1]$, with $m = 100$, $\lambda_{10} = \lambda_{20} = 2.0$, and $\sigma_{11} = \sigma_{22} = 0.5$

[a] Percentage of subjects with missing event type

[b] Empirical biases of the point estimate

[c] Empirical coverage probabilities of the 95% confidence intervals

[d] Empirical rate of rejection the null hypothesis $\beta_j = 0$, for $j = 1$ or 2, respectively

$H_{20}: \beta_2 = 0$ and $H_{30}: \sigma_{12} = 0$ are reported. Table 1 (upper panel) displays the results for the regression coefficients β_1 and β_2 based on 1000 simulations. For the mixed Poison distribution, under all settings considered, the empirical biases are all very small for the regression coefficients β_1 and β_2. We also conducted several simulations for sample size $m = 50$ and 200 (data not show), and observed that the empirical biases decrease when the sample size increases. The empirical coverage probabilities for parameters β_1 and β_2 are all close to the nominal level of 95%. Table 1 also shows the power of rejecting the null hypothesis H_{10} or H_{20}.

Table 2 (upper panel) displays the empirical biases and the empirical coverage probabilities of the 95% confidence intervals for the variance–covariance parameters. The empirical biases are all relatively small for σ_{11}, σ_{22} and σ_{12} While the performances of the coverage probabilities for parameters σ_{11} and σ_{22} are not as good as what we

Table 2 Empirical bias, coverage probabilities and rejection rates for the variance and covariance parameters

e^{β_1}	e^{β_2}	σ_{12}	MT[a]	$\hat{\sigma}_{11}$		$\hat{\sigma}_{22}$		$\hat{\sigma}_{12}$		
				BIAS[b]	CP[c]	BIAS	CP	BIAS	CP	REJ[d]
Mixed Poisson processes										
1.0	1.0	0.0	0	−0.017	92.9	−0.014	94.1	−0.006	96.2	3.8
2.0	1.0	0.0	0	−0.019	93.4	−0.003	93.7	0.001	95.9	4.1
2.0	2.0	0.0	0	−0.020	94.0	−0.017	93.3	−0.006	94.7	5.3
1.0	1.0	0.2	0	−0.025	92.3	−0.021	93.1	−0.010	94.1	49.8
2.0	1.0	0.2	0	−0.019	92.8	0.023	92.7	−0.009	95.2	50.2
2.0	2.0	0.2	0	−0.024	93.3	−0.017	93.8	−0.009	95.8	50.6
1.0	1.0	0.0	10	−0.023	93.4	−0.030	94.7	−0.020	96.7	3.3
2.0	1.0	0.0	10	−0.026	94.3	−0.023	93.3	−0.016	95.5	4.5
2.0	2.0	0.0	10	−0.021	94.8	−0.021	94.2	−0.020	95.0	5.0
1.0	1.0	0.2	10	−0.035	94.9	−0.029	94.6	−0.052	94.3	16.2
2.0	1.0	0.2	10	−0.027	94.4	0.021	95.2	−0.048	94.1	17.6
2.0	2.0	0.2	10	−0.033	92.5	−0.030	93.3	−0.037	93.8	20.2
1.0	1.0	0.0	20	−0.029	93.5	−0.018	94.3	−0.037	94.5	5.5
2.0	1.0	0.0	20	−0.028	92.6	0.021	93.4	−0.048	95.1	4.9
2.0	2.0	0.0	20	−0.026	93.2	−0.037	94.7	−0.038	94.9	5.1
1.0	1.0	0.2	20	−0.022	95.8	−0.027	94.6	−0.059	96.0	15.4
2.0	1.0	0.2	20	−0.025	92.3	−0.025	95.2	0.049	96.2	16.0
2.0	2.0	0.2	20	−0.024	93.9	−0.041	92.2	−0.054	94.5	19.6
Bivariate binary frailty										
1.0	1.0	0.2	10	−0.123	81.1	−0.122	81.4	−0.051	90.4	33.4
2.0	1.0	0.2	10	−0.106	85.4	−0.112	81.6	−0.045	91.9	37.5
2.0	2.0	0.2	10	−0.120	83.2	−0.110	82.1	−0.046	90.1	36.6
Additive gamma frailty										
1.0	1.0	0.2	10	−0.066	88.9	−0.076	88.8	−0.127	75.3	9.2
2.0	1.0	0.2	10	−0.068	89.2	−0.070	88.9	−0.128	75.0	8.2
2.0	2.0	0.2	10	−0.073	89.6	−0.066	90.2	−0.105	97.0	2.9

Results are based on 1,000 simulation samples. Subjects were observed in interval $[0, 1]$, with $m = 100$, $\lambda_{10} = \lambda_{20} = 2.0$, and $\sigma_{11} = \sigma_{22} = 0.5$

[a] Percentage of subjects with missing event type

[b] Empirical biases of the point estimate

[c] Empirical coverage probabilities of the 95% confidence intervals

[d] Empirical rate of rejection the null hypothesis $\sigma_{12} = 0$

observed for β_1 and β_2, the empirical coverage probability for σ_{12} is satisfactory. Since the variance parameters σ_{11} and σ_{22} were restricted to be positive, we used the logarithm transformation to improve the performance of their empirical coverage probabilities. Table 2 also shows the power of rejecting the null hypothesis H_{30}.

Compared to the results when $MT = 0\%$, when $MT = 10$ and 20%, the empirical biases and coverage probabilities are quite stable. However, when the percentage of missing event types increases, the power of rejecting the null hypothesis: $H_{10} : \beta_1 = 0$, $H_{20} : \beta_2 = 0$ or $H_{30} : \sigma_{12} = 0$ decreases. We also do some additional simulations to evaluate the power of rejecting the null hypothesis that the overall event rates are the same between two groups when the event types are not distinguished. Our simulation shows that under the setting of $\beta_1 = \log(2)$ and $\beta_2 = 0$ the proposed method has

higher power to reject either H_{10} or H_{20} with Bonferroni correction than the method that does not distinguish the event types.

To study the performances of the random effect model when the distribution of the frailty variables are misspecified, we assessed the finite sample properties of the proposed log normal frailty model (2.1) when the distribution of the frailty variable was as follows. The following two scenarios were considered: bivariate binary and additive Gamma frailties. For the bivariate binary frailty, data were generated the same way as the log normal frailty model, with $\boldsymbol{u}_i = (u_{i1}, u_{i2})'$, where

$$\begin{cases} P(u_{i1} = \sqrt{\sigma_{11}}, u_{i2} = \sqrt{\sigma_{22}}) & = \frac{\sqrt{\sigma_{11}\sigma_{22}}+\sigma_{12}}{4\sqrt{\sigma_{11}\sigma_{22}}} \\ P(u_{i1} = \sqrt{\sigma_{11}}, u_{i2} = -\sqrt{\sigma_{22}}) & = \frac{\sqrt{\sigma_{11}\sigma_{22}}-\sigma_{12}}{4\sqrt{\sigma_{11}\sigma_{22}}} \\ P(u_{i1} = -\sqrt{\sigma_{11}}, u_{i2} = \sqrt{\sigma_{22}}) & = \frac{\sqrt{\sigma_{11}\sigma_{22}}-\sigma_{12}}{4\sqrt{\sigma_{11}\sigma_{22}}} \\ P(u_{i1} = -\sqrt{\sigma_{11}}, u_{i2} = -\sqrt{\sigma_{22}}) & = \frac{\sqrt{\sigma_{11}\sigma_{22}}+\sigma_{12}}{4\sqrt{\sigma_{11}\sigma_{22}}}. \end{cases}$$

This gives the discrete frailty variable $\boldsymbol{u}_i$ mean zero and covariance matrix Σ. For the additive Gamma frailty, we let $v_{i1} = v_{i0} + v_{i1}^*$ and $v_{i2} = v_{i0} + v_{i2}^*$, where v_{i0}, v_{i1}^*, and v_{i2}^* are independent Gamma random variables with the same shape parameter. Under this formulation one can select the scale parameters such that $\boldsymbol{v}_i$ has mean one and covariance matrix Σ.

Simulation studies (Table 2, lower panel) suggest that the variance parameters for the random effects were underestimated when the distribution of the frailty variable was misspecified. The coverage probabilities of the 95% confidence intervals for the variance parameters of the random effects also depart from the nominal level. Biases for the regression parameters remained very small (Table 1, lower panel), however, and the coverage probabilities of the 95% confidences intervals for the regression parameters remained consistent with the nominal level. This indicates that the likelihood based method still provides a good point estimator for the regression coefficient and the corresponding standard errors even when the distribution of the frailty variable is misspecified. This is consistent with the findings of Ng and Cook (2000) and Cook and Wei (2003) who found that misspecification of the random effect distribution for mixed Poisson models did not materially affect the performance of estimators of regression coefficients, but did lead to poorly behaved estimators of the variance parameters of the random effect distribution.

5 Application to an asthma trial

Here we consider data from a recent clinical trial involving different treatment strategies for asthma patients. The traditional clinical strategy (CS) involved prophylactic dosing based on respiratory symptoms and spirometry readings, while the experimental sputum strategy (SS) was based on the measurement of airway inflammation by sputum cell counts. In this study, 100 adults with asthma were randomly assigned to one of these two treatment groups, with 52 in the CS group (control, $x_i = 0$), and 48 in the SS group (experimental, $x_i = 1$). Patients were followed up for 24 months after their maintenance visit (randomization), although some were withdrawn early.

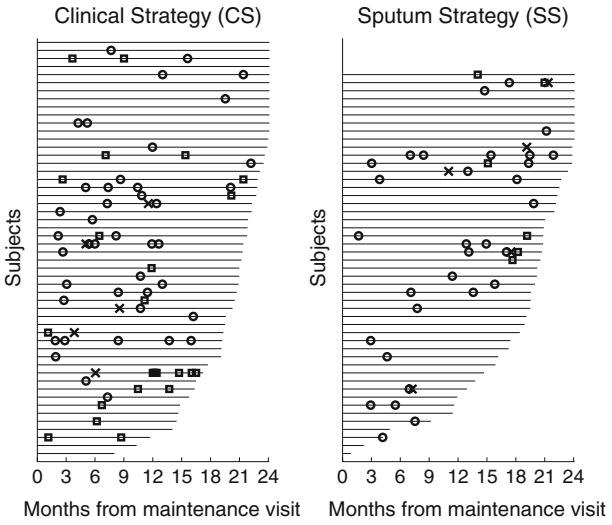

Fig. 2 Plot of sample profiles

The length of followup was represented as a horizontal line for each subject in Fig. 2. There were total 125 (73 in CS and 42 in SS) exacerbations during the 1966.8 person-months of followup (1069.5 for CS and 897.3 for SS).

Exacerbations were made up of two sub-types: 84 (48 in CS and 36 in SS) non-eosinophilic exacerbations (shown as circle in Fig. 2) and 31 (25 in CS and 6 in SS) eosinophilic exacerbations (shown as square in Fig. 2). If rescue medication was taken for exacerbations before sputum samples were collected, the cellular analysis of sputum samples were invalidated and hence the sub-type were unknown. In this study, there was 10 (5 in CS and 5 in SS) exacerbations of unknown subtype (shown as cross in Fig. 2).

The random effect model (2.1) was fitted to these data with a single treatment versus control covariate x as described above. Cut-points for the conditional baseline intensity functions were chosen to be at the 6, 12, 18, and 24 months after the initial maintenance visit. Similar results were obtained with other cut-points. The top-left panel of Table 3 provides the results for both the random effect approach without (Independent Model (1)) and with (Joint Model (2)) accounting for the correlation between two event types (i.e. $\sigma_{12} \neq 0$ and $\sigma_{12} = 0$). Both models suggest that sputum strategy reduces the occurrence of eosinophilic exacerbations. Wald test fails to reject the null hypothesis of $\sigma_{12} = 0$ (p-value 0.327), so there is not much evidence against the two subtypes of exacerbations arise independently.

Conditional on the frailty variable, Model (2) suggests that eosinophilic asthma exacerbations were reduced in the sputum strategy group compare to the clinical strategy group, with relative risk $RR = 0.319$ (95% confidence interval [0.109, 0.936]). The p-value for the two-sided test of null hypothesis $H_0 : \beta_1 = 0$ is 0.0375. For the non-eosinophilic exacerbations, we have the conditional relative risk $RR = 0.952$ with (95% confidence interval [0.575, 1.576]).

Table 3 Point estimates and standard errors for regression coefficients and variance–covariance parameters in the asthma study in Jayaram et al. (2006)

Type of exacerbation	Random effect models				Omitting events with missing type			
	Independent(1)		Joint(2)		Independent(3)		Joint(4)	
	EST.	S.E.	EST.	S.E.	EST.	S.E.	EST.	S.E.
Eosinophilic								
β_1	−1.165	0.560	−1.142	0.549	−1.217	0.586	−1.181	0.583
σ^2_{11}	1.685	0.508	1.722	0.805	1.867	0.772	1.814	0.871
Non-eosinophilic								
β_2	−0.049	0.257	−0.025	0.261	−0.092	0.264	−0.068	0.267
σ^2_{22}	0.449	0.219	0.464	0.205	0.425	0.223	0.439	0.324
σ_{12}			−0.396	0.374			−0.484	0.480
Any type								
β	−0.367	0.217						
σ^2	0.261	0.117						

The top-right panel of Table 3 gives the results when the events with missing types are removed from the data (Independent Model (3) and Joint Model (4)). By comparing the standard errors from Model 1 and 2 with those from Model (3) and (4), one can see the efficiency gain when information on the occurrence of all events is used.

Finally, we considered an analysis of total event count, which has the advantage of no suffer from the missing event type problem. The lower panel of Table 3 reports the results when the two event types were combined into one. The sputum strategy treatment reduced the total number of exacerbations ($RR = 0.693$, 95% confidence interval [0.453, 1.061]), but the outcome was not statistically significant (p-value 0.091).

The fitted cumulative mean functions for both CS and SS arms from Model (2) based on piecewise constant baseline rates are plotted in Fig. 3.

6 Discussion

The analysis of multi-type recurrent events is sometimes complicated by the fact that event types are not always available. In some settings biases arises when one ignores the events with missing type, and certainly there is a loss in efficiency if one does not use all the available data. In this paper we propose a likelihood based approach to address this problem when the unobserved event type is missing at random. The motivation of our approach is to formulate a model in which the observed data influence the "imputation" at the E-step of an EM algorithm. The multivariate random effect formulation does this quite nicely in the sense that the conditional distribution of the random effects given the data reflects the frequency and types of events observed over $[0, \tau_i]$. Simulation studies suggest that the proposed methods yield unbiased estimates and that valid covariance estimates are obtained with moderate samples. Moreover, estimators for the regression coefficients remain excellent and the

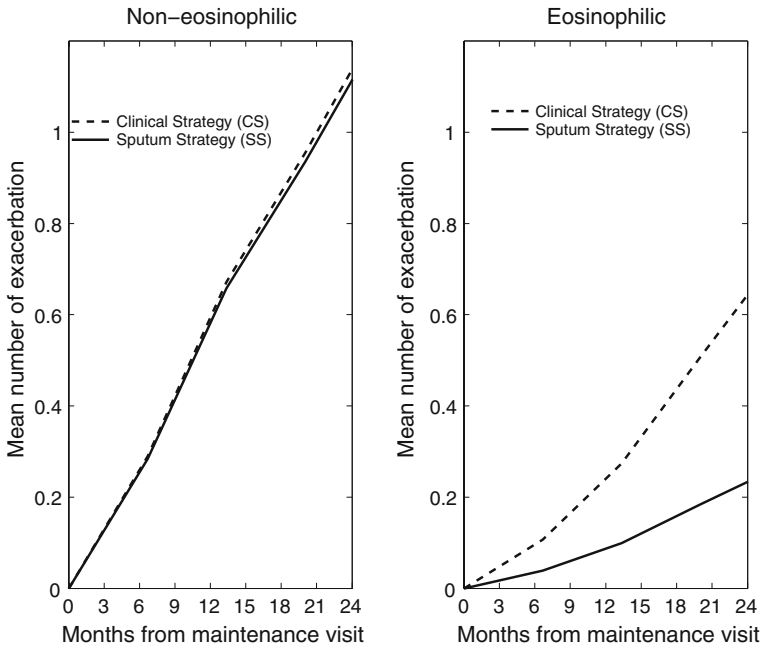

Fig. 3 Plot of cumulative mean functions for the number of exacerbations by type and treatment strategy

corresponding standard errors still lead to valid inferences when the distribution of the random effects is misspecified.

The computational challenges arising in the implementation of the MCEM algorithm was addressed in two ways. First, to maximize the acceptance probability of the rejection sampling procedure, we found the smallest envelope function for the target function (3.3) in order to obtain a high acceptance rate. Second, we varied the number of burn-in samples and Gibbs' samples in the Monte-Carlo part of the algorithm to maintain a balance between speed of the algorithm and the accuracy of parameter estimation.

When all recurrent event types are observed, the problem discussed in the paper reduces to the correlated multivariate recurrent event processes. The software developed in this paper for likelihood analyses can be easily modified to fit random effects models for the interval-censored multivariate recurrent event data (Chen et al. 2005), which was previous handled by a Bayesian approach using software BUGS (Gilks et al. 1994). In this way, one can conduct frequentist inferences for the analysis of interval-censored multivariate recurrent event data. Furthermore, the reduction in the computational burden also makes it possible to conduct simulation studies for this kind of problem, which is almost impossible to do in the Bayesian framework. Tools for model diagnostics are of importance in this and other settings to assess the plausibility of the assumption of proportionality of intensity functions in the presence of missing event types.

In conclusion, we applied the proposed method to a clinical trial involving the use of sputum cell counts to guide prophylactic treatment of asthma exacerbations. We

found that the new treatment based on the sputum cell count (SS) significantly reduces the number of eosinophilic exacerbations compared to the clinical treatment strategy (CS). But, there was little evidence of an effect on the incidence of non-eosinophilic exacerbations. This finding, together with the large and significant frailty variance for the eosinophilic outcome, represent useful information for the guidance of treatment in asthma. Specifically, patients with a history of frequent eosinophilic exacerbations are at increased risk of future eosinophilic exacerbations. These patients have the greatest potential to benefit from a treatment strategy informed by cellular analysis of sputum samples.

Acknowledgements We thank the two reviewers and the Associate Editor for their careful review of our paper and useful comments, and Dr. Liesel D'Silva for permission to use the data from the asthma study. This research was supported by grants to the second author from the Natural Sciences and Engineering Research Council and the Canadian Institutes for Health Research. RJ Cook is Canada Research Chair in Statistical Methods for Health Research. The simulation studies and data analysis for this paper were conducted using Matlab. The Matlab programs for the "core functions" are available from the author upon request.

References

Abu-Libdeh H, Turnbull BW, Clark LC (1990) Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial. Biometrics 46(4):1017–1034

Andersen PK, Borgan O, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer-Verlag, New York

Cai J, Schaubel DE (2004) Marginal means/rates models for multiple type recurrent event data. Lifetime Data Anal 10(2):121–138

Chen BE, Cook RJ (2003) Regression modeling with recurrent events and time-dependent interval-censored marker data. Lifetime data Anal 9:275–291

Chen BE, Cook RJ, Lawless JF, Zhan M (2005) Statistical methods for multivariate interval-censored recurrent events. Stat Med 24:671–691

Cook RJ, Lawless JF (2007) The statistical analysis of recurrent events. Springer, New York

Cook RJ, Wei W (2003) Conditional analysis of mixed poisson processes with baseline counts: implications for trial design and analysis. Biostatistics 4:479–494

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

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the em algorithm. J Roy Stat Soc Ser B 39:1–38

Efron B, Hinkley DV (1978) Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. Biometrika 65:457–482

Flehinger BJ, Reiser B, Yashchin E (2002) Parametric modeling for survival with competing risks and masked failure causes. Lifetime Data Anal 8(2):177–203

Geman S, Geman D (1984) Gibbs distributions and the bayesian restoration of images. IEEE Trans Pattn Anal Mach Intell 6:721–741

Gilks WR, Thomas A, Spiegelhalter DJ (1994) A language and program for complex bayesian modelling. statistician 43:169–178

Jayaram L, Pizzichini MM, Cook RJ, Boulet LP, Lemière C, Pizzichini E, Cartier A, Hussack P, Goldsmith CH, Laviolette M, Parameswaran K, Hargreave FE (2006) Determining asthma treatment by monitoring sputum cell counts: effect on exacerbations. Eur Respir J 27(3):483–494

Lawless JF, Zhan M (1998) Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. Can J Stat 26:549–565

Louis TA (1982) Finding observed information using the em algorithm. J Roy Stat Soc Ser B 44:226–233

Lu K, Tsiatis AA (2001) Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. Biometrics 57:1191–1197

Miyakawa M (1984) Analysis of incomplete data in competing risks model. IEEE Trans Reliab 33:293–296

Molenberghs G, Kenward MG (2007) Missing data in clinical studies. Wiley, New York
Ng ETM, Cook RJ (1999) Robust inference for bivariate point processes. Can J Stat 27:509–524
Ng ETM, Cook RJ (2000) A comparison of some random effect models for parameter estimation in recurrent events. In: Haynatzki GR, Haynatzka VR, Gani J, Rachev ST (eds) A special issue on stochastic models in mathematical biology. Mathematical and computer modeling. Elsevier, San Diego, pp 11–26
Reiser B, Guttman I, Lin DKJ, Guess FM, Usher HS (1995) Bayesian inference for masked system lifetime data. Appl Stat 44:79–90
Robert CP, Casella G (2004) Monte Carlo statistical methods, 2nd edn. Springer-Verlag, New York
Schaubel D, Cai J (2006) Rate/mean regression for multiple-sequence recurrent events data with missing event category. Scand J Stat 33:191–207
Simes RJ (1986) An improved bonferroni procedure for multiple tests of significance. Biometrika 73:751–754
Wei LJ, Johnson WE (1985) Combining dependent tests with incomplete repeated measurements. Biometrika 72:359–364
Wei GCG, Tanner MA (1990) A Monte Carlo implementation of the em algorithm and the poor man's data augmentation algorithm. J Am Stat Assoc 85:699–704