Biometrika (2009), **96**, 2, *pp*. 445–456
doi: 10.1093/biomet/asp010

Advance Access publication 24 April 2009
Printed in Great Britain

Marginal analysis of panel counts through estimating functions

BY X. JOAN HU

Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6
joanh@stat.sfu.ca

STEPHEN W. LAGAKOS

Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.
lagakos@biostat.harvard.edu

AND RICHARD A. LOCKHART

Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6
lockhart@stat.sfu.ca

SUMMARY

We develop nonparametric estimation procedures for the marginal mean function of a counting process based on periodic observations, using two types of self-consistent estimating equations. The first is derived from the likelihood studied by Wellner & Zhang (2000), assuming a Poisson counting process. It gives a nondecreasing estimator, which equals the nonparametric maximum likelihood estimator of Wellner & Zhang and is consistent without the Poisson assumption. Motivated by the construction of parametric generalized estimating equations, the second type is a set of data-adaptive quasi-score functions, which are likelihood estimating functions under a mixed-Poisson assumption. We evaluate the procedures using simulation, and illustrate them with the data from a bladder cancer study.

Some key words: Counting process; Interval censoring; Marginal mean function; Nonparametric estimation; Quasi-score function.

1. INTRODUCTION

Panel counts are common in medical research, field reliability and other areas. A familiar example is the recurring bladder tumour data presented by Byar (1980). Lawless & Zhan (1998) refer to such data as interval-grouped recurrent events. Sun (2006) provides a thorough review of related data structures, including various types of interval censoring.

Let $N_i(\cdot) = \{N_i(t) : t > 0\}$ with $N_i(0) = 0$ be the underlying counting process associated with individual i in a study with n independent subjects. Our interest is in nonparametric estimation of the marginal mean function $\Lambda_0(t) = E\{N_i(t)\}$, $t > 0$, with the periodic observations, i.e. panel counts,

$$\{(t_{i,k}, N_i(t_{i,k})) : k = 1, \ldots, K_i; i = 1, \ldots, n\}, \tag{1}$$

where $0 < t_{i,1} < \cdots < t_{i,K_i}$ are the observation times for individual i, and are independent of $N_i(\cdot)$. The observation times can vary from individual to individual. When $N_1(\cdot), \ldots, N_n(\cdot)$ are independent realizations of a Poisson process and are observed subject to right censoring, the Nelson–Aalen estimator (Andersen et al., 1992, p. 55) is commonly used for estimating the cumulative intensity function, which is the marginal mean function $\Lambda_0(\cdot)$. Lawless & Nadeau (1995) and Lin et al. (2000) note that the Nelson–Aalen estimator is consistent for $\Lambda_0(\cdot)$ without the Poisson assumption. When the observations are panel counts, Sun & Kalbfleisch (1995) present a monotone estimator of the mean function based on the isotonic regression. Under the assumption that the counting process is Poisson, Wellner & Zhang (2000) indicate that Sun & Kalbfleisch's estimator is a maximum pseudolikelihood estimator, and derive the nonparametric maximum likelihood estimator from panel counts, with the constraint that the mean function is nondecreasing. Wellner & Zhang show that both estimators are consistent without the Poisson assumption, and demonstrate via simulation that the nonparametric maximum likelihood estimator is more efficient than the nonparametric maximum pseudolikelihood estimator. Hu et al. (2009) consider a consistent estimator derived from a generalized sum of squares subject to the monotonicity constraint. Their estimator with a particular weight reduces to the estimator given by Sun & Kalbfleisch (1995); with another weight, it is closely related to the nonparametric maximum likelihood estimator of Wellner & Zhang (2000). To achieve monotonicity of the estimated mean function, both Wellner & Zhang (2000) and Hu et al. (2009) apply the iterative convex minorant algorithm or its modification (Jongbloed, 1998).

This paper proposes alternative estimation procedures through estimating functions, aiming at easily implementable and transparent procedures. We extend ideas used by Kaplan & Meier (1958), Turnbull (1976) and Nadeau & Lawless (1998), and estimate the mean function $\Lambda_0(\cdot)$ by taking the potentially best applicable estimator of the finite number of parameters $\lambda_j = \Lambda_0(s_j) - \Lambda_0(s_{j-1})$, $\hat{\lambda}_j$ say, for $j = 1, \ldots, J$, where $0 = s_0 < \cdots < s_J < \infty$ are the distinct observation times in the study, and assuming there is no change in the mean response at other times. The two estimators of $\Lambda_0(\cdot)$ studied in this paper can be expressed as

$$\hat{\Lambda}(t) = \sum_{j:s_j \leqslant t} \hat{\lambda}_j \quad (t > 0), \tag{2}$$

with $\hat{\lambda}_j$ obtained from two sets of estimating functions. Our estimating functions may be viewed as derived with data-dependent piecewise-constant mean function models, and thus look similar to those in Lawless & Zhan (1998) and their extensions in Chen et al. (2005) with the parametric assumption of piecewise-constant rate functions. Our estimators are derived from the expectations of efficient estimating functions based on more complete datasets conditional on the available data.

2. A SELF-CONSISTENT ALGORITHM

With $0 = s_0 < \cdots < s_J$ denoting the distinct values of the observation times $t_{i,k}$ ($k = 1, \ldots, K_i$; $i = 1, \ldots, n$), let $\lambda_j = \Lambda_0(s_j) - \Lambda_0(s_{j-1})$ ($j = 1, \ldots, J$) denote the corresponding increments of $\Lambda_0(\cdot)$. The likelihood function with panel counts under the Poisson assumption,

$$L_{\mathrm{WZ}}(\Lambda) = \prod_{i=1}^{n} \prod_{k=1}^{K_i} \{\Lambda(t_{i,k}) - \Lambda(t_{i,k-1})\}^{N_i(t_{i,k})-N_i(t_{i,k-1})} \exp[-\{\Lambda(t_{i,k}) - \Lambda(t_{i,k-1})\}], \tag{3}$$

is now a function of $\lambda = (\lambda_1, \ldots, \lambda_J)'$, where $t_{i,0} = 0$ ($i = 1, \ldots, n$). The resulting nonparametric maximum likelihood estimator (Wellner & Zhang, 2000), denoted by $\hat{\Lambda}_{\mathrm{WZ}}(\cdot)$, is the constrained solution of

$$\frac{\partial \log L_{\mathrm{WZ}}(\hat{\Lambda})}{\partial \lambda_l} \leqslant 0 \quad (l = 1, \ldots, J), \qquad \sum_{j=1}^{J} \hat{\lambda}_j \frac{\partial \log L_{\mathrm{WZ}}(\hat{\Lambda})}{\partial \lambda_j} = 0, \tag{4}$$

subject to $\hat{\lambda}_j \geqslant 0$. The solution of

$$\frac{\partial \log L_{\mathrm{WZ}}(\Lambda)}{\partial \lambda_j} = \sum_{i=1}^{n} \sum_{k=1}^{K_i} I\{s_j \in (t_{i,k-1}, t_{i,k}]\} \left\{ \frac{N_i(t_{i,k}) - N_i(t_{i,k-1})}{\Lambda(t_{i,k}) - \Lambda(t_{i,k-1})} - 1 \right\} = 0 \quad (j = 1, \ldots, J), \tag{5}$$

satisfies (4), but its components are not clearly nonnegative. Thus, Wellner & Zhang (2000) apply the iterative convex minorant algorithm to (3) to obtain $\hat{\Lambda}_{\mathrm{WZ}}(\cdot)$.

Define $\Delta N_i(s_j) = N_i(s_j) - N_i(s_{j-1})$, $t_{i,K_i} = C_i$ and $Y_i(t) = I(t \leqslant C_i)$. When the response observations are prescheduled to take place at the same fixed set of times s_j for all individuals with monotone pattern of missing data, in the terminology of Robins & Rotnitzky (1995), (5) reduces to the estimating equations given by the score function

$$\sum_{i=1}^{n} Y_i(s_j)\{\Delta N_i(s_j) - \lambda_j\} = 0 \quad (j = 1, \ldots, J), \tag{6}$$

whose solution has nonnegative components. When the observation times become dense, the resulting estimator

$$\hat{\Lambda}(t) = \sum_{j:s_j \leqslant t} \hat{\lambda}_j = \sum_{j:s_j \leqslant t} \sum_{i=1}^{n} \frac{Y_i(s_j)\Delta N_i(s_j)}{\sum_{l=1}^{n} Y_l(s_j)} \quad (t > 0),$$

approaches the Nelson–Aalen estimator with right-censored Poisson counts, where $\hat{\lambda}_j$ are the solutions to the estimating equations (6). The estimator cannot be evaluated with panel counts when individuals have different sets of observation times. We propose a new set of estimating equations using the structure of (6).

Let $\tilde{\Delta} N_i(s_j) = N_i\{r_i(s_j)\} - N_i\{l_i(s_j)\}$, and $\tilde{\Delta}\Lambda_i(s_j) = \Lambda_0\{r_i(s_j)\} - \Lambda_0\{l_i(s_j)\}$; that is, $\tilde{\Delta}\Lambda_i(s_j) = \sum_{k:l_i(s_j)<s_k\leqslant r_i(s_j)} \lambda_k$, where $r_i(s_j)$ and $l_i(s_j)$ are the most recent observation times of individual i not before and before s_j, respectively. Note that $r_i(s_j) = \min\{t_{i,k} : k = 1, \ldots, K_i; t_{i,k} \geqslant s_j\}$ and $l_i(s_j) = \max\{t_{i,k} : k = 1, \ldots, K_i; t_{i,k} < s_j\}$. Here $r_i(s_j) = s_{J+1} = \infty$ if $s_j > t_{i,K_i}$, and $l_i(s_j) = s_0 = 0$ if $s_j < t_{i,1}$. We consider the nonparametric estimator of $\Lambda_0(\cdot)$ in the form of (2), where $\{\hat{\lambda}_j : j = 1, \ldots, J\}$ is the nontrivial solution of the estimating equations

$$\sum_{i=1}^{n} Y_i(s_j)\{\hat{\Delta} N_i(s_j) - \lambda_j\} = 0 \quad (j = 1, \ldots, J), \tag{7}$$

with $\hat{\Delta} N_i(s_j) = \lambda_j \tilde{\Delta} N_i(s_j)/\tilde{\Delta}\Lambda_i(s_j)$. The estimating equations in (7) are unbiased, since the expectation of $\hat{\Delta} N_i(s_j)$, conditional on the observation mechanism, is λ_j regardless of the underlying probability model of the counting process. Moreover, since

$$E\{\Delta N_i(s_j) \mid \text{data of individual } i\} = \lambda_j \frac{\tilde{\Delta} N_i(s_j)}{\tilde{\Delta}\Lambda_i(s_j)} = \hat{\Delta} N_i(s_j)$$

under the Poisson assumption, the estimating functions in (7) can be viewed as the expectation of the likelihood estimating functions in (6) conditional on the panel counts.

The resulting estimator $\hat{\Lambda}(\cdot)$ from (7) is then the solution of the equations

$$\Lambda(t) = \sum_{j:s_j \leqslant t} \sum_{i=1}^{n} \frac{Y_i(s_j)}{\sum_{l=1}^{n} Y_l(s_j)} \tilde{\Delta} N_i(s_j) \frac{\lambda_j}{\tilde{\Delta}\Lambda_i(s_j)} \quad (t > 0). \tag{8}$$

A self-consistent algorithm for obtaining $\hat{\Lambda}(\cdot)$ is as follows. Given the estimate $\Lambda^{(\nu)}(\cdot)$ of $\Lambda_0(\cdot)$ at the νth iteration, the left-hand side of (8) gives a new estimate $\Lambda^{(\nu+1)}(\cdot)$ with $\Lambda^{(\nu)}(\cdot)$ inserted into the right-hand side of (8). The limit of the sequence $\{\Lambda^{(\nu)}(\cdot) : \nu = 1, \ldots\}$ is then $\hat{\Lambda}(\cdot)$, provided that all the entries of the initial estimate $\lambda^{(0)}$ associated with $\Lambda^{(0)}(\cdot)$ are positive; see the Appendix for the convergence of this algorithm.

Any sample path of $\hat{\Lambda}(\cdot)$ obtained by the algorithm is nondecreasing if the iterative procedure starts with a nondecreasing initial estimate $\Lambda^{(0)}(\cdot)$. This and the fact that (5) is equivalent to (7) indicate that the estimator $\hat{\Lambda}(\cdot)$ is the same as $\hat{\Lambda}_{WZ}(\cdot)$; see the Appendix. Obtaining $\hat{\Lambda}(\cdot)$ through (7) provides a simpler-to-implement alternative approach to obtaining the nonparametric maximum likelihood estimator. It can be viewed as a nonparametric generalization of the expectation-solution algorithm proposed by Rosen et al. (2000), not requiring a specification of the underlying probability model for the counting process.

To assess the finite-sample properties of $\hat{\Lambda}(\cdot)$, we ran a simulation study using an underlying Poisson process with the intensity function $\lambda_0(t) = 6\gamma t^{\gamma-1}$ $(0 < t \leqslant 1)$ to generate time-homogeneous, $\gamma = 1$, or time-nonhomogeneous, $\gamma = 1/2$ or 2, Poisson processes. We simulated studies with $n = 100$, a follow-up period of [0, 1] and two different observation processes: a time-homogeneous Poisson process, and a process with events independently occurring at times $0{\cdot}05, 0{\cdot}10, \ldots, 1{\cdot}00$ with pr(an event occurrence at t) $= 1{\cdot}02 - t^{1/4}$. We intended to simulate studies where the observation times vary among individuals by the first scheme, or where the observation times are prescheduled and the data-missing rate is increasing as the study progresses through the second scheme. The parameters for each observation scheme were chosen to obtain four observations on average per individual. We refer to the two observation processes as 'Poisson' and 'Bernoulli', respectively.

The computer programs for this and the subsequent simulation studies were written in C, using the generators of Weibull, Poisson, gamma, normal and beta random variables in Splus to generate random variables. All the generated times in the simulation were rounded to two digits after the decimal point. Iterations were terminated when the sup-norm of the difference between successive estimates of $\Lambda_0(\cdot)$ became less than 10^{-5}.

Each of the six experimental settings was simulated 200 times. Our new procedure and that of Wellner & Zhang yielded almost the same realization of $\hat{\Lambda}(\cdot)$ for each dataset. Our new procedure was faster than Wellner & Zhang's, and it had efficiency close to that of the Nelson–Aalen estimator based on a continuously observed process subject to right censoring. The detailed simulation results are available upon request.

In the next section, we introduce a different procedure for estimating the mean function $\Lambda_0(\cdot)$.

3. QUASI-SCORE-BASED PROCEDURE

3·1. *Estimation procedure*

The quasi-score function for $\lambda = (\lambda_1, \ldots, \lambda_J)'$ based on observations of $\Delta N(s_j) = N(s_j) - N(s_{j-1})$ with a monotone pattern of missing data is

$$U(\lambda) = \sum_{i=1}^{n} Y_i' \operatorname{var}(Y_i \Delta N_i)^{-} Y_i(\Delta N_i - \lambda), \tag{9}$$

where Y_i is the diagonal matrix with the diagonal elements $\{Y_i(s_1), \ldots, Y_i(s_J)\}$, ΔN_i is the J-dimensional vector with components $\Delta N_i(s_j)$ $(j = 1, \ldots, J)$ and $\text{var}(Y_i \Delta N_i)^-$ is the Moore–Penrose generalized inverse of the matrix $\text{var}(Y_i \Delta N_i)$. When $N(\cdot)$ is a Poisson process, (9) gives the likelihood estimating equations (6).

The estimating function $U(\lambda)$ in (9) cannot be evaluated with the panel counts, since not all $\Delta N_i(s_j)$ are available. To overcome this, we substitute an estimator of ΔN_i based on the data and obtain an estimating function of λ; that is,

$$\tilde{U}(\lambda) = \sum_{i=1}^{n} Y_i' \text{var}(Y_i \Delta N_i)^- Y_i(\hat{\Delta} N_i - \lambda), \tag{10}$$

where the jth component of $\hat{\Delta} N_i$ is $\hat{\Delta} N_i(s_j) = \lambda_j \tilde{\Delta} N_i(s_j)/\tilde{\Delta}\Lambda_i(s_j)$. The estimating function $\tilde{U}(\lambda)$ in (10) is unbiased, since, as indicated in § 2, the expectation of the jth component of $\hat{\Delta} N_i$ is λ_j for all j such that $s_j \leqslant t_{i,K_i} = C_i$, regardless of the underlying probability model of the counting process. Furthermore, $\tilde{U}(\lambda) = 0$ is equivalent to

$$\lambda = \left\{ \sum_{i=1}^{n} Y_i' \text{var}(Y_i \Delta N_i)^- Y_i \right\}^{-1} \sum_{i=1}^{n} Y_i' \text{var}(Y_i \Delta N_i)^- Y_i \hat{\Delta} N_i. \tag{11}$$

Similar to (8), (11) suggests a self-consistent procedure for obtaining $\hat{\lambda}_j$ in (2). We denote the resulting estimator by $\hat{\Lambda}_2(\cdot)$.

It is easy to verify that (10) reduces to (6) of § 2 when $\text{var}(\Delta N_i)$ is diagonal. For Poisson processes, the increments are independent, and $\text{var}(Y_i \Delta N_i)$ in (9) is a diagonal matrix with diagonal elements λ_j for $s_j \leqslant C_i$ and 0 for other j. In other settings, such as when $N(\cdot)$ is a one-jump process, $\text{var}(Y_i \Delta N_i)$ is not diagonal, and (10) differs from (6) of § 2.

There are situations, such as one-jump processes, where $\text{var}(Y_i \Delta N_i)$ is known or assumed known up to λ. The self-consistent procedure based on (11) can be used to obtain $\hat{\Lambda}_2(\cdot)$ in those situations. When $\text{var}(Y_i \Delta N_i)$ involves parameters, θ say, other than λ, we need to consider (10) coupled with further estimating equations. For example, we may extend the ideas in Breslow (1990) and consider the quadratic estimating equation

$$\begin{aligned} 0 = \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{l=1}^{J} W_{ijl} Y_i(s_j) Y_i(s_l) [\{\tilde{\Delta} N_i(s_j) - \tilde{\Delta}\Lambda_i(s_j)\}\{\tilde{\Delta} N_i(s_l) - \tilde{\Delta}\Lambda_i(s_l)\} \\ - \text{cov}\{\tilde{\Delta} N_i(s_j), \tilde{\Delta} N_i(s_l)\}], \end{aligned} \tag{12}$$

where the covariance is conditional on the observation mechanism, and W_{ijl} are the weights.

The components of the solution of (11) are not necessarily all nonnegative; thus, some sample paths of $\hat{\Lambda}_2(\cdot)$ might not be monotone. We could then consider its isotonic regression with an appropriate set of weights, such as $n_j = \sum_{i=1}^{n} Y_i(s_j)$ $(j = 1, \ldots, J)$ with n_j the number of individuals in the study at time s_j or $n_j^* = \sum_{i=1}^{n} I[s_j \in \{t_{i,k} : k = 1, \ldots, K_i\}]$ with n_j^* the number of total observations at time s_j in the study. This gives an estimator close to the generalized least-squares estimator of Hu et al. (2009) with the corresponding weight, and avoids the intensive computation involved in their approach.

3·2. *Mixed-Poisson panel counts*

Suppose the intensity function of $N_i(\cdot)$, given nonnegative random effect α_i, is $\alpha_i \gamma(t)$ $(i = 1, \ldots, n)$, where α_i has mean 1 and variance θ. Then $E\{N(t)\} = \Lambda_0(t) = \int_0^t \gamma(s) ds$, and $\text{var}\{N(t)\} = \Lambda_0(t)\{1 + \theta\Lambda_0(t)\}$. Conditional on the observation mechanism,

$\mathrm{var}(Y_i\Delta N_i)^- = \mathrm{diag}(Y_i\lambda)^- - \theta Y_i 1_J 1_J' Y_i'/\{1+\theta\Lambda(C_i)\}$, where 1_J is the J-dimensional vector with all components equal to 1. The components in (10) are then

$$\tilde{U}(\lambda;\theta)_j = \sum_{i=1}^{n} Y_i(s_j)\left\{\frac{\tilde{\Delta}N_i(s_j)}{\tilde{\Delta}\Lambda_i(s_j)} - 1\right\} - \sum_{i=1}^{n} Y_i(s_j)\frac{\theta\{N_i(C_i)-\Lambda(C_i)\}}{1+\theta\Lambda(C_i)} \quad (j=1,\ldots,J). \tag{13}$$

The estimating functions in (13) lead to the equations

$$\lambda_j = \left\{\sum_{i=1}^{n} Y_i(s_j)\frac{1+\theta N_i(C_i)}{1+\theta\Lambda(C_i)}\right\}^{-1} \sum_{i=1}^{n} Y_i(s_j)\lambda_j \frac{\tilde{\Delta}N_i(s_j)}{\tilde{\Delta}\Lambda_i(s_j)} \quad (j=1,\ldots,J), \tag{14}$$

which suggest a self-consistent algorithm for computing the estimator $\hat{\Lambda}_2(\cdot)$ with fixed θ. The algorithm gives a nondecreasing estimate of $\Lambda_0(\cdot)$ if the initial estimate is nondecreasing.

Lawless (1987) presents a mixed-Poisson process with α_i generated from the Gamma density $g(\alpha;\theta) = \alpha^{\phi-1}e^{-\alpha/\theta}\theta^{-\phi}\Gamma(\phi)^{-1}$ ($\alpha > 0$), with $\phi = 1/\theta$. The corresponding loglikelihood function based on the panel counts is

$$\log L(\Lambda;\theta) = \sum_{i=1}^{n}\left[\sum_{k=1}^{K_i}\{N_i(t_{i,k}) - N_i(t_{i,k-1})\}\log\left(\sum_{j:t_{i,k-1}<s_j\leqslant t_{i,k}} \lambda_j\right) - \left\{N_i(C_i)+\frac{1}{\theta}\right\}\log\{1+\theta\Lambda(C_i)\} + \sum_{k=0}^{N_i(C_i)-1}\log(1+k\theta)\right]. \tag{15}$$

With the Gamma-distributed random effects, $\tilde{U}(\lambda;\theta)$ in (13) is the gradient of $\log L(\Lambda;\theta)$ in (15) with respect to λ with fixed θ. Therefore, $\hat{\Lambda}_2(\cdot)$ is the nonparametric maximum likelihood estimator derived from (15) with known θ and subject to the monotonicity constraint. Extending the arguments in Wellner & Zhang (2000), we may show that $\hat{\Lambda}_2(\cdot)$ is consistent without the Gamma-Poisson assumption. Similar arguments to those in the Appendix can be used to prove the convergence of the self-consistent algorithm with the additional moment assumption $\mathrm{var}\{N(t)\} = \Lambda_0(t)\{1+\theta\Lambda_0(t)\}$.

When the missing-data pattern is monotone, $\tilde{U}(\lambda;\theta)$ in (10) is the same as $U(\lambda;\theta)$ in (9). The solution of $U(\lambda;\theta) = 0$, denoted by $\hat{\Lambda}_{\mathrm{NL}}(\cdot)$, may be viewed as a nonparametric version of the estimator given in Nadeau & Lawless (1998). Here $\hat{\Lambda}_{\mathrm{NL}}(\cdot)$ satisfies

$$\Lambda(t) = \sum_{s_j\leqslant t} \frac{\sum_{i=1}^{n} Y_i(s_j)\Delta N_i(s_j)}{\sum_{i=1}^{n} Y_i(s_j)\{1+\theta N_i(C_i)\}/\{1+\theta\Lambda(C_i)\}},$$

and is a generalization of the Nelson–Aalen estimator, weighting more or less individuals with the final counts $N_i(C_i)$ larger than the average in the denominator when θ is positive or negative, respectively.

In practice, θ is usually unknown. We could maximize the loglikelihood function $\log L(\Lambda;\theta)$ in (15) with respect to both the nuisance parameter θ and $\Lambda(\cdot)$; this can be computationally difficult. An extension of the EM algorithm to frailty models (Andersen et al., 1992, Ch. IX) may be considered. Zhang & Jamshidian (2003) apply the approach to computing their extension of the maximum pseudolikelihood estimator of Wellner & Zhang (2000) under the Gamma-Poisson assumption. The procedure discussed in § 3·1, which uses the estimating equations in (14) combined with a quadratic estimating equation, can be a preferred alternative in practice. To be specific, the following estimating function, proposed for the situations with right-censored observations, could be used to supplement $\tilde{U}(\lambda;\theta)$ in (13):

$$U_{\rm D}(\theta;\lambda)=\sum_{i=1}^{n}\frac{\{N_i(C_i)-\Lambda(C_i)\}^2-\mathrm{var}\{N_i(C_i)\}}{\{1+\theta\Lambda(C_i)\}^2}.$$

The estimating function $U_{\rm D}(\theta;\lambda)$, advocated by Dean (1991), is (12) with a set of weights suggested by Breslow (1990). A different supplementary estimating function

$$U_{\rm NL}(\theta;\lambda)=U_{\rm D}(\theta;\lambda)-\sum_{i=1}^{n}\frac{\{1+2\theta\Lambda(C_i)\}\{N_i(C_i)-\Lambda(C_i)\}}{\{1+\theta\Lambda(C_i)\}^2}$$

is the optimal quadratic estimating function derived by Nadeau & Lawless (1998). As pointed out by them, using $U_{\rm D}(\theta;\lambda)$ or $U_{\rm NL}(\theta;\lambda)$ gives similar results.

3·3. *Simulation studies*

Two simulation studies were conducted to examine efficiency and robustness of $\hat{\Lambda}_2(\cdot)$, compared to $\hat{\Lambda}(\cdot)$, which is the same as $\hat{\Lambda}_{\rm WZ}(\cdot)$, the Nelson–Aalen estimator based on right-censored count data, and the nonparametric version of the Nadeau–Lawless estimator $\hat{\Lambda}_{\rm NL}(\cdot)$.

First we study efficiency. We use $n=100$ and the observation processes considered in the previous simulation. The counting process $N_i(\cdot)$ was generated as a mixed-Poisson process with rate $\lambda=6\alpha_i$, where α_i was generated from the Gamma distribution with density $g(\alpha;\theta)=\alpha^{\phi-1}e^{-\alpha/\theta}\theta^{-\phi}\Gamma(\phi)^{-1}$, with $\phi=1/\theta$ and the shape parameter $\theta=0{\cdot}5$, $1{\cdot}0$ or $5{\cdot}0$. Coupled with the two observation processes, there are six experimental settings, each with mean function $\Lambda_0(t)=6t$ and nonindependent increments. The correlations between response increments increase with θ. In each simulation setting, we evaluated $\hat{\Lambda}(\cdot)$ and $\hat{\Lambda}_2(\cdot)$ using (2) with the component estimates obtained by the procedures presented in § 2 and § 3, respectively. The estimates associated with $\hat{\Lambda}_2(\cdot)$ were obtained from (11) coupled with the estimating function $U_{\rm D}(\theta;\lambda)$ given in § 3·2. For comparison, we also computed $\hat{\Lambda}_{\rm NA}(\cdot)$ and $\hat{\Lambda}_{\rm NL}(\cdot)$ based on a right-censored observation process in which each individual is censored at his largest observation time.

All estimators are essentially unbiased, which also confirms that $\hat{\Lambda}(\cdot)$, a special case of $\hat{\Lambda}_2(\cdot)$ under the Poisson assumption, is consistent without the Poisson assumption. Figure 1 presents the sample mean squared errors of the estimators, and indicates a clear trend of $\hat{\Lambda}_2(\cdot)$ performing better than $\hat{\Lambda}(\cdot)$, as θ increases. The results indicate that there is no apparent advantage of the estimators based on right-censored data over those based on the panel data when observation times are not very sparse. In fact, when $\theta=5{\cdot}0$, it appears that $\hat{\Lambda}_2(\cdot)$ has higher efficiency than the Nelson–Aalen estimator, which is probably due to the use of neighbourhood information by the $\hat{\Lambda}(\cdot)$ and $\hat{\Lambda}_2(\cdot)$ estimators.

The estimators of θ from the quadratic estimating function $U_{\rm D}(\theta;\lambda)$ associated with $\hat{\Lambda}_2(\cdot)$ and $\hat{\Lambda}_{\rm NL}(\cdot)$ performed well when the random effect is light. While their sample standard errors were sometimes large, the corresponding estimators $\hat{\Lambda}_2(\cdot)$ and $\hat{\Lambda}_{\rm NL}(\cdot)$, using the θ estimators, performed quite well in all the simulations, echoing the comment by Nadeau & Lawless (1998) about the robustness of their estimation procedure for the primary parameter to different estimates of θ.

We conducted another simulation study to assess the robustness of $\hat{\Lambda}_2(\cdot)$ when the response does not follow the mixed-Poisson model with Gamma-distributed random effect. We simulated the same settings as before, except that $N(\cdot)$ was generated as one of the following three counting processes, given a total number of events L over the time period (0, 1]: (a) the process with L independent Un(0, 1) event times, where L is Poisson with mean 6; (b) the process with L

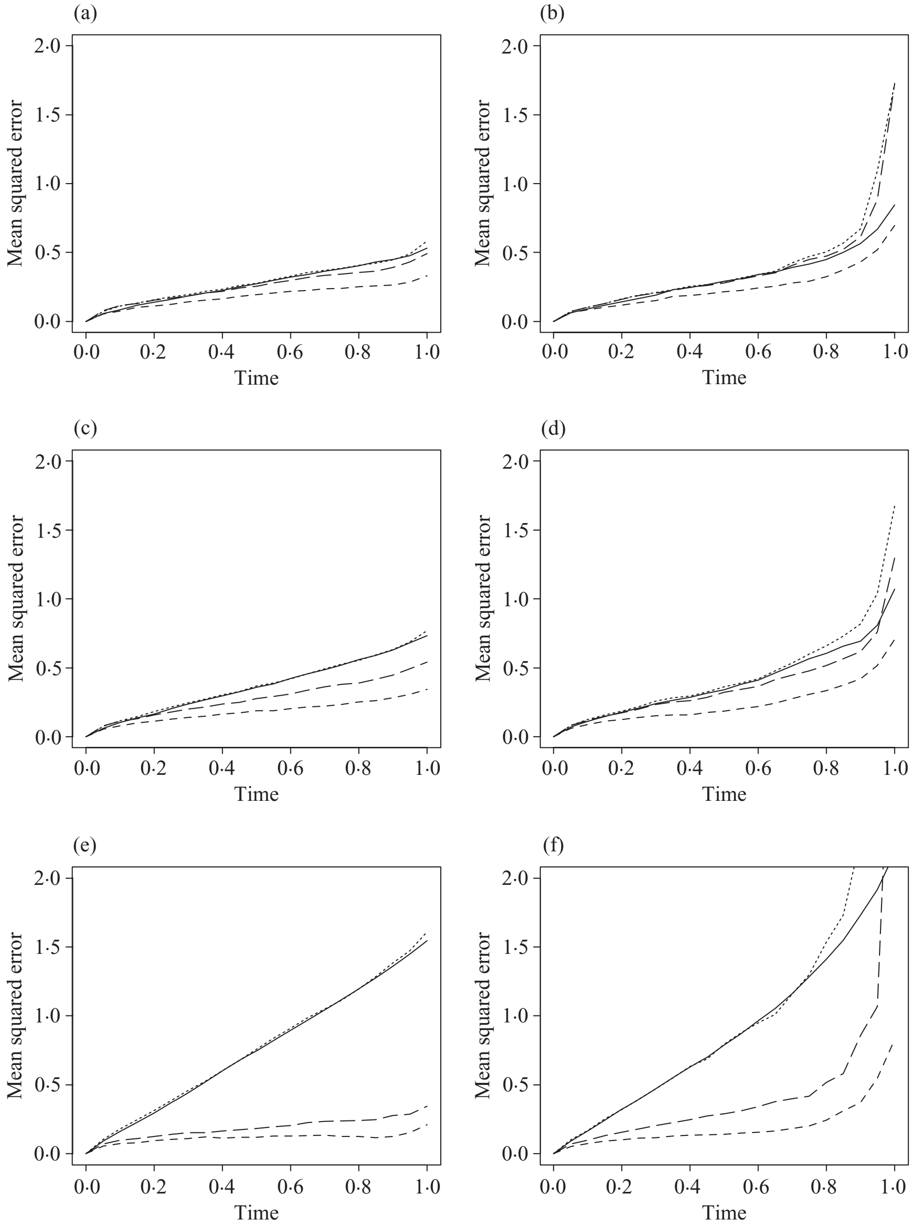

Fig. 1. Evaluations of estimators in the simulation study of efficiency, for the Gamma-Poisson process with (a)–(b) $\theta = 0{\cdot}5$, (c)–(d) $\theta = 1$ and (e)–(f) $\theta = 5$, and observation mechanisms (a), (c), (e) Poisson and (b), (d), (f) Bernoulli. The lines represent the Nelson–Aalen estimates (solid), Wellner & Zhang's $\hat{\Lambda}(\cdot)$ estimates (dotted), Nadeau–Lawless estimates (small dashed) and our $\hat{\Lambda}_2(\cdot)$ estimates (dashed).

independent Be(0·5, 0·5) event times, where L is Poisson with mean 6α and α is lognormal with mean 1 and variance 3; (c) the process with L_1 independent event Un(0, 0·5) event times and L_2 independent Un(0·5, 1) event times, where $L_1 + L_2 = L$, L_1 and L_2 are independently Poisson with means $6\alpha_1$ and $6\alpha_2$, and $(\log\alpha_1, \log\alpha_2)$ is normal with $\mathrm{cov}(\log\alpha_1, \log\alpha_2) = 1$, and $E(\alpha_1) = E(\alpha_2) = 1$, $\mathrm{var}(\alpha_1) = 2$, $\mathrm{var}(\alpha_2) = 4$.

The sample means of the Nelson–Aalen, $\hat{\Lambda}(\cdot)$, Nadeau–Lawless and $\hat{\Lambda}_2(\cdot)$ estimators were very close to the true mean functions, confirming their consistency. The sample mean squared errors of the estimators indicate the following conclusions. The Nelson–Aalen method performs best with the Poisson response, while the other estimators perform well and similarly to each other. The relative performances of the four estimators with the processes of type (b) are similar to what we observe from Fig. 1, where the random effect is Gamma distributed, and Nadeau–Lawless is the maximum likelihood estimator with right-censored counts and $\hat{\Lambda}_2(\cdot)$ is close to the maximum likelihood estimator with panel counts. With the processes of type (c), the Nelson–Aalen and $\hat{\Lambda}(\cdot)$ estimators perform considerably worse than the $\hat{\Lambda}_2(\cdot)$ and Nadeau–Lawless estimators. The detailed results are available from the authors.

4. Example

We illustrate the proposed methods using results from the bladder cancer study reported by Byar (1980), and analyzed by Wei et al. (1989), Lawless & Zhan (1998), Wellner & Zhang (2000), Jin et al. (2006) and others. We use the version of the data given by Hu et al. (2003), focusing on the placebo and thiotepa study groups, with respective sample sizes of 47 and 38.

We evaluated the $\hat{\Lambda}(\cdot)$ estimator, the $\hat{\Lambda}_2(\cdot)$ estimator and Sun & Kalbfleisch's estimator with the panel counts from the two groups. The $\hat{\Lambda}(\cdot)$ and Sun & Kalbfleisch's estimates were in close agreement with those presented by Wellner & Zhang (2000), but the $\hat{\Lambda}_2(\cdot)$ estimates were quite different. This is analogous to the big difference shown by Zhang & Jamshidian (2003) between the evaluations of Sun & Kalbfleisch's estimator, i.e. the nonparametric maximum pseudolikelihood estimator, and the extension proposed by Zhang & Jamshidian to address intracorrelation between the panel counts of a counting process. The estimates of the mixed-effect parameter θ associated with the $\hat{\Lambda}_2(\cdot)$ estimates were $2{\cdot}705$ and $6{\cdot}570$ for the placebo and thiotepa groups, respectively. We obtained the estimates for the mean function of the cumulative visit numbers in each group, based on the Nelson–Aalen estimator and the Nadeau–Lawless estimator, which also differ. The corresponding estimates of θ, the degree of the random effect, associated with the Nadeau–Lawless estimator were $7{\cdot}661$ and $4{\cdot}691$ in the two groups, respectively. We suspect that the discrepancies between the $\hat{\Lambda}(\cdot)$ and $\hat{\Lambda}_2(\cdot)$ estimates, and between the Nelson–Aalen and Nadeau–Lawless estimates, resulted from the nonhomogeneous responses over time within each individual and among the individuals in each group. As pointed out by a referee, the difference could also be due to the possibly dependent loss to follow-up, which may bias the Sun & Kalbfleisch, $\hat{\Lambda}(\cdot)$ and Nelson–Aalen estimates more than the $\hat{\Lambda}_2(\cdot)$ and Nadeau–Lawless estimates.

To estimate standard errors for the estimators, we evaluated each of the estimators mentioned above on 100 bootstrap samples and obtained the bootstrap variance estimates, from which we constructed pointwise interval estimates with the upper and lower and limits $\hat{\Lambda}(t) \pm 1{\cdot}96\,\mathrm{SE}\{\hat{\Lambda}(t)\}$, $t \in (0, 48)$, respectively, for the mean functions of the cumulative numbers of new tumours in the two groups; see Fig. 2. We compared the limits with the $2{\cdot}5\%$ and $97{\cdot}5\%$ percentiles of the bootstrap realizations, i.e. the bootstrap percentile intervals, and found that they were quite similar. The interval estimates associated with Sun & Kalbfleisch's, the $\hat{\Lambda}(\cdot)$ and the $\hat{\Lambda}_2(\cdot)$ estimators indicate some difference between the placebo and thiotepa groups. The $\hat{\Lambda}_2(\cdot)$ estimates in Fig. 2(c) clearly suggest a significant difference between the two treatment groups over time, while the other two sets in Fig. 2(a) and (b) do not. The simulation results in § 3 show relatively lower variation associated with $\hat{\Lambda}_2(\cdot)$, especially when the observations are not too sparse, compared to Sun & Kalbfleisch's and Wellner & Zhang's estimators; this may explain the discrepancy in the results.

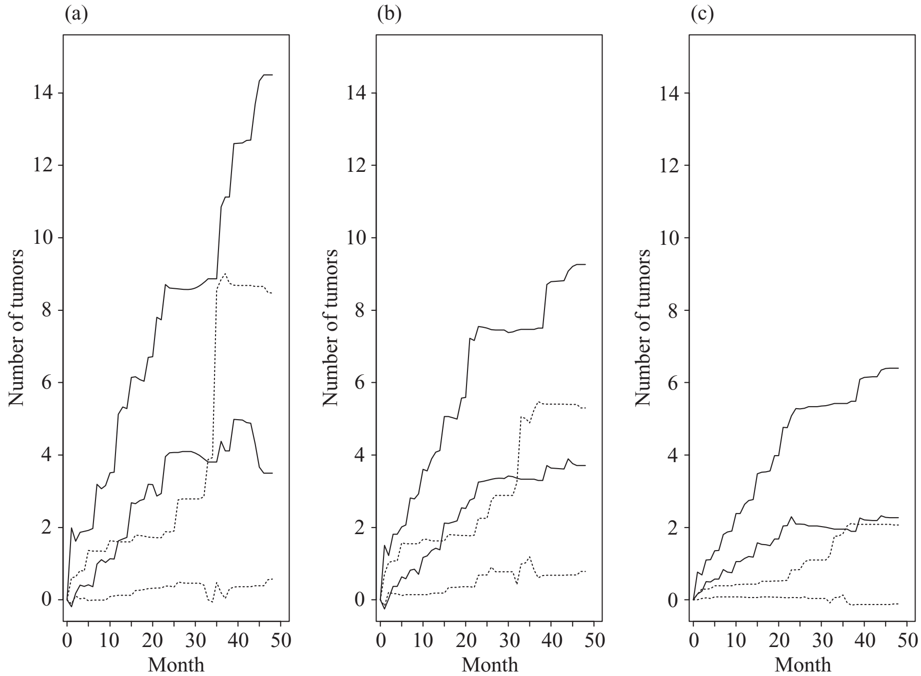

Fig. 2. Tumour data. Pointwise 95% interval estimates for number of tumours, for placebo (solid) and thiotepa (dotted) treatment groups. (a) Sun & Kalbfleisch's estimates, (b) Wellner & Zhang's $\hat{\Lambda}(\cdot)$ estimates and (c) our $\hat{\Lambda}_2(\cdot)$ estimates.

The bootstrap standard errors, associated with the estimates of the mixed-effect parameter θ for the placebo and thiotepa groups, are 4·848 and 7·500 with $\hat{\Lambda}_2(\cdot)$ from the tumour response data and 5·758 and 5·381 with Nadeau–Lawless from the clinic-visit data, respectively. This indicates considerable overdispersion in the responses, and may partly explain the discrepancy among the three estimates.

5. Discussion

Our empirical studies confirm the convergence of the self-consistent algorithm in § 2, and indicate convergence of the algorithm in § 3. The corresponding results associated with the EM algorithm suggest that the algorithm in § 3 converges under the mixed-Poisson assumption with Gamma-distributed random effect. Proof of the convergence of the algorithm in general remains as a theoretical challenge, as does proof of consistency and derivation of weak convergence of the resulting estimator $\hat{\Lambda}_2(\cdot)$ with the nuisance parameter θ.

In some situations, the second set of estimating functions involves nuisance parameters, which require the use of supplementary estimating functions. The method in § 3·2, of combining the second estimating function set with another estimating function employed in Nadeau & Lawless (1998), gave satisfactory estimates for the mean function based on both the $\hat{\Lambda}_2(\cdot)$ and the Nadeau–Lawless estimators in the simulated settings. However, our simulation and the example in § 4 indicate the need for further investigation to improve the estimation in § 3·2 for the nuisance parameter θ. Robustness and asymptotic properties of the resulting estimator $\hat{\Lambda}_2(\cdot)$ with different estimates of the nuisance parameters can also be investigated further.

The approaches developed in this paper are readily extendible to situations with a general nondecreasing response process, or where there are covariates and a semiparametric model is assumed for the expectation of the counting process conditional on the covariates. The latter allows for response-dependent observations, provided that the response and the observation mechanism are independently conditional on the covariates. It would be worthwhile to explore how to deal with panel counts from different informative observation mechanisms. As suggested by a referee, the inverse probability weighting may be adopted in some of the situations.

Acknowledgement

The research was partially supported by the U.S. National Institute of Allergy and Infectious Diseases and the Natural Sciences and Engineering Research Council of Canada. We thank Professor J. F. Lawless for helpful discussions, and two referees and Professor D. M. Titterington for their constructive comments.

Appendix

Some theoretical details for the algorithm in § 2

Equivalence to the EM *algorithm under the Poisson assumption*. The estimating equations (6) are derived from the loglikelihood function, conditional on the observation times, based on the Poisson panel counts with monotone pattern of missing data $\ell(\lambda) = \sum_{i=1}^{n} \sum_{l=1}^{J} Y_i(s_l)\{\Delta N_i(s_l)\log(\lambda_l) - \lambda_l\}$.
Denote the actual observed data, i.e. the panel counts in (1), by $\mathcal{O}$. Under the Poisson assumption,

$$Q(\lambda \mid \lambda^*) \equiv E_{\lambda^*}\{\ell(\lambda) \mid \mathcal{O}\} = \sum_{i=1}^{n}\sum_{l=1}^{J} Y_i(s_l)\left\{\frac{\tilde{\Delta} N_i(s_l)}{\tilde{\Delta}\Lambda_i^*(s_l)}\lambda_l^* \log(\lambda_l) - \lambda_l\right\},$$

and thus the components of the gradient of $Q(\lambda \mid \lambda^*)$ with respect to λ are

$$\sum_{i=1}^{n} Y_i(s_l)\left\{\frac{\tilde{\Delta} N_i(s_l)}{\tilde{\Delta}\Lambda_i^*(s_l)}\frac{\lambda_l^*}{\lambda_l} - 1\right\} \quad (l = 1, \ldots, J).$$

Setting the above functions equal to 0 gives the update step of our algorithm. Note that $Q(\lambda \mid \lambda^*)$ is concave as a function of its first argument and strictly so unless there is an s_l such that $\sum_i Y_i(s_l)\tilde{\Delta} N_i(s_l) = 0$. In either case, it is easily seen that at each step our algorithm maximizes $Q(\lambda \mid \lambda^*)$ with respect to λ. Thus, our algorithm is the EM algorithm applied to the Poisson panel counts.

Global convergence of the algorithm. Since convergence of an algorithm for a given dataset does not depend on the mechanism which generated the data, we may apply general results for convergence of the EM algorithm without assuming that the true underlying counting processes are Poisson.

Denote the loglikelihood based on the panel counts $\mathcal{O}$ under the Poisson assumption by $\ell_O(\lambda) = \log L_{\text{WZ}}(\lambda)$, where $L_{\text{WZ}}(\lambda)$ is given in (3). Let $\lambda^{(\nu)}$ denote the value of λ after ν iterates of the algorithm. It follows from Lemma 1 and Theorem 1 of Dempster et al. (1977) that our algorithm has the property $\ell_O(\lambda^{(\nu+1)}) - \ell_O(\lambda^{(\nu)}) \geqslant 0$ at each stage ν. Furthermore, the second derivative matrix of ℓ_O has l_1, l_2 entry given by

$$-\sum_{i=1}^{n}\sum_{k=1}^{K_i} \frac{\{N(t_{i,k}) - N(t_{i,k-1})\}}{\left(\sum_{l:s_l\in(t_{i,k-1},t_{i,k}]} \lambda_l\right)^2} I_{ikl_1} I_{ikl_2},$$

where $I_{ikl} = I(t_{i,k-1} < s_l \leqslant t_{i,k})$. It follows that ℓ_O is strictly concave provided that the matrix M with entries $m_{l_1,l_2} = \sum_{i=1}^{n} \sum_{k=1}^{K_i} \{N(t_{i,k}) - N(t_{i,k-1})\} I_{ikl_1} I_{ikl_2}$ is positive definite.

The following proposition establishes that the algorithm in § 2 converges from any starting point under a reasonable condition on the available data.

Proposition A1. *Assume that the matrix M is positive definite.*

(i) *The function ℓ_O has a unique maximum at some parameter vector $\hat{\lambda}$ which must be a fixed point of our algorithm.*

(ii) *For each subset* L *of* $\{1, \ldots, n\}$*, let* $\Theta_L = \{\lambda : \lambda_l = 0, l \notin L\}$*. Then* ℓ_O *has a unique maximum over* Θ_L *at* $\hat{\lambda}_L$*, say, which also will be a fixed point of our algorithm.*
(iii) *Starting from any parameter vector* $\lambda^{(0)}$ *such that* $\ell_O(\lambda^{(0)}) > -\infty$*, the sequence of iterates* $\lambda^{(\nu)}$ *converges as* $\nu \to \infty$ *to some* $\hat{\lambda}_L$*.*
(iv) *If all entries of* $\lambda^{(0)}$ *are positive then the sequence of iterates converges to* $\hat{\lambda}$*.*

This proposition can be proved using arguments similar to those in Wu (1983) for the convergence of the EM algorithm. A detailed proof is available upon request.

References

Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N. (1992). *Statistical Models Based on Counting Processes*. New York: Springer.

Breslow, N. (1990). Regression and other quasi-likelihood models. *J. Am. Statist. Assoc.* **85**, 565–71.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: Comparison of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology*, Ed. M. Pavone-Macaluso, P. H. Smith and F. Edsmyn, pp. 363–70. New York: Plenum.

Chen, B. E., Cook, R. J., Lawless, J. F. & Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statist. Med.* **24**, 671–91.

Dean, C. B. (1991). Estimating functions for mixed Poisson models. In *Estimating Functions*, Ed. V. P. Godambe, pp. 35–46. Oxford: Clarendon.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *J. R. Statist. Soc.* B **39**, 1–38.

Hu, X. J., Lagakos, S. W. & Lockhart, R. A. (2009). Generalized least squares estimation of the mean function of a counting process based on panel counts. *Statist. Sinica*. **19**, 561–80.

Hu, X. J., Sun, J. & Wei, L. J. (2003). Regression parameter estimation from panel counts. *Scand. J. Statist.* **30**, 25–43.

Jin, Z., Lin, D. Y. & Ying, Z. (2006). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scand. J. Statist.* **33**, 1–23.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comp. Graph. Statist.* **28**, 161–83.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.* **53**, 457–81.

Lawless, J. F. (1987). Negative binomial regression models. *Can. J. Statist.* **15**, 209–26.

Lawless, J. F. & Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158–68.

Lawless, J. F. & Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Can. J. Statist.* **26**, 549–65.

Lin, D. Y., Wei, L. J., Yang, I. & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Statist. Soc.* B **62**, 711–30.

Nadeau, C. & Lawless, J. F. (1998). Inference for means and covariances of point processes through estimating functions. *Biometrika* **85**, 893–906.

Robins, J. M. & Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Am. Statist. Assoc.* **90**, 122–9.

Rosen, O., Jiang, W. & Tanner, M. A. (2000). Mixtures of marginal models. *Biometrika* **87**, 391–404.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. New York: Springer.

Sun, J. & Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica* **5**, 279–90.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Statist. Soc.* B **38**, 290–5.

Wei, L. J., Lin, D. Y. & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065–73.

Wellner, J. A. & Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779–814.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11**, 95–103.

Zhang, Y. & Jamshidian, M. (2003). The Gamma-frailty Poisson model for the nonparametric estimation of panel count data. *Biometrics* **59**, 1099–106.

[*Received April* 2007. *Revised August* 2008]