Biostatistics (2014), **15**, 3, *pp.* 555–568
doi:10.1093/biostatistics/kxu009
Advance Access publication on March 19, 2014

Regression analysis of mixed recurrent-event and panel-count data

LIANG ZHU

Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

XINWEI TONG

School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China

JIANGUO SUN*

Department of Statistics, University of Missouri, Columbia, MO 65211, USA and School of Mathematics, Jilin University, Changchun 130012, China

sunj@missouri.edu

MAN-HUA CHEN

Department of Statistics, Tamkang University, Tamsui, New Taipei 25137, Taiwan

DEO KUMAR SRIVASTAVA

Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

WENDY LEISENRING

Department of Biostatistics, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

LESLIE L. ROBISON

Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

Summary

In event history studies concerning recurrent events, two types of data have been extensively discussed. One is recurrent-event data (Cook and Lawless, 2007. *The Analysis of Recurrent Event Data*. New York: Springer), and the other is panel-count data (Zhao *and others*, 2010. Nonparametric inference based on panel-count data. *Test* **20**, 1–42). In the former case, all study subjects are monitored continuously; thus, complete information is available for the underlying recurrent-event processes of interest. In the latter case, study subjects are monitored periodically; thus, only incomplete information is available for the processes of interest. In reality, however, a third type of data could occur in which some study subjects are monitored continuously, but others are monitored periodically. When this occurs, we have mixed recurrent-event and panel-count data. This paper discusses regression analysis of such mixed data and presents two estimation procedures for the problem. One is a maximum likelihood estimation procedure, and the other

*To whom correspondence should be addressed.

is an estimating equation procedure. The asymptotic properties of both resulting estimators of regression parameters are established. Also, the methods are applied to a set of mixed recurrent-event and panel-count data that arose from a Childhood Cancer Survivor Study and motivated this investigation.

Keywords: Estimating equation-based approach; Maximum likelihood approach; Regression analysis.

1. INTRODUCTION

Event history studies on recurrent events are conducted in many fields including clinical and longitudinal studies, reliability experiments, and sociological studies. Examples of clinical recurrent events include hospitalizations, infections, acute myocardial infarctions, and tumor metastases. For the analysis of these studies, two types of data have been extensively discussed. One is recurrent-event data, and the other is panel-count data. In the former, all study subjects are monitored continuously; thus, complete information is available for the recurrent-event processes of interest. In the latter, study subjects are monitored only periodically; thus, incomplete information is available. More specifically, recurrent-event data provide the time points of all occurrences of the events of interest, and panel-count data provide only the numbers of occurrences of the events between observation times. Panel-count data usually occur when continuous observation is too expensive or impractical.

In reality, a third type of data can occur in which some study subjects are monitored continuously, while others are monitored periodically (Zhu *and others*, 2013). In this case, we have mixed recurrent-event and panel-count data, and such mixed data often occur in medical researches among others. For example, in a long-term follow-up study on hospitalizations, some subjects may remember or provide the dates of all their hospitalizations, while others may remember or provide only the numbers of hospitalizations from time to time. Another situation where mixed data can occur is a chronic disease study in which some patients are observed continuously, while others are monitored intermittently due to their health conditions. In Section 5, we discuss in details a more specific example of mixed data that motivated this study and arose from a Childhood Cancer Survivor Study (CCSS).

Many authors have discussed the analysis of recurrent-event data or panel-count data. For example, Cook and Lawless (2007) gave a relatively thorough review of the literature on the analysis of recurrent-event data. Other authors who have investigated recurrent-event data include Cai and Schaubel (2004), Lawless and Nadeau (1995), Lin *and others* (2000), Pepe and Cai (1993), and Wang and Chen (2000). In regard to panel-count data, among others, Sun and Kalbfleisch (1995), Wellner and Zhang (2000), and Hu *and others* (2009) considered non-parametric estimation of the mean function of the underlying recurrent-event process. Cheng and Wei (2000), Sun and Wei (2000), Wellner and Zhang (2007), and Zhang (2002) investigated regression analysis of panel-count data. More recently, Zhao *and others* (2010) reviewed non-parametric approaches on analyzing panel-count data, and Sun and Zhao (2013) provided a relatively complete review of the existing literature on the topic.

In comparison, there exists little research on the analysis of mixed recurrent-event and panel-count data except Zhu *and others* (2013), who presented a simple and intuitively appealing approach for regression analysis of the data. As pointed out in Zhu *and others* (2013), a main difficulty for the analysis of mixed data is that one needs to model or take into account two different data structures. To avoid this problem, it is apparent that a naive approach is to base the analysis only on the observed recurrent-event or panel-count data. A similar approach is to transform the mixed data into recurrent-event data by imputation or panel-count data by summarizing and perform the analysis accordingly. Zhu *and others* (2013) showed by the simulation study that these two methods could either give biased results or lose efficiency.

In the following sections, we propose two approaches that make use of all available information and do not rely on imputation procedures. In Section 2, we first present a maximum likelihood approach under

the Poisson process assumption about the underlying recurrent-event process of interest. The resulting maximum likelihood estimator of regression parameters is efficient and follows an asymptotic normal distribution. Sometimes the Poisson assumption may not hold. To address this, in Section 3, we present an estimating equation-based approach. As with the maximum likelihood estimator, the asymptotic properties of the new estimator of regression parameters are established. Section 4 gives some results from an extensive simulation study and they suggest that both estimation approaches seem to work well in practice. In Section 5, we apply the methods to the mixed data arising from the CCSS and Section 6 contains some discussion and concluding remarks.

2. Maximum likelihood approach

Consider a recurrent-event study that consists of n independent subjects. Suppose that some of the subjects are observed continuously and provide recurrent-event data, while others are observed only periodically and give panel-count data. For subject i, define $r_i = 1$ if the subject is observed continuously and 0 otherwise. Thus, the r_is are observation-type indicators. Also for each subject, suppose that there exists a p-dimensional vector of covariates X_i and a follow-up time C_i, and assume that both C_i and r_i are independent of the underlying recurrent-event process of interest.

Let $N_i^*(t)$ denote the underlying recurrent-event process, representing the cumulative number of the events that subject i has experienced up to time t, $i = 1, \ldots, n$. Define $N_i(t) = N_i^*(t \wedge C_i)$, the observed recurrent-event process. For subject i, let $T_{i1} < T_{i2} < \cdots < T_{iK_i}$ denote the times at which the recurrent event of interest occurs if $r_i = 1$ or the times where the subject is observed if $r_i = 0$, where K_i denotes the total number of events or observations. Then the observed data have the form $\{O_i = (r_i, \mathbf{T}_i, \mathbf{N}_i, K_i, C_i, X_i);\ i = 1, \ldots, n\}$ with $\mathbf{T}_i = \{T_{i1}, \ldots, T_{iK_i}\}$ and $\mathbf{N}_i = \{N_i(T_{i1}), \ldots, N_i(T_{iK_i})\}$.

In this section, we assume that $N_i^*(t)$ is a non-homogeneous Poisson process with the following intensity function:

$$\lambda_i(t|X_i) = \lambda(t)\, \mathrm{e}^{X_i'\beta}, \tag{2.1}$$

given X_i. In this function, β denotes the regression parameters, and $\lambda(t)$ is an unspecified baseline intensity function. Define $\Lambda(t) = \int_0^t \lambda(s)\, \mathrm{d}s$ and $\theta = (\beta, \Lambda)$. Then under the Poisson assumption, the likelihood function of θ has the form

$$L_n(\theta) = \prod_{i=1}^{n} \left\{ \mathrm{e}^{-\Lambda(C_i)\, \mathrm{e}^{X_i'\beta}}\, \mathrm{e}^{K_i X_i'\beta} \prod_{j=1}^{K_i} \lambda(T_{ij}) \right\}^{r_i} \left[\mathrm{e}^{N_{iK_i} X_i'\beta}\, \mathrm{e}^{-\Lambda(T_{iK_i})\, \mathrm{e}^{X_i'\beta}} \prod_{j=1}^{K_i} \{\Lambda(T_{ij}) - \Lambda(T_{ij-1})\}^{\Delta N_{ij}} \right]^{1-r_i},$$

where $T_{i0} = 0$, $N_{iK_i} = N_i(T_{iK_i})$, and $\Delta N_{ij} = N_i(T_{ij}) - N_i(T_{ij-1})$ for $j = 1, \ldots, K_i$, $i = 1, \ldots, n$. Correspondingly, the log-likelihood function has the form

$$\sum_{i=1}^{n} \sum_{j=1}^{K_i} r_i \log \lambda(T_{ij}) + \sum_{i=1}^{n} \Bigg[-r_i \Lambda(C_i)\, \mathrm{e}^{X_i'\beta} + r_i K_i X_i'\beta + (1 - r_i) N_{iK_i} X_i'\beta - (1 - r_i)\Lambda(T_{iK_i})\, \mathrm{e}^{X_i'\beta} + \sum_{j=1}^{K_i} (1 - r_i) \Delta N_{ij} \log\{\Lambda(T_{ij}) - \Lambda(T_{ij-1})\} \Bigg]. \tag{2.2}$$

Thus, for estimation of θ, it is natural to maximize this log-likelihood function.

Let $0 < t_1 < \cdots < t_M$ denote the ordered distinction time points of all $\{T_{ij}\}$ from the subjects with $r_i = 1$. Define the parameter space $\Theta_n = \{\theta = (\beta, \Lambda) : \beta \in B, \Lambda$ is a right-continuous and step functions having jumps only at the t_j's$\}$. To maximize the log-likelihood function given in (2.2), it is easy to see that one

cannot consider all non-decreasing functions $\Lambda(t)$ and instead, we should focus on these $\Lambda(t)$ in Θ_n. For this, we will consider the modified log-likelihood function

$$l_n(\theta) = \sum_{m=1}^{M} \tilde{N}_m \log \Delta\Lambda(t_m) + \sum_{i=1}^{n} \Bigg[-r_i \Lambda(C_i)\, \mathrm{e}^{X_i'\beta} + r_i K_i X_i'\beta + (1-r_i) N_{iK_i} X_i'\beta - (1-r_i)\Lambda(T_{iK_i})\, \mathrm{e}^{X_i'\beta} + \sum_{j=1}^{K_i} (1-r_i)\Delta N_{ij} \log\{\Lambda(T_{ij}) - \Lambda(T_{ij-1})\} \Bigg], \tag{2.3}$$

where $\tilde{N}_m = \sum_{i=1}^{n} r_i \sum_{j=1}^{K_i} I(T_{ij} = t_m)$, and $\Delta\Lambda(t_m) = \Lambda(t_m) - \Lambda(t_{m-1})$ with $\Lambda(t_0) = 0$. Define the estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n(t))$ of θ to be the value of θ that maximizes $l_n(\theta)$ over Θ_n. Let $\theta_0 = (\beta_0, \Lambda_0(t))$ denote the true value of θ. Then under the conditions (C1)–(C6) given in supplementary material available at *Biostatistics* online, one can show that $\hat{\beta}_n \to \beta_0$ a.s. and $\sup_{t\in[0,\tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)| \to 0$ a.s., where τ denotes the longest follow-up time. Furthermore, it can be shown that under the same conditions, $n^{1/2}(\hat{\Lambda}_n(t) - \Lambda_0(t), \hat{\beta}_n - \beta_0)$ converges weakly to a zero-mean Gaussian process and $\hat{\beta}_n$ is asymptotically efficient. The proofs of these results are sketched in supplementary material available at *Biostatistics* online.

For the determination of $\hat{\theta}_n$, we propose the following iterative algorithm. Let $\theta^{(0)} = (\beta^{(0)}, \Lambda^{(0)}(t))$ denote an initial estimator. At the kth iteration, define the updated estimator of $\Lambda(t)$ as

$$\Lambda^{(k)}(t) = \sum_{m=1}^{M} h_m^{(k)} I(t \geqslant t_m),$$

where

$$h_m^{(k)} = \left\{ \sum_{i=1}^{n} \sum_{j=1}^{K_i} I(T_{ij} = t_m) \right\} \left\{ \sum_{i=1}^{n} I(C_i \geqslant t_m) r_i\, \mathrm{e}^{X_i'\beta^{(k-1)}} + (1-r_i) I(T_{iK_i} \geqslant t_m)\, \mathrm{e}^{X_i'\beta^{(k-1)}} - \sum_{j=1}^{K_i} (1-r_i)\Delta N_{ij} \frac{I(t_m \in (T_{ij-1}, T_{ij}])}{\Lambda^{(k-1)}(T_{ij}) - \Lambda^{(k-1)}(T_{ij-1})} \right\}^{-1}.$$

Also the updated estimator $\beta^{(k)}$ can be obtained by solving the following equation:

$$\sum_{i=1}^{n} r_i (K_i - \Lambda^{(k)}(C_i)\, \mathrm{e}^{X_i'\beta}) X_i + (1-r_i)(N_{iK_i} - \Lambda^{(k)}(T_{iK_i})\, \mathrm{e}^{X_i'\beta}) X_i = 0. \tag{2.4}$$

To choose the initial estimator, one way is to consider only the recurrent-event data and apply the corresponding estimation procedure such as that given in Lin *and others* (2000).

Note that in the above, the t_j's are defined based only on the observed recurrent-event data. In the case where the proportion of recurrent-event data is small, the estimate of $\Lambda(t)$ given above may not be accurate, and the algorithm may not converge. In that case, one can use all T_{ij}s for the definition of the t_js. That is, let $0 < t_1 < \cdots < t_l < \cdots < t_L$ denote the distinct time points of all $\{T_{ij}\}$, At the kth step, one can estimate $\Lambda^{(k)}(t_l)$ by $\sum_{j:t_j \leqslant t_l} \lambda(t_j)$ with

$$\lambda(t_l) = \frac{\sum_{i=1}^{n} Y_i(t_l) \left\{ r_i dN_i(t_l) + (1-r_i)\tilde{\Delta}N(t_l)\lambda(t_l)/\tilde{\Delta}\Lambda^{(k-1)}(t_l) \right\}}{\sum_{i=1}^{n} Y_i(t_l)\, \mathrm{e}^{X_i\beta^{(k-1)}}}.$$

Here $Y_i(t_l) = I(t_l \leqslant C_i)$, $\tilde{\Delta}\Lambda_i(t_l) = \Lambda(R_i(t_l)) - \Lambda(L_i(t_l))$, and $\tilde{\Delta}N_i(t_l) = N_i(R_i(t_l)) - N_i(L_i(t_l))$ with $R_i(t_l) = \min\{t_{ij},\ \ j = 1, \ldots, K_i;\ T_{ij} \geqslant t_l\}$ and $L_i(t_l) = \max\{t_{ij},\ \ j = 1, \ldots, K_i;\ T_{ij} < t_l\}$, $l = 1, \ldots, L$ (Hu *and others*, 2009). The two estimates of $\Lambda(t)$ given above are asymptotically equivalent. To estimate the covariance matrix of $\hat{\beta}_n$, by following Zeng and Lin (2006), one can regard β and the $\Delta\Lambda(t_j)$s as unknown parameters in (2.3) and compute the inverse of the observed information matrix evaluated at $\hat{\Lambda}_n$ and $\hat{\beta}_n$. It follows that the covariance matrix of $\hat{\beta}_n$ can be estimated by the submatrix of the obtained inverse matrix corresponding to $\hat{\beta}_n$.

3. Estimating equation-based approach

Although the estimator $\hat{\beta}_n$ given above is asymptotically efficient, it is well known that the Poisson process assumption used may not hold in practice. In this section, we present a different estimation procedure that does not rely on this assumption. To describe the covariate effects, we assume that given X_i, the mean function of $N_i^*(t)$ has the form

$$E\{N_i^*(t)|X_i\} = \Lambda(t)\, \mathrm{e}^{X_i'\beta}.$$

Here $\Lambda(t)$ denotes the baseline mean function and β represents covariate effects as before. The model above is often referred to as the proportional mean model (Cook and Lawless, 2007).

To develop an estimating equation for β, it is natural to consider (2.4). To use it, we need to estimate the function $\Lambda(t)$. For this, let the t_l's be defined as before, d_l denote the number of the event time points equal to t_l, and n_l the number of the event time points satisfying $T_{ij} \leqslant s_l \leqslant C_i$ among all subjects with $r_i = 1$. Wang *and others* (2001) and Huang and Wang (2004) suggested that one can estimate $F(t) = \Lambda(t)/\Lambda(\tau)$ by

$$\hat{F}_n(t) = \prod_{s_l > t}(1 - d_l/n_l).$$

Let $\theta_1 = (\beta', \log\Lambda(\tau))'$. Then motivated by (2.4) and the estimator above, we propose the following estimating equation:

$$U_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} X_{1i}\{r_i(K_i - \hat{F}_n(C_i)\, \mathrm{e}^{X_{1i}'\theta_1}) + (1 - r_i)(N_{iK_i} - \hat{F}_n(T_{iK_i})\, \mathrm{e}^{X_{1i}'\theta_1})\}, \tag{3.1}$$

where $X_{1i} = (X_i', 1)'$.

Let $\hat{\theta}_n^*$ denote the estimator of θ_1 given by the solution to (3.1). Define

$$Q_n(t) = \left\{\sum_{i=1}^{n} r_i\right\}^{-1} \sum_{i=1}^{n} r_i \sum_{j=1}^{K_i} I(T_{ij} \leqslant t),$$

$$R_n(t) = \left\{\sum_{i=1}^{n} r_i\right\}^{-1} \sum_{i=1}^{n} r_i \sum_{j=1}^{K_i} I(T_{ij} \leqslant t \leqslant C_i),$$

and

$$b_{in}(t) = \sum_{j=1}^{K_i}\left\{\int_t^{\tau} \frac{I(T_{ij} \leqslant u \leqslant C_i)}{R_n(u)^2}\, \mathrm{d}Q_n(u) - \frac{I(t \leqslant T_{ij} \leqslant \tau)}{R_n(T_{ij})}\right\}.$$

Then one can show that as $\hat{\theta}_n$, under the conditions (C1)–(C5) given in supplementary material available at *Biostatistics* online, $\hat{\theta}_n^*$ is also consistent. In addition, one can approximate the distribution of

$n^{1/2}(\theta_n^* - \theta_{10})$ by the normal distribution with mean zero and the covariance matrix $\hat{A}_n^{-1}\hat{\Sigma}_n\hat{A}_n^{-1}$. In the above, θ_{10} denotes the true value of θ_1,

$$\hat{A}_n = \frac{1}{n}\sum_{i=1}^{n} X_{1i}X_{1i}'\{r_i\hat{F}_n(C_i)\,\mathrm{e}^{X_{1i}'\hat{\theta}_n^*} + (1-r_i)\hat{F}_n(T_{iK_i})\,\mathrm{e}^{X_{1i}'\hat{\theta}_n^*}\},$$

and $\hat{\Sigma}_n = (1/n)\sum_{i=1}^{n} u_i u_i'$ with

$$u_i = X_{1i}[r_i\{K_i - \hat{F}_n(C_i)\,\mathrm{e}^{X_{1i}'\hat{\theta}_n^*}\} + (1-r_i)\{N_{iK_i} - \hat{F}_n(T_{iK_i})\,\mathrm{e}^{X_{1i}'\hat{\theta}_n^*}\}]$$
$$- \frac{1}{n}\sum_{j=1}^{n} X_{1j}\,\mathrm{e}^{X_{1j}'\hat{\theta}_n^*}\{r_j\hat{F}_n(C_j)b_{in}(C_j) + (1-r_j)\hat{F}_n(T_{jK_j})b_{in}(T_{jK_j})\}.$$

The sketch of the proof is given in supplementary material available at *Biostatistics* online.

It is worth noting that although the estimating (3.1) is the same as the score function used in the maximum likelihood approach, the two estimation procedures give different estimators even under the Poisson assumption. This is because different estimators are used for $\Lambda(t)$ in the two procedures. Especially, in the procedure given in this section, for estimation of $\Lambda(t)$, we employ the relationship $F(t) = \Lambda(t)/\Lambda(\tau)$ and the estimator given in Wang *and others* (2001). Also note that although we have assumed that the indicator of the observation type is independent of the underlying counting process, the simulation results below indicate that the two proposed methods seem to be valid as long as r_i and $N_i^*(t)$ are independent given covariates. Under this situation, one can show that both the estimating (2.4) and (3.1) are unbiased.

4. SIMULATION STUDY

An extensive simulation study was conducted to assess the finite sample performance of the estimation procedures proposed in the previous sections for mixed recurrent-event and panel-count data. In the study, the covariate X_i was assumed to follow the Bernoulli distribution with the probability of success being 0.5, and the censoring time C_i was generated from the uniform distribution $U(\tau/2, \tau)$ with $\tau = 1$. For the data-type indicator r_i, we generated it from the Bernoulli distribution with the percentage of the subjects giving recurrent-event data, denoted by p_r, being 0.3, 0.5, 0.7, or 0.9, independent of covariate X_i. In addition, to assess the robustness of the procedures, we considered the case where the r_is were generated in the same way, but p_r was assumed to be related to X_i. Specifically, we set p_r =0.2, 0.4, 0.6, or 0.81 for the subjects with $X_i = 0$ and 0.4, 0.6, 0.8, or 0.99 otherwise. Note that in the latter case, on average, p_r is still equal to 0.3, 0.5, 0.7, or 0.9.

For the underlying recurrent-event process $N_i^*(t)$, we also considered two situations. In the first, we assumed that the process is a Poisson process satisfying model (2.1) with $\lambda(t) = 3$. In the second, we assumed that $N_i^*(t)$ is a mixed Poisson process with the mean function $3t\,\mathrm{e}^{X_i\beta}v_i$ given v_i, where the v_i's are i.i.d. random variables from the gamma distribution $\Gamma(2, \frac{1}{2})$. Finally, for the subjects with $r_i = 0$, the observation time points were generated from the Poisson process with the mean function $3t$. The results given below are based on 1000 replications with the sample size of 100 or 200.

Tables 1 and 2 present the results on estimation of β based on the simulated mixed data from the Poisson process with the true value β_0 being -0.5, 0, or 0.5. Table 1 shows the case where p_r is independent of the covariate, and Table 2 shows the case where p_r depends on the covariate. The results include the averages of the point estimates (Estimate) given by the two estimation procedures, the sample standard errors of the estimates (SSEs), the averages of the estimated standard errors (ESEs), and the 95% empirical coverage probabilities (CPs). For comparison, we also obtained and include in the tables the estimates given by

Table 1. *Simulation results on estimation of β by the two proposed estimation procedures and the method given in Zhu and others (2013) based on the simulated data generated under the Poisson assumption with p_r independent on X_i. In the table, $\hat{\beta}_n$ denotes the estimator given by the maximum likelihood approach proposed in Section 2, $\hat{\beta}_n^*$ the estimator given by the estimating equation method proposed in Section 3, Est the average value of the estimates in 1000 Monte Carlo samples of size n, SSE the sample standard error of the same estimates, ESE the average of the estimated standard errors, and CP the coverage probability of nominal 95% confidence intervals by using the normal distribution*

		$\hat{\beta}_n$				$\hat{\beta}_n^*$				Zhu *and others* (2013)			
β	p_r	Est	SSE	ESE	CP	Est	SSE	ESE	CP	Est	SSE	ESE	CP
						$n = 100$							
0.5	0.3	0.49	0.14	0.13	0.95	0.49	0.14	0.14	0.94	0.51	0.21	0.20	0.94
	0.5	0.48	0.14	0.13	0.94	0.49	0.13	0.13	0.95	0.49	0.19	0.18	0.94
	0.7	0.49	0.13	0.13	0.95	0.50	0.13	0.13	0.94	0.49	0.17	0.16	0.94
	0.9	0.49	0.12	0.12	0.95	0.49	0.12	0.12	0.94	0.50	0.13	0.13	0.94
0	0.3	0.00	0.15	0.14	0.93	−0.00	0.15	0.16	0.94	−0.01	0.22	0.21	0.94
	0.5	0.01	0.17	0.15	0.93	0.01	0.15	0.15	0.94	−0.01	0.20	0.19	0.94
	0.7	0.00	0.15	0.14	0.96	0.00	0.14	0.14	0.94	0.01	0.19	0.17	0.93
	0.9	−0.01	0.13	0.14	0.96	0.00	0.13	0.14	0.93	−0.01	0.14	0.14	0.95
−0.5	0.3	−0.47	0.18	0.17	0.93	−0.50	0.18	0.18	0.94	−0.50	0.25	0.24	0.93
	0.5	−0.48	0.22	0.17	0.93	−0.49	0.17	0.17	0.95	−0.50	0.23	0.22	0.93
	0.7	−0.48	0.19	0.17	0.94	−0.50	0.16	0.17	0.94	−0.51	0.20	0.19	0.94
	0.9	−0.50	0.16	0.16	0.95	−0.50	0.16	0.16	0.93	−0.50	0.17	0.16	0.94
						$n = 200$							
0.5	0.3	0.50	0.10	0.11	0.97	0.49	0.10	0.10	0.94	0.48	0.14	0.14	0.95
	0.5	0.48	0.10	0.09	0.94	0.49	0.09	0.09	0.95	0.49	0.14	0.13	0.93
	0.7	0.48	0.09	0.09	0.94	0.49	0.09	0.09	0.94	0.49	0.12	0.11	0.93
	0.9	0.49	0.08	0.09	0.96	0.49	0.09	0.09	0.95	0.50	0.10	0.09	0.94
0	0.3	0.01	0.10	0.11	0.96	0.00	0.11	0.11	0.95	0.00	0.16	0.15	0.95
	0.5	0.00	0.12	0.11	0.93	0.00	0.10	0.10	0.95	0.00	0.14	0.14	0.94
	0.7	−0.01	0.10	0.10	0.95	0.00	0.10	0.10	0.95	0.00	0.13	0.13	0.95
	0.9	0.00	0.10	0.10	0.96	0.00	0.10	0.10	0.95	0.01	0.11	0.10	0.94
−0.5	0.3	−0.50	0.13	0.13	0.96	−0.49	0.12	0.12	0.96	−0.49	0.18	0.17	0.93
	0.5	−0.50	0.14	0.12	0.94	−0.51	0.12	0.12	0.95	−0.49	0.16	0.16	0.93
	0.7	−0.50	0.12	0.12	0.96	−0.49	0.11	0.12	0.94	−0.50	0.14	0.14	0.94
	0.9	−0.49	0.11	0.11	0.97	−0.50	0.11	0.11	0.95	−0.50	0.12	0.12	0.96

the estimation procedure proposed in Zhu *and others* (2013). One can see from the tables that the two estimates proposed above appear to be unbiased, and the proposed ESEs are comparable with the SSEs in both cases. Also as expected, the estimates became better or more efficient when the sample size or p_r increased. Note that the increasing of p_r means that more information is available. The results given in the tables also show that both proposed estimates are more efficient than that given in Zhu *and others* (2013). Note that as pointed out by a referee, the SSE for $\hat{\beta}_n$ seems to be a little bigger than that for $\hat{\beta}_n^*$ in general. The main reason for this is that the second estimation procedure is more stable in general than the first estimation procedure since the latter involves estimation of many more parameters than the former.

The results obtained under the mixed Poisson process are given in Tables 3 and 4. Table 3 shows the case where p_r is independent of X_i, and Table 4 shows the situation where p_r depends on X_i. Note that

Table 2. *Simulation results on estimation of β by the two proposed estimation procedures and the method given in Zhu and others (2013) based on the simulated data generated under the Poisson assumption with p_r dependent on X_i*

		$\hat{\beta}_n$				$\hat{\beta}_n^*$				Zhu *and others* (2013)			
β	p_r	Est	SSE	ESE	CP	Est	SSE	ESE	CP	Est	SSE	ESE	CP
						$n = 100$							
0.5	0.3	0.50	0.14	0.13	0.92	0.49	0.14	0.14	0.94	0.49	0.23	0.21	0.91
	0.5	0.48	0.14	0.13	0.93	0.50	0.13	0.13	0.95	0.50	0.20	0.19	0.92
	0.7	0.49	0.13	0.13	0.95	0.50	0.13	0.13	0.94	0.50	0.18	0.16	0.93
	0.9	0.49	0.12	0.13	0.95	0.49	0.12	0.12	0.95	0.49	0.13	0.13	0.94
0	0.3	0.00	0.16	0.13	0.91	0.00	0.16	0.15	0.95	0.01	0.23	0.22	0.94
	0.5	0.02	0.16	0.15	0.94	0.01	0.15	0.15	0.94	0.01	0.21	0.20	0.93
	0.7	0.00	0.14	0.14	0.95	0.01	0.14	0.14	0.94	0.00	0.17	0.17	0.94
	0.9	0.00	0.13	0.14	0.96	0.01	0.14	0.13	0.94	0.01	0.15	0.14	0.94
−0.5	0.3	−0.49	0.18	0.16	0.93	−0.50	0.18	0.18	0.94	−0.51	0.24	0.24	0.95
	0.5	−0.47	0.21	0.17	0.93	−0.49	0.17	0.17	0.95	−0.49	0.23	0.22	0.93
	0.7	−0.49	0.18	0.17	0.95	−0.50	0.17	0.16	0.94	−0.50	0.20	0.19	0.94
	0.9	−0.50	0.15	0.16	0.97	−0.50	0.16	0.16	0.93	−0.50	0.17	0.16	0.94
						$n = 200$							
0.5	0.3	0.49	0.10	0.11	0.97	0.50	0.10	0.10	0.94	0.49	0.16	0.15	0.93
	0.5	0.48	0.10	0.09	0.95	0.49	0.09	0.09	0.95	0.50	0.15	0.14	0.93
	0.7	0.49	0.09	0.09	0.95	0.50	0.09	0.09	0.95	0.49	0.12	0.12	0.94
	0.9	0.49	0.09	0.09	0.96	0.49	0.09	0.09	0.95	0.49	0.09	0.09	0.95
0	0.3	0.01	0.11	0.11	0.96	0.00	0.11	0.11	0.94	0.00	0.16	0.16	0.93
	0.5	−0.01	0.12	0.11	0.95	0.01	0.10	0.10	0.96	0.00	0.15	0.14	0.95
	0.7	−0.01	0.10	0.10	0.95	0.00	0.10	0.10	0.96	0.00	0.14	0.13	0.93
	0.9	0.00	0.10	0.10	0.95	0.01	0.10	0.10	0.95	0.00	0.10	0.10	0.96
−0.5	0.3	−0.50	0.12	0.12	0.94	−0.49	0.12	0.12	0.95	−0.49	0.18	0.17	0.93
	0.5	−0.49	0.14	0.12	0.94	−0.50	0.12	0.12	0.95	−0.50	0.16	0.16	0.95
	0.7	−0.50	0.12	0.12	0.95	−0.49	0.12	0.11	0.94	−0.49	0.15	0.14	0.93
	0.9	−0.50	0.12	0.11	0.95	−0.49	0.11	0.11	0.95	−0.49	0.12	0.11	0.95

here we considered only the estimation procedures given in Section 3 and Zhu *and others* (2013). As with Tables 1 and 2, the results again indicate that the estimator presented in Section 3 appears to be unbiased and more efficient than that proposed in Zhu *and others* (2013). To evaluate the normal approximation to the distributions of the two proposed estimates, we studied the quantile plots of the standardized estimates against the standard normal distribution. The plots, not given here, indicate that the approximation performs well.

In addition, suggested by a referee, we also conducted two other studies. One is to allow r_i to depend on $N_i^*(t)$ even conditional on X_i and the other is to compare the proposed methods to the naive method that bases the analysis only on the subjects giving recurrent-event data or panel-count data. Table 5 presents the results on estimation of β based on the simulated data with r_i being dependent on $N_i^*(t)$. Specifically, for subject i, we took $p_{r_i} = p_r + 0.2$ if the total number of the observed recurrent events is equal to or greater than 2 and $p_{r_i} = p_r + 0.2$ otherwise. The other set-ups were the same as with Table 1. One can see that the results are similar to those given in Table 1 and suggest that the proposed estimation procedures seem still to be valid. Table 6 gives the results on estimation of β given by the two proposed procedures and obtained based only on recurrent-event data ($\hat{\beta}_r$) or panel-count data ($\hat{\beta}_p$). Here we used the same set-up as with

Table 3. *Simulation results on estimation of β by the proposed estimating equation procedure and the method given in Zhu and others (2013) based on the simulated data generated under the mixed Poisson assumption with p_r independent on X_i*

		$\hat{\beta}_n^*$				Zhu *and others* (2013)			
β	p_r	Est	SSE	ESE	CP	Est	SSE	ESE	CP
				$n = 100$					
0.5	0.3	0.50	0.16	0.16	0.94	0.49	0.23	0.22	0.94
	0.5	0.50	0.16	0.16	0.94	0.49	0.22	0.20	0.93
	0.7	0.51	0.16	0.15	0.94	0.51	0.20	0.18	0.93
	0.9	0.50	0.15	0.15	0.95	0.50	0.16	0.15	0.94
0	0.3	0.01	0.18	0.17	0.93	0.01	0.24	0.22	0.92
	0.5	−0.01	0.16	0.16	0.95	−0.02	0.22	0.20	0.94
	0.7	0.00	0.16	0.15	0.94	−0.01	0.20	0.18	0.93
	0.9	0.00	0.15	0.15	0.94	0.00	0.17	0.16	0.94
−0.5	0.3	−0.50	0.18	0.17	0.94	−0.50	0.25	0.23	0.93
	0.5	−0.51	0.17	0.17	0.95	−0.51	0.23	0.21	0.93
	0.7	−0.50	0.16	0.16	0.94	−0.51	0.20	0.19	0.92
	0.9	−0.50	0.16	0.15	0.94	−0.51	0.17	0.16	0.94
				$n = 200$					
0.5	0.3	0.51	0.12	0.12	0.94	0.50	0.17	0.16	0.93
	0.5	0.50	0.11	0.11	0.96	0.50	0.14	0.15	0.95
	0.7	0.50	0.11	0.11	0.94	0.50	0.14	0.13	0.94
	0.9	0.50	0.10	0.10	0.96	0.50	0.11	0.11	0.95
0	0.3	0.00	0.12	0.12	0.94	−0.01	0.17	0.16	0.94
	0.5	0.01	0.11	0.11	0.95	0.02	0.15	0.15	0.95
	0.7	−0.01	0.11	0.11	0.94	0.00	0.14	0.13	0.93
	0.9	−0.01	0.10	0.11	0.95	−0.01	0.11	0.11	0.94
−0.5	0.3	−0.50	0.12	0.12	0.95	−0.50	0.17	0.17	0.95
	0.5	−0.50	0.12	0.12	0.95	−0.50	0.16	0.15	0.95
	0.7	−0.51	0.11	0.11	0.95	−0.51	0.14	0.14	0.94
	0.9	−0.50	0.11	0.11	0.95	−0.51	0.12	0.12	0.94

Table 1 and only calculated the sample standard errors of the obtained estimators. They indicate that the proposed estimators seem to be more efficient than $\hat{\beta}_r$ and $\hat{\beta}_p$ and as expected, the efficiency of $\hat{\beta}_r$ and $\hat{\beta}_p$ depends on p_r. We also considered other set-ups and obtained similar results.

5. Analysis of the CCSS

In this section, we apply the two proposed estimation procedures to the mixed recurrent-event and panel-count data arising from the CCSS. The CCSS is a multicenter longitudinal cohort study (Robison *and others*, 2002), and since 1996, it has distributed summary questionnaires periodically to more than 13 000 childhood cancer survivors who were diagnosed between 1970 and 1986 and have survived for at least 5 years since diagnosis. Questionnaires are also sent to a random sample of the survivors' siblings, who serve as a control group. One objective of the CCSS is to compare the pregnancy rates of survivors and siblings to determine the effect of prior childhood cancer treatment on reproductive function. The summary questionnaire asked the participants to report pregnancies, including the age range at the

Table 4. *Simulation results on estimation of β by the proposed estimating equation procedure and the method given in Zhu and others (2013) based on the simulated data generated under the mixed Poisson assumption with p_r dependent on X_i*

		$\hat{\beta}_n^*$				Zhu *and others* (2013)			
β	p_r	Est	SSE	ESE	CP	Est	SSE	ESE	CP
				$n = 100$					
0.5	0.3	0.49	0.17	0.17	0.94	0.50	0.24	0.23	0.95
	0.5	0.50	0.16	0.16	0.95	0.50	0.22	0.21	0.95
	0.7	0.51	0.16	0.15	0.94	0.51	0.20	0.18	0.93
	0.9	0.50	0.15	0.15	0.94	0.50	0.16	0.15	0.93
0	0.3	0.00	0.16	0.17	0.95	−0.01	0.24	0.23	0.93
	0.5	0.00	0.16	0.16	0.94	0.00	0.22	0.21	0.94
	0.7	0.00	0.16	0.15	0.95	0.01	0.20	0.18	0.92
	0.9	0.00	0.15	0.15	0.94	−0.01	0.16	0.15	0.95
−0.5	0.3	−0.50	0.17	0.17	0.96	−0.49	0.25	0.24	0.94
	0.5	−0.50	0.18	0.17	0.93	−0.51	0.23	0.21	0.93
	0.7	−0.51	0.16	0.16	0.95	−0.51	0.20	0.19	0.93
	0.9	−0.50	0.16	0.15	0.94	−0.50	0.17	0.16	0.94
				$n = 200$					
0.5	0.3	0.50	0.12	0.12	0.94	0.50	0.17	0.17	0.94
	0.5	0.50	0.11	0.11	0.95	0.50	0.16	0.15	0.94
	0.7	0.51	0.11	0.11	0.95	0.51	0.14	0.14	0.94
	0.9	0.50	0.11	0.10	0.94	0.50	0.11	0.11	0.94
0	0.3	0.00	0.12	0.12	0.94	−0.01	0.17	0.17	0.95
	0.5	0.00	0.11	0.11	0.95	0.00	0.16	0.15	0.94
	0.7	0.00	0.11	0.11	0.95	0.00	0.14	0.13	0.94
	0.9	0.00	0.11	0.11	0.94	0.00	0.11	0.11	0.95
−0.5	0.3	−0.50	0.12	0.12	0.95	−0.50	0.18	0.17	0.94
	0.5	−0.50	0.12	0.12	0.95	−0.50	0.16	0.15	0.95
	0.7	−0.50	0.12	0.11	0.94	−0.50	0.14	0.14	0.94
	0.9	−0.51	0.11	0.11	0.94	−0.51	0.11	0.11	0.96

beginning of each pregnancy (under 15, 15–20, 21–25, 26–30, 31–35, 36 and over). If a pregnancy (after cancer treatment) was reported, a detailed pregnancy questionnaire was sent to obtain further information, including the precise age at pregnancy.

The data considered here consist of 3966 female participants who were at least 25 years when the study began and returned the summary questionnaires up to 2006. Among them, 697 participants who reported at least one pregnancy on their summary questionnaires did not return the pregnancy questionnaires. All others returned both questionnaires. Thus, 3269 participants provided recurrent-event data ($r_i = 1$), and 697 participants provided only panel-count data ($r_i = 0$). Of the 3966 subjects, 2765 were cancer survivors, and 1201 were siblings; the average pregnancy counts for the two groups were 1.498 and 2.049, respectively. More specifically, among all cancer survivors, the percentage of the subjects with 0, 1, 2 or more than 2 pregnancies is about 41%, 15%, 19%, or 25%, respectively, while the corresponding percentages for the siblings are 23%, 15%, 28%, and 34%, respectively.

For the analysis, define $X_i = 1$ if the ith subject is a survivor and 0 otherwise. The application of the two proposed estimation approaches yielded $\hat{\beta}_n = -0.396$ and $\hat{\beta}_n^* = -0.334$ with the ESEs being 0.025 and 0.031, respectively. Both results give a p-value close to zero for testing no difference between the

Table 5. *Simulation results on estimation of β based on the simulated data with p_r depending on $N(t)$*

			$\hat{\beta}_n$				$\hat{\beta}_n^*$			
n	β	p_r	EST	SSE	ESE	CP	EST	SSE	ESE	CP
					Case 1					
100	0.5	0.5	0.48	0.12	0.18	0.96	0.47	0.12	0.12	0.94
		0.7	0.49	0.12	0.19	0.97	0.47	0.12	0.12	0.95
	0	0.5	0.00	0.14	0.17	0.96	0.00	0.14	0.14	0.96
		0.7	0.01	0.13	0.19	0.98	0.00	0.13	0.13	0.95
	−0.5	0.5	−0.48	0.17	0.17	0.95	−0.48	0.17	0.16	0.95
		0.7	−0.47	0.16	0.18	0.97	−0.48	0.16	0.15	0.95
200	0.5	0.5	0.48	0.09	0.12	0.96	0.47	0.09	0.09	0.94
		0.7	0.49	0.09	0.11	0.96	0.47	0.08	0.08	0.94
	0	0.5	0.00	0.10	0.11	0.95	0.00	0.10	0.10	0.93
		0.7	0.01	0.10	0.12	0.97	0.00	0.10	0.09	0.94
	−0.5	0.5	−0.48	0.11	0.12	0.96	−0.49	0.11	0.11	0.94
		0.7	−0.48	0.11	0.12	0.96	−0.49	0.11	0.11	0.95
					Case 2					
100	0.5	0.5	0.52	0.14	0.14	0.94	0.51	0.14	0.14	0.95
		0.7	0.52	0.13	0.16	0.97	0.51	0.13	0.13	0.95
	0	0.5	0.00	0.15	0.15	0.94	−0.01	0.15	0.16	0.96
		0.7	0.00	0.14	0.16	0.97	−0.01	0.14	0.15	0.96
	−0.5	0.5	−0.50	0.18	0.17	0.94	−0.52	0.18	0.18	0.95
		0.7	−0.50	0.17	0.17	0.95	−0.51	0.17	0.17	0.95
200	0.5	0.5	0.52	0.10	0.11	0.96	0.52	0.10	0.10	0.95
		0.7	0.53	0.09	0.13	0.97	0.52	0.09	0.09	0.94
	0	0.5	0.01	0.11	0.11	0.95	0.01	0.11	0.11	0.94
		0.7	0.01	0.11	0.12	0.96	0.01	0.11	0.11	0.95
	−0.5	0.5	−0.50	0.12	0.12	0.95	−0.51	0.12	0.12	0.95
		0.7	−0.50	0.12	0.13	0.96	−0.51	0.12	0.12	0.95

pregnancy rates of survivors and siblings and indicate that the pregnancy rate for cancer survivors was significantly lower than that for their siblings. Note that by using a different scale, we have $e^{\hat{\beta}_n} = 0.673$ and $e^{\hat{\beta}_n^*} = 0.716$, suggesting that the average pregnancy number of cancer survivors is about 70% of that of their siblings. In other words, the cancer treatment appears to have some negative effect on reproductive functioning.

6. Discussion and concluding remarks

This paper considered regression analysis of mixed recurrent-event and panel-count data, a type of data that often arises from recurrent-event studies but has not been discussed much. We proposed two estimation procedures, a maximum likelihood approach based on the Poisson process assumption and an estimating equation approach that does not rely on the Poisson assumption. The simulation study demonstrated that both procedures work well in practical situations. The code for the numerical study is available upon request.

Table 6. *Comparison of the proposed estimators with the estimators based only on recurrent-event data or panel-count data*

β	p_r	$\hat{\beta}_n$	SSE($\hat{\beta}_n$)	$\hat{\beta}_n^*$	SSE($\hat{\beta}_n^*$)	$\hat{\beta}_r$	SSE($\hat{\beta}_r$)	$\hat{\beta}_p$	SSE($\hat{\beta}_p$)
					$n = 100$				
0.5	0.3	0.49	0.14	0.49	0.13	0.55	0.23	0.50	0.18
	0.5	0.50	0.13	0.49	0.13	0.53	0.17	0.50	0.23
	0.7	0.51	0.13	0.49	0.12	0.53	0.14	0.51	0.31
	0.9	0.52	0.12	0.49	0.12	0.53	0.13	0.58	0.76
0	0.3	0.00	0.15	−0.01	0.15	0.04	0.27	−0.01	0.20
	0.5	0.00	0.14	−0.01	0.14	0.03	0.20	−0.01	0.24
	0.7	0.01	0.14	0.00	0.14	0.03	0.17	−0.01	0.33
	0.9	0.02	0.13	0.00	0.13	0.02	0.14	0.08	0.83
−0.5	0.3	−0.50	0.18	−0.50	0.18	−0.45	0.32	−0.51	0.23
	0.5	−0.49	0.17	−0.50	0.17	−0.46	0.22	−0.52	0.28
	0.7	−0.49	0.17	−0.50	0.17	−0.47	0.19	−0.53	0.37
	0.9	−0.48	0.16	−0.50	0.16	−0.47	0.17	−0.50	1.00
					$n = 200$				
0.5	0.3	0.50	0.10	0.50	0.10	0.53	0.17	0.50	0.12
	0.5	0.50	0.09	0.49	0.09	0.53	0.13	0.50	0.15
	0.7	0.50	0.09	0.49	0.09	0.53	0.10	0.51	0.20
	0.9	0.52	0.09	0.49	0.08	0.53	0.09	0.53	0.41
0	0.3	0.00	0.11	0.00	0.11	0.04	0.18	0.00	0.15
	0.5	0.00	0.11	0.00	0.11	0.03	0.14	0.00	0.17
	0.7	0.01	0.10	0.00	0.10	0.03	0.12	0.00	0.23
	0.9	0.02	0.10	0.00	0.10	0.03	0.11	0.00	0.45
−0.5	0.3	−0.50	0.12	−0.50	0.12	−0.48	0.21	−0.50	0.16
	0.5	−0.50	0.12	−0.50	0.12	−0.48	0.16	−0.50	0.19
	0.7	−0.49	0.11	−0.50	0.11	−0.48	0.14	−0.50	0.25
	0.9	−0.49	0.11	−0.50	0.11	−0.48	0.12	−0.53	0.55

As pointed out above, the main advantage of the maximum likelihood approach is that the resulting estimate is asymptotically efficient, but the Poisson assumption used could be questionable in practice. In contrast, the estimating equation approach does not depend on the Poisson assumption, and its implementation is simpler than that of the maximum likelihood approach. Also with the maximum likelihood approach, when the percentage of recurrent-event data is low, sometimes one may have a non-convergence issue due to estimating $\Lambda(t)$. On the other hand, the estimating equation approach does not appear to have a non-convergence issue. Also as pointed out above, the estimating equation approach is usually more stable than the maximum likelihood approach as the latter involves estimation of many more parameters.

In this paper, we discussed mixed recurrent-event and panel-count data in which a subject is observed either continuously or at discrete time points over the entire period of follow-up. In practice, as discussed in Zhu *and others* (2013), some subjects may be observed continuously during some periods but then only at discrete time points during others. Thus, one subject may yield both recurrent-event data and panel-count data. It is apparent that the resulting data structure would be much more complicated than that considered herein, and the two estimation procedures proposed would not be applicable anymore. One would need new approaches that are beyond the scope of this paper.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

The authors wish to thank the editor, Dr Molenberghs, and three reviewers for their many useful comments and suggestions that greatly improved the manuscript. We also wish to thank Dr Wenbin Lu for the discussion on supplementary material available at *Biostatistics* online. *Conflict of Interest*: None declared.

Funding

This work was supported by the National Institutes of Health (R03 CA169150) to L.Z., and by the National Natural Science Foundation of China (11371062), China Zhongdian Project (11131002), Beijing Center for Mathematics and Information Interdisciplinary Sciences, and the Fundamental Research Funds for the Central Universities to X.T.

References

Cai, J. and Schaubel, D. E. (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis* **10**, 121–138.

Cheng, S. C. and Wei, L. J. (2000). Inferences for a semiparametric model with panel data. *Biometrika* **87**, 89–97.

Cook, R. J. and Lawless, J. F. (2007). *The Analysis of Recurrent Event Data*. New York: Springer.

Hu, X. J., Lagakos, S. W. and Lockhart, R. A. (2009). Marginal analysis of panel counts through estimating functions. *Biometrika* **96**, 445–456.

Huang, C. Y. and Wang, M. C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* **99**, 1153–1165.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158–168.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate function of recurrent events. *Journal of Royal Statistical Society, Series B* **69**, 711–730.

Pepe, M. S. and Cai, J. (1993). Some graphic displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* **88**, 811–820.

Robison, L. L., Mertens, A. C., Boice, J. D., Breslow, N. E., Donaldson, S. S., Green, D. M., Li, F. P., Meadows, A. T., Mulvihill, J. J., Neglia, J. P. *and others* (2002). Study design and cohort characteristics of the childhood cancer survivor study: a multi-institutional collaborative project. *Medical and Pediatric Oncology* **38**(4), 229–239.

Sun, J. and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistical Sinica* **5**, 279–290.

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society, Series B* **62**, 293–302.

Sun, J. and Zhao, X. (2013). *The Statistical Analysis of Panel Count Data*. New York: Springer.

Wang, M. C. and Chen, Y. Q. (2000). Nonparametric and semiparametric trend analysis of stratified recurrent time data. *Biometrics* **56**, 789–794.

WANG, M. C., QIN, J. AND CHIANG, C. T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* **96**(455), 1057–1065.

WELLNER, J. A. AND ZHANG, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Annals of Statistics* **28**, 779–814.

WELLNER, J. A. AND ZHANG, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Annals of Statistics* **35**, 2106–2142.

ZENG, D. AND LIN, D. Y. (2006). Maximum likelihood estimation in semiparametric transformation models for counting processes. *Biometrika* **93**, 627–640.

ZHANG, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89**, 39–48.

ZHAO, X., BALAKRISHNAN, N. AND SUN, J. (2010). Nonparametric inference based on panel count data. *Test* **20**, 1–42.

ZHU, L., TONG, X., ZHAO, H., SUN, J., SRIVASTAVA, D. K., LEISENRING, W. AND ROBISON, L. L. (2013). Statistical analysis of mixed recurrent event data with application to cancer survivor study. *Statistics in Medicine* doi:10.1002/sim.5674.

[*Received April 22, 2013; revised February 18, 2014; accepted for publication February 19, 2014*]