Statistics in Medicine

Research Article

Received 3 December 2010, Accepted 23 May 2011 Published online 22 July 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4306

Semiparametric transformation models for joint analysis of multivariate recurrent and terminal events

Liang Zhu,*† Jianguo Sun, Deo Kumar Srivastava, Xingwei Tong, Wendy Leisenring, Hui Zhang and Leslie L. Robison

Recurrent event data occur in many clinical and observational studies, and in these situations, there may exist a terminal event such as death that is related to the recurrent event of interest. In addition, sometimes more than one type of recurrent events may occur, that is, one may encounter multivariate recurrent event data with some dependent terminal event. For the analysis of such data, one must take into account the dependence among different types of recurrent events and that between the recurrent events and the terminal event. In this paper, we extend a method for univariate recurrent and terminal events and propose a joint modeling approach for regression analysis of the data and establish the finite and asymptotic properties of the resulting estimates of unknown parameters. The method is applied to a set of bivariate recurrent event data arising from a long-term follow-up study of childhood cancer survivors. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: joint modeling; multivariate analysis; regression analysis; survival analysis

1. Introduction

Recurrent event data usually refer to data in which an event of interest can occur more than once. They arise in many fields such as clinical and longitudinal studies, reliability experiments, and sociological studies [1–4]. Examples of such events include hospitalizations, infections, acute myocardial infections, and tumor metastases. Many authors have discussed the analysis of recurrent event data. For example, [5] and [6] developed some intensity-based methods. Subsequently, [7–9] proposed some marginal mean and rate-based approaches. However, in real life, the analysis of recurrent event data can be more complicated as commented in the following text.

This study was motivated by the Childhood Cancer Survivor Study (CCSS). In the last 30 years, the 5-year childhood cancer survivor rate has been improved significantly. However, as a result, the late effects of cancer and its associated treatment have become an important topic of study in cancer research. For this purpose, CCSS follows more than 14,000 survivors who have survived at least 5 years from their diagnosis of childhood cancer and collects information on participants' late medical effects experienced as well as demographic information [10, 11]. Among these late complications, the most serious ones include recurrence of the original cancer and development of a subsequent new cancer. Both of these late effects may occur repeatedly and place survivors at increased risk for late morbidity or mortality. The two types of cancers are treated differently because a recurrence of the original cancer usually results from the presence of residual cancer cells, whereas new cancer is often a side effect of cancer treatment such as intensive radiation. As a result, there are two correlated recurrent event processes instead of one. Another complicating factor is that during the study, approximately 15% of the patients died: 46% of deaths were due to the recurrence of the original cancer and 15% due to subsequent new cancers. Hence, the terminal event, death, may depend on the underlying recurrent event process.

Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105, U.S.A.

**Correspondence to: Liang Zhu, Department of Biostatistics, St. Jude ChildrenŠs Research Hospital, Memphis, TN 38105, U.S.A.*

†*E-mail: liang.zhu@stjude.org*

Some methods have been proposed for the analysis of recurrent event data with a terminal event. For example, [12–15] developed frailty-based joint modeling procedures that model the recurrent event process and the terminal event process together, whereas [16] and [17] studied the same problem by relying on marginal models instead of the frailty model. However, these approaches apply only to univariate recurrent event data. References [18–21] investigated the analysis of multivariate recurrent event data but only for the situation where the follow-up period is independent of the underlying recurrent process. In this article, we will extend the method of [15] and propose a joint modeling approach that can deal with both multivariate recurrent events and a dependent terminal event.

We organize the remainder of the article as follows. We will begin with introducing some notation and assumptions in Section 2. To present the inference idea, we will first consider the Cox type models for joint regression analysis of the recurrent event processes of interest and the dependent terminal event. For estimation of unknown parameters, we will take the maximum likelihood approach, and we will develop an expectation–maximization (EM) algorithm following [15]. We will then generalize the inference procedure to transformation models. Section 3 gives some results obtained from a simulation study conducted to assess the finite sample properties of the proposed estimates. In Section 4, we apply the proposed methodology to the CCSS data discussed previously. Section 5 contains some concluding remarks.

2. Inference procedures

2.1. Notation and models

Consider a recurrent event study that consists of n independent subjects and involves K types of recurrent events and a terminal event. Let $N_{ik}(t)$ denote the number of the kth type of recurrent events that the ith subject has experienced by time t and T_i the time to the terminal event for subject i, $k = 1, \dots, K$, $i = 1, \dots, n$. For subject i, assume that there exists a vector of possibly time-dependent covariates denoted by $X_i(t)$ and a vector of subject-specific random effects $b_i = (b_{i1}^T, \dots, b_{iK}^T)^T$ whose density function $f(b; \gamma)$ is known up to some unknown parameters γ. Also assume that there exists a censoring time C_i independent of T_i,$N_{ik}(t)$ and b_i, given X_i. Then, the observed data are $O = \{O_i = \{Y_i, \Delta_i, N_{ik}(t), X_i(t); t \leqslant Y_i, k = 1, \dots, K\}\}_{i=1}^n$, where $Y_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leqslant C_i)$ with $I(\cdot)$ being the indicator function.

To model the effects of covariates $X_i(t)$, in this subsection, we will assume that given b_i and X_i, the cumulative intensity function of the recurrent event process $N_{ik}(t)$ and the cumulative hazard function of terminal event time T_i have the forms

$$A_{ik}(t|X_i; b_{ik}) = \int_0^t e^{\alpha_k^T X_i(s) + b_{ik}^T \tilde{X}_i(s)} \mathrm{d}A_{0k}(s) \tag{1}$$

and

$$\Lambda_i(t|X_i; b_i) = \int_0^t e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s). \tag{2}$$

Here, $\tilde{X}_i(s)$ is a subset of $X_i(s)$ plus the unit component, α_k and β are regression parameters, $A_{0k}(\cdot)$ and $\Lambda_0(\cdot)$ are arbitrary increasing functions, ϕ_k is a set of unknown constants, and $(\phi_k \circ b_{ik})$ denotes the component-wise product of ϕ_k and b_{ik}. Here, ϕ indicates the degree of the dependence between the recurrent event and the terminal event processes. Note that in model (1), we use different α_k for different types of recurrent events, and an alternative is to use the same parameter. To look more closely at the models given earlier, consider the situation where $K = 2$ and $\tilde{X}_i(t) = 1$. That is, we have bivariate recurrent event data and only the intercept has a random part. Models (1) and (2) then have the forms

$$\begin{aligned} A_{i1}(t|X_i; b_{i1}) &= \int_0^t e^{\alpha_1^T X_i(s) + b_{i1}} \mathrm{d}A_{01}(s), \\ A_{i2}(t|X_i; b_{i2}) &= \int_0^t e^{\alpha_2^T X_i(s) + b_{i2}} \mathrm{d}A_{02}(s); \\ \Lambda_i(t|X_i; b_i) &= \int_0^t e^{\beta^T X_i(s) + \phi^T b_i} \mathrm{d}\Lambda_0(s), \end{aligned}$$

In the following text, we will assume that given b_{ik} and X_i, $N_{i1} \dots, N_{iK}$ and T_i are independent.

2.2. Estimation of parameters

For estimation of the unknown parameters $\theta = (\alpha_1^T, ..., \alpha_K^T, \beta^T, \phi^T, \gamma^T, A_{01}^T, ..., A_{0K}^T, \Lambda_0^T)^T$, we will take the maximum likelihood approach. For this, note that the full log-likelihood function has the form

$$\begin{aligned}\log L(O) &= \sum_i^n \log \int_b f(O_i|b_i) f(b)\mathrm{d}b \\ &= \sum_i^n \log \int_b \Bigg\{ \prod_{k=1}^K \left[\prod_t \left\{ a_{0k}(t) e^{\alpha_k^T X_i(t) + b_{ik}^T \tilde{X}_i(t)} \right\}^{R_i(t)\Delta N_{ik}(t)} e^{-\int_0^{Y_i} e^{\alpha_k^T X_i(s)+b_{ik}^T \tilde{X}_i(s)} \mathrm{d}A_{0k}(s)} \right] \\ &\qquad \left[\left\{ \lambda_0(Y_i) e^{\beta^T X_i(Y_i) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(Y_i)} \right\}^{\Delta_i} \right. \\ &\qquad \left. \times\ e^{-\int_0^{Y_i} e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s)} \right] \Bigg\} f(b;\gamma)\mathrm{d}b,\end{aligned}$$

where $R_i(t) = I(Y_i \geqslant t)$, $\Delta N_{ik}(t)$ denotes the jump size of $N_{ik}(t)$ at t and $a_{0k}(t)$ and $\lambda_0(t)$ denote the first derivatives of $A_{0k}(t)$ and $\Lambda_0(t)$, respectively. To maximize $\log L(O)$, following [15], we assume that $A_{0k}(t)$ is a step function with jumps only at the observed recurrent event times for the kth type of events and $\Lambda_0(t)$ is a step function with jumps only at the observed terminal event times. In addition, similar to [15], we will replace $a_{0k}(t)$ and $\lambda_0(t)$ in the likelihood function by the jump sizes of $A_{0k}(t)$ and $\Lambda_0(t)$ at time t, denoted by $A_{0k}\{t\}$ and $\Lambda_0\{t\}$, respectively. The conditional density function $f(O_i|b_i)$ in the log-likelihood function can then be rewritten as

$$\begin{aligned}f(O_i|b_i) = \prod_{k=1}^K &\left[\prod_t \left\{ A_{0k}\{t\} e^{\alpha_k^T X_i(t) + b_{ik}^T \tilde{X}_i(t)} \right\}^{R_{ik}(t)\Delta N_{ik}(t)} e^{-\int_0^{Y_i} e^{\alpha_k^T X_i(s)+b_{ik}^T \tilde{X}_i(s)} \mathrm{d}A_{0k}(s)} \right] \\ &\times \left[\left\{ \Lambda_0\{Y_i\} e^{\beta^T X_i(Y_i) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(Y_i)} \right\}^{\Delta_i} e^{-\int_0^{Y_i} e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s)} \right].\end{aligned} \tag{3}$$

It is easy to see that the direct maximization of the log-likelihood function is not straightforward because the integration over b_i has no closed form, and thus we will turn to the EM algorithm by regarding the subject-specific random effects b_{ik}s ($k = 1, ..., K$) as missing data. For this, note that in the E-step of the EM algorithm, we need to evaluate the conditional expectations of functions of b_i, given the observed data. In the M-step, we need to maximize the conditional expectation of the pseudo-complete data log-likelihood function given the observed data. Specifically,

E-step. Let $g(b_i)$ denote a function of b_i and $\Gamma_i(b)$ represent the part of $f(O_i|b_i)$ that involves b_i in Equation (3). Then, the conditional expectation of $g(b_i)$ given the observed data has the form $E\{g(b_i)|O_i\} = \int g(b_i) f(O_i|b_i) f(b_i)\mathrm{d}b / \int f(O_i|b_i) f(b_i)\mathrm{d}b$ and

$$\begin{aligned}\Gamma_i(b) = \prod_{k=1}^K &\left[\prod_t \left\{ e^{b_{ik}^T \tilde{X}_i(t)} \right\}^{R_{ik}(t)\Delta N_{ik}(t)} e^{-\int_0^{Y_i} e^{\alpha_k^T X_i(s)+b_{ik}^T \tilde{X}_i(s)} \mathrm{d}A_{0k}(s)} \right] \\ &\times \left[\left\{ e^{\sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(Y_i)} \right\}^{\Delta_i} e^{-\int_0^{Y_i} e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s)} \right].\end{aligned}$$

This suggests that $E\{g(b_i)|O_i\}$ can be approximated by

$$\hat{E}\{g(b_i)|O_i\} = \int g(b_i)\Gamma_i(b) f(b_i)\mathrm{d}b \Big/ \int \Gamma_i(b) f(b_i)\mathrm{d}b$$

based on the Monte Carlo simulation of b_i or the Gaussian-quadrature approximation when b_i follows the normal distribution.

M-step. To update the estimates for all parameters θ, we will apply the Newton–Raphson method. Also because of the large number of parameters involved in $A_{0k}(t)$ and $\Lambda_0(t)$, we will derive some recursive formulae similar to those in [15], which will involve much less number of parameters.

Let $t_{1,k} < t_{2,k} < t_{3,k} < ... < t_{m_k,k} (k = 1, ..., K)$ be the ordered time points, where the kth type of recurrent events are observed, and let $a_{1,k} < a_{2,k} < a_{3,k} < ... < a_{m_k,k} (k = 1, ..., K)$ be the jump size of A_{0k} at those time points. Define $f_{1k}(t_{j,k}) = a_{j,k}/A_{0k}(t_{m_k,k})$ $(j = 1, ..., m_k)$. Because $A_{0k}(t) = a_{1,k} + ... + a_{j,k}$ for t between $t_{j,k}$ and $t_{j,k+1}$, we have $\sum_j f_{1k}(t_{j,k}) = 1$. If the values of $f_{1k}(t_{j,k})$ for $j = 1, ..., m_k$ and $A_{0k}(t_{m_k,k})$ are known, then A_{0k} can be obtained. Similarly, let $s_1 < s_2 < s_3 < ... < s_{m_T}$ be the ordered observation time points on the terminal event and let $d_1, d_2, d_3, ..., d_{m_T}$ be the jump size of Λ_0 at those time points. Define $f_2(s_j) = d_j/\Lambda_0(s_{m_T})$ $(j = 1, ..., m_T)$, and we have $\sum_j f_2(s_j) = 1$. If the values of $f_2(s_j)$ for $j = 1, ..., m_T$ and $\Lambda_0(t_{m_T})$ are known, then Λ_0 can be obtained.

To estimate $f_{1k}(t_{j,k})$ for $j = 1, ..., m_k$ and $f_2(s_j)$ for $j = 1, ..., m_T$, one can take the derivative of the log likelihood with respect to $f_{1k}(t_{j,k})$ and $f_2(s_j)$ and set the derivatives equal to zero. This leads to the following recursive formulae:

$$\begin{aligned}\frac{1}{f_{1k}(t_{j,k})} =& \frac{1}{f_{1k}(t_{j+1,k})} + A_{0k}(t_{m_k,k}) \sum_{i=1}^{n} \Big[\hat{E}[e^{\alpha_k^T X_i(t_{j,k}) + b_{ik}^T \tilde{X}_i(t_{j,k})}] I(t_{j,k} \leqslant Y_i) \\ & - \hat{E}[e^{\alpha_k^T X_i(t_{j+1,k}) + b_{ik}^T \tilde{X}_i(t_{j+1,k})}] I(t_{j+1,k} \leqslant Y_i) \Big]\end{aligned} \tag{4}$$

When X and $\tilde{X}$ are time independent, Equation (4) can be simplified as

$$\frac{1}{f_{1k}(t_{j,k})} = \frac{1}{f_{1k}(t_{j+1,k})} + A_{0k}(t_{m_k,k}) \sum_{i=1}^{n} \Big[\hat{E}[e^{\alpha_k^T X_i + b_{ik}^T \tilde{X}_i}] I(t_{j,k} \leqslant Y_i < t_{j+1,k}) \Big]$$

If $f_{1k}(t_{m_k})$ is known, $f_{1k}(t_{j,k})$ for $j = 1, ..., m_k - 1$ can be derived. Similarly, for the terminal event, one can derive

$$\begin{aligned}\frac{1}{f_2(s_j)} =& \frac{1}{f_2(s_{j+1})} + \Lambda_0\left(t_{m_T}\right) \sum_{i=1}^{n} \Big[\hat{E}[e^{\beta^T X_i(s_j) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s_j)}] I(s_j \leqslant Y_i) \\ & - \hat{E}[e^{\beta^T X_i(s_{j+1}) + \sum_n (\phi_k \circ b_{ik})^T \tilde{X}_i(s_{j+1})}] I(s_{j+1} \leqslant Y_i) \Big].\end{aligned} \tag{5}$$

When X and $\tilde{X}$ are time-independent, Equation (5) can be simplified as

$$\frac{1}{f_2(s_j)} = \frac{1}{f_2(s_{j+1})} + \Lambda_0(t_{m_T}) \sum_{i=1}^{n} \Big[\hat{E}[e^{\beta^T X_i + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i}] I(s_j \leqslant Y_i < s_{j+1}) \Big].$$

It can be seen that the aforementioned recursive formulae can reduce the large number of parameters in the function of $A_{0k}(t)$ and $\Lambda_0(t)$ to several key parameters such as $A_{0k}(t_{m_k})$, $\Lambda_0(t_{m_T})$, $f_{1k}(t_{m_k})$, and $f_2(t_{m_T})$. We will obtain the updating equations for other parameters including $\left(\alpha_1^T, ..., \alpha_K^T, \beta^T, \phi^T, \gamma^T\right)$ by taking first derivatives of the log-likelihood function. For estimation of variances and covariances of the estimated parameters, one can use the inverse of the observed information matrix and follow the formula given in [22]. We sketch the required conditions and the validity of this approach in Appendix A.

2.3. Generalization to transformation models

Models (1) and (2) are Cox-type models, and it is well known that sometimes they may not fit the data well in practice. For this, a class of transformation models have been proposed and investigated under various contexts [15, 23, 24]. They can be written as

$$A_{ik}(t|X_i; b_{ik}) = H_k \left(\int_0^t e^{\alpha_k^T X_i(s) + b_{ik}^T \tilde{X}_i(s)} \mathrm{d}A_{0k}(s) \right) \tag{6}$$

and

$$\Lambda_i^T(t|X_i; b_i) = G \left(\int_0^t e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s) \right) \tag{7}$$

for the cumulative intensity function of $N_{ik}(t)$ and the cumulative hazard function of T_i, respectively, where H_k and G are known transformation functions. It is easy to see that these models include models (1) and (2) as special cases. Under them, the log-likelihood function has the form

$$\begin{aligned}\log L(O)=&\sum_i^n \log\int_b\left\{\prod_{k=1}^K\left[\prod_t\left\{a_{0k}(t)e^{\alpha_k^T X_i(t)+b_{ik}^T\tilde{X}_i(t)}H_k'\left(\int_0^t e^{\alpha_k^T X_i(s)+b_{ik}^T\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)\right\}^{R_{ik}(t)\Delta N_{ik}(t)}\right.\right.\\ &\left.\times\, e^{-H_k\left(\int_0^{Y_i} e^{\alpha_k^T X_i(s)+b_{ik}^T\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)}\right]\left\{\lambda_0(Y_i)e^{\beta^T X_i(Y_i)+\sum_k(\phi_k\circ b_{ik})^T\tilde{X}_i(Y_i)}\right.\\ &\left.\times\, G'\left(\int_0^{Y_i} e^{\beta^T X_i(s)+\sum_k(\phi_k\circ b_{ik})^T\tilde{X}_i(s)}\mathrm{d}\Lambda_0(s)\right)\right\}^{\Delta_i}\\ &\left.\times\, e^{-G\left(\int_0^{Y_i} e^{\beta^T X_i(s)+\sum_k(\phi_k\circ b_{ik})^T\tilde{X}_i(s)}\mathrm{d}\Lambda_0(s)\right)}\right\} f(b;\gamma)\mathrm{d}b,\end{aligned}$$

where g' denotes the derivative of g.

For estimation of the parameters θ, we can develop an EM algorithm similar to that given in the previous subsection, and in this situation, we have

$$\begin{aligned}\Gamma_i(b)=&\prod_{k=1}^K\left[\prod_t\left\{e^{b_{ik}^T\tilde{X}_i(t)}H_k'\left(\int_0^t e^{\alpha_k^T X_i(s)+b_{ik}^T\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)\right\}^{R_{ik}(t)\Delta N_{ik}(t)}\right.\\ &\left.\times\, e^{-H_k\left(\int_0^{Y_i} e^{\alpha_k^T X_i(s)+b_{ik}\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)}\right]\\ &\times\left[\left\{e^{\sum_k(\phi_k\circ b_{ik})^T\tilde{X}_i(Y_i)}G'\left(\int_0^{Y_i} e^{\beta^T X_i(s)+\sum_k(\phi_k\circ b_{ik})^T\tilde{X}_i(s)}\mathrm{d}\Lambda_0(s)\right)\right\}^{\Delta_i}\right.\\ &\left.\times\, e^{-G\left(\int_0^{Y_i} e^{\beta^T X_i(s)+\sum_k(\phi_k\circ b_{ik})^T\tilde{X}_i(s)}\mathrm{d}\Lambda_0(s)\right)}\right].\end{aligned}$$

The recursive formulae given in Equations (4) and (5) take the forms

$$\begin{aligned}\frac{1}{f_{1k}(t_{j,k})}=&\frac{1}{f_{1k}(t_{j+1,k})}-A_{0k}(t_{m_k,k})\sum_{i=1}^n\hat{E}\left[\int R_{ik}(t)\frac{H_k''\left(\int_0^t e^{\alpha_k^T X_i(s)+b_{ik}^T\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)}{H_k'\left(\int_0^t e^{\alpha_k^T X_i(s)+b_{ik}^T\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)}\right.\\ &\times\, e^{\alpha_k^T X_i(t_{j,k})+b_{ik}^T\tilde{X}_i(t_{j,k})}\times I(t\geqslant t_{j,k})\mathrm{d}N_{ik}(t)-H_k'\left(\int_0^{Y_i} e^{\alpha_k^T X_i(s)+b_{ik}^T\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)\\ &\times\, e^{\alpha_k^T X_i(t_{j,k})+b_{ik}^T\tilde{X}_i(t_{j,k})}I(t_{j+1,k}\leqslant Y_i)-\int R_{ik}(t)\frac{H_k''\left(\int_0^t e^{\alpha_k^T X_i(s)+b_{ik}^T\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)}{H_k'\left(\int_0^t e^{\alpha_k^T X_i(s)+b_{ik}^T\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)}\\ &\times\, e^{\alpha_k^T X_i(t_{j+1,k})+b_{ik}^T\tilde{X}_i(t_{j+1,k})}\times I(t\geqslant t_{j+1,k})\mathrm{d}N_{ik}(t)+H_k'\left(\int_0^{Y_i} e^{\alpha_k^T X_i(s)+b_{ik}^T\tilde{X}_i(s)}\mathrm{d}A_{0k}(s)\right)\\ &\left.\times\, e^{\alpha_k^T X_i(t_{j+1,k})+b_{ik}^T\tilde{X}_i(t_{j+1,k})}I(t_{j+1,k}\leqslant Y_i)\right]\end{aligned}$$

and

$$\frac{1}{f_2(s_j)} = \frac{1}{f_2(s_{j+1})} - \Lambda_0(t_{m_T}) \sum_{i=1}^{n} \hat{E} \left[\left(\Delta_i \frac{G''\left(\int_0^{Y_i} e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s)\right)}{G'\left(\int_0^{Y_i} e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s)\right)} \right.\right.$$
$$\times\, e^{\alpha_k^T X_i(s_j) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s_j)} - G'\left(\int_0^{Y_i} e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s)\right)$$
$$\left. \times\, e^{\beta^T X_i(s_j) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s_j)} \right) I(s_j \leqslant Y_i)$$
$$- \left(\Delta_i \frac{G''\left(\int_0^{Y_i} e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s)\right)}{G'\left(\int_0^{Y_i} e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s)\right)} e^{\alpha_k^T X_i(s_{j+1}) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s_{j+1})} \right.$$
$$\left. - G'\left(\int_0^{Y_i} e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0(s)\right) e^{\beta^T X_i(s_{j+1}) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s_{j+1})} \right)$$
$$\left. \times\, I(s_{j+1} \leqslant Y_i) \right],$$

respectively, where g'' represents the second derivative of g. The updating equations for other parameters can be obtained similarly. As before, their variances can be estimated by the inverse of the observed information matrix and the formula given in [22].

In addition to estimation of regression parameters, sometimes one may be also interested in estimating the cumulative rate function for each type of recurrent events. For this, note that

$$E(N_{ik}(t) | T_i > t, X_i] = \frac{\int_b H_k\left(\int_0^t e^{\alpha_k^T X_i(s) + b_{ik}^T \tilde{X}_i(s)} \mathrm{d}A_{0k}\right) e^{-G\left(\int_0^t e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0\right)} f(b;\gamma)\mathrm{d}b}{\int_b e^{-G\left(\int_0^t e^{\beta^T X_i(s) + \sum_k (\phi_k \circ b_{ik})^T \tilde{X}_i(s)} \mathrm{d}\Lambda_0\right)} f(b;\gamma)\mathrm{d}b}.$$

Thus, one can estimate it by its empirical estimate with the unknown parameters replaced by their nonparametric maximum likelihood estimators.

3. Simulation studies

We conducted extensive simulation studies to examine the performance of the proposed methods in practical settings. In the studies, we considered the situations with two types of recurrent events and $K = 2$. Similar to the setup in [15], we assumed that X_i is two dimensional and consists of a Bernoulli variable with 0.5 success probability and a uniform variable over $(-1, 1)$. The data on recurrent and terminal events were generated under models (6) and (7) with the same parameters of α for two types of recurrent events with $\alpha = \beta = (1, 0.5)^T$ and $\tilde{X}_i = 1$. Furthermore, it was assumed that b followed a bivariate normal distribution with mean 0 and the covariance matrix $\Sigma = [0.5, 0.2; 0.2, 0.5]$. The independent censoring time C_i was generated from the uniform distribution over $(0, \tau)$ with $\tau = 5$. We took $A_1(t) = \xi_1 t$, $A_2(t) = \xi_2 t$, and $\Lambda(t) = \xi_3 t^2$ with the values of ξ_1, ξ_2, and ξ_3 chosen to give approximately 35% censoring for the terminal event and 0.5 to 2 recurrent events for each type. The results given in the following text are based on 1000 replications.

Table I presents the results obtained for estimation of θ with $n = 200$, ϕ being equal to [−0.3, 0.5], [0, 0], or [0.3, 0.5], and $H_k(x) = G(x) = x$, which gives models (1) and (2). Here, for A_k and Λ, we give their estimates at three different time points $\tau/4$, $\tau/2$, and $3\tau/4$. The results include the estimated bias given by the average of the estimates minus the true value (BIAS), the sample standard error of the estimates (SE), the average of the estimated SEs (SEE), and the 95% empirical coverage probability (CP). Table II gives the results obtained under the same setup except for $n = 400$. These results indicate that all estimates are virtually unbiased, and the estimated SEs are close to the sample SEs. Also, the

Table I. Summary statistics for the simulation studies under $H(x)=G(x)=x$ for $n=200$.

	$\phi=[-0.3,0.5]$				$\phi=[0,0]$				$\phi=[0.3,0.5]$			
Parameters	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
α_1	0.001	0.127	0.126	0.95	−0.004	0.130	0.125	0.94	−0.004	0.131	0.126	0.94
α_2	−0.003	0.111	0.109	0.95	0.002	0.111	0.108	0.94	0.001	0.112	0.109	0.95
$\Lambda_1(\tau/4)$	0.005	0.121	0.123	0.95	0.010	0.122	0.122	0.94	0.011	0.127	0.123	0.94
$\Lambda_1(\tau/2)$	0.009	0.161	0.165	0.95	0.012	0.162	0.163	0.95	0.013	0.173	0.164	0.94
$\Lambda_1(3\tau/4)$	0.020	0.271	0.272	0.95	0.022	0.275	0.272	0.95	0.014	0.290	0.277	0.94
$\Lambda_2(\tau/4)$	0.006	0.120	0.122	0.96	0.009	0.128	0.122	0.94	0.004	0.124	0.122	0.95
$\Lambda_2(\tau/2)$	0.003	0.160	0.162	0.96	0.010	0.171	0.163	0.94	0.005	0.165	0.164	0.94
$\Lambda_2(3\tau/4)$	−0.003	0.276	0.273	0.95	0.018	0.281	0.271	0.94	0.007	0.286	0.279	0.94
β_1	0.021	0.192	0.194	0.95	0.017	0.184	0.184	0.95	0.011	0.213	0.206	0.95
β_2	0.005	0.170	0.166	0.95	0.011	0.165	0.158	0.93	0.005	0.176	0.176	0.95
$\Lambda(\tau/4)$	−0.007	0.069	0.068	0.94	−0.008	0.065	0.065	0.95	−0.005	0.072	0.071	0.95
$\Lambda(\tau/2)$	−0.005	0.112	0.110	0.94	−0.005	0.112	0.106	0.93	0.003	0.119	0.117	0.95
$\Lambda(3\tau/4)$	0.036	0.291	0.271	0.94	0.032	0.254	0.248	0.95	0.059	0.296	0.289	0.95
ϕ_1	−0.009	0.315	0.311	0.97	0.012	0.249	0.264	0.97	0.014	0.287	0.301	0.96
ϕ_2	0.029	0.327	0.329	0.96	−0.010	0.263	0.268	0.96	0.044	0.307	0.315	0.97
σ_{11}^2	−0.010	0.124	0.125	0.97	−0.006	0.123	0.125	0.97	−0.014	0.128	0.126	0.96
σ_{12}^2	−0.013	0.087	0.088	0.95	−0.014	0.086	0.088	0.95	−0.016	0.090	0.091	0.95
σ_{22}^2	−0.016	0.124	0.124	0.96	−0.013	0.124	0.124	0.96	−0.007	0.125	0.128	0.96

CP, coverage probability of the 95% confidence interval; SE, standard error; SEE, mean of the SE estimator. The confidence intervals for $\Lambda_1(\cdot)$, $\Lambda_2(\cdot)$, and $\Lambda(\cdot)$ are constructed on the basis of the log transformation, whereas the confidence intervals for σ_{11}^2 and σ_{22}^2 are based on the Satterthwaite approximation: $(\nu\hat{\sigma}^2/\chi^2_{\nu,0.975}, \nu\hat{\sigma}^2/\chi^2_{\nu,0.025})$, where $\nu=2\{\hat{\sigma}^2/\hat{se}(\hat{\sigma}^2)\}^2$ and $\chi^2_{\nu,\alpha}$ is the α-quantile of the χ^2 distribution with ν degrees of freedom.

Table II. Summary statistics for the simulation studies under $H(x)=G(x)=x$ for $n=400$.

	$\phi=[-0.3,0.5]$				$\phi=[0,0]$				$\phi=[0.3,0.5]$			
Parameters	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
α_1	−0.001	0.089	0.089	0.95	0.000	0.091	0.089	0.95	0.000	0.090	0.089	0.95
α_2	0.002	0.077	0.077	0.95	−0.001	0.077	0.077	0.94	−0.002	0.078	0.077	0.95
$\Lambda_1(\tau/4)$	0.004	0.086	0.086	0.95	−0.003	0.084	0.085	0.95	0.002	0.087	0.086	0.95
$\Lambda_1(\tau/2)$	0.005	0.113	0.115	0.96	−0.004	0.114	0.114	0.95	0.001	0.117	0.116	0.95
$\Lambda_1(3\tau/4)$	0.012	0.188	0.190	0.96	−0.005	0.189	0.190	0.96	0.005	0.195	0.195	0.95
$\Lambda_2(\tau/4)$	0.002	0.086	0.086	0.95	−0.001	0.087	0.086	0.94	0.000	0.086	0.086	0.95
$\Lambda_2(\tau/2)$	0.005	0.115	0.115	0.95	0.001	0.114	0.115	0.95	0.000	0.116	0.116	0.95
$\Lambda_2(3\tau/4)$	0.000	0.197	0.193	0.95	−0.002	0.188	0.190	0.95	−0.006	0.194	0.195	0.96
β_1	0.012	0.130	0.134	0.96	0.007	0.130	0.127	0.95	0.014	0.147	0.142	0.94
β_2	0.008	0.119	0.114	0.95	0.000	0.109	0.108	0.95	0.007	0.122	0.121	0.95
$\Lambda(\tau/4)$	−0.006	0.047	0.047	0.95	−0.003	0.047	0.045	0.95	−0.005	0.050	0.049	0.95
$\Lambda(\tau/2)$	−0.005	0.075	0.077	0.95	−0.002	0.075	0.074	0.95	−0.002	0.082	0.081	0.94
$\Lambda(3\tau/4)$	0.019	0.181	0.182	0.95	0.021	0.174	0.169	0.95	0.021	0.194	0.191	0.95
ϕ_1	0.000	0.196	0.197	0.96	0.003	0.167	0.172	0.96	0.003	0.177	0.187	0.96
ϕ_2	0.013	0.209	0.205	0.96	−0.001	0.171	0.171	0.95	0.020	0.200	0.198	0.95
σ_{11}^2	−0.004	0.089	0.088	0.95	−0.005	0.088	0.088	0.95	0.000	0.088	0.090	0.96
σ_{12}^2	−0.011	0.063	0.063	0.94	−0.010	0.061	0.063	0.95	−0.007	0.062	0.065	0.96
σ_{22}^2	−0.004	0.086	0.088	0.96	−0.002	0.084	0.088	0.95	−0.001	0.085	0.090	0.97

See the note in Table I.

normal distribution approximation seems to perform well, and as expected, the results become better when the sample size increases.

Following a reviewer's suggestion, we also investigated the performance of the naive approach that ignores the dependence between the two types of recurrent events and the dependence between the recurrent and terminal events. For this, we estimated the regression parameters in models (1) and (2) by the mean and rate-based approach given in [8] and the Cox model, respectively, and we present some of the results in Table III. They indicated that as expected, the naive approach is clearly not appropriate.

We obtained the results given previously under models (1) and (2). For models (6) and (7), we considered the situation with $H(x) = x$ and $G(x) = \log(x+1)$, and the results obtained are given in Tables IV and V for $n = 200$ and $n = 400$, respectively. Here, all other setups were the same as in Tables I and II. Tables VI and VII present the estimation results for models (6) and (7) with $H_1(x) = \log(x+1)$, $H_2 = x$, and $G(x) = \log(x+1)$. It can be seen that the results give similar conclusions as those by Tables I and II. For all cases, the proposed approach could overestimate ϕ and some variance parameters for small sample sizes.

Table III. Summary statistics for simulation studies under $H(x)=G(x)=x$ based on the naive approach for $n=200$ and $n=400$.

	$\phi = [-0.3, 0.5]$				$\phi = [0, 0]$				$\phi = [0.3, 0.5]$			
θ	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
					$n = 200$							
α_{11}	0.018	0.178	0.113	0.785	0.001	0.173	0.114	0.813	−0.028	0.160	0.119	0.858
α_{12}	0.000	0.150	0.097	0.794	−0.004	0.153	0.098	0.789	−0.021	0.144	0.101	0.819
α_{21}	−0.032	0.167	0.118	0.820	−0.006	0.176	0.114	0.797	−0.039	0.173	0.120	0.800
α_{22}	−0.013	0.148	0.101	0.814	0.004	0.154	0.098	0.795	−0.021	0.143	0.103	0.842
β_1	−0.028	0.185	0.177	0.945	0.004	0.174	0.177	0.952	−0.068	0.175	0.177	0.938
β_2	−0.009	0.150	0.152	0.952	0.012	0.158	0.153	0.943	−0.028	0.153	0.152	0.948
					$n = 400$							
α_{11}	0.005	0.126	0.079	0.781	−0.003	0.125	0.080	0.797	−0.034	0.120	0.083	0.795
α_{12}	0.007	0.112	0.068	0.764	0.005	0.107	0.068	0.788	−0.017	0.104	0.071	0.820
α_{21}	−0.026	0.113	0.083	0.833	0.008	0.124	0.080	0.795	−0.043	0.115	0.084	0.808
α_{22}	−0.016	0.104	0.071	0.821	0.004	0.107	0.068	0.791	−0.016	0.101	0.072	0.833
β_1	−0.030	0.121	0.124	0.941	−0.001	0.126	0.124	0.940	−0.058	0.128	0.124	0.926
β_2	−0.015	0.109	0.107	0.939	0.002	0.110	0.106	0.948	−0.026	0.113	0.107	0.937

α_{11} and α_{12} are for the first type of recurrent event, and α_{21} and α_{22} are for the second type of recurrent event.

Table IV. Summary statistics for the simulation studies under $H(x)=x$ and $G(x)=\log(x+1)$ for $n=200$.

	$\phi = [-0.3, 0.5]$				$\phi = [0, 0]$				$\phi = [0.3, 0.5]$			
Parameters	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
α_1	0.002	0.129	0.129	0.95	−0.006	0.134	0.129	0.94	−0.003	0.128	0.130	0.95
α_2	−0.001	0.112	0.111	0.95	0.002	0.108	0.112	0.95	−0.003	0.116	0.112	0.94
$A_1(\tau/4)$	0.004	0.125	0.129	0.95	0.007	0.128	0.129	0.95	0.007	0.132	0.130	0.95
$A_1(\tau/2)$	0.008	0.174	0.175	0.95	0.009	0.174	0.176	0.95	0.008	0.179	0.177	0.95
$A_1(3\tau/4)$	0.013	0.287	0.283	0.94	0.022	0.291	0.286	0.95	0.009	0.291	0.287	0.95
$A_2(\tau/4)$	−0.002	0.129	0.128	0.95	0.012	0.134	0.129	0.94	0.008	0.129	0.130	0.95
$A_2(\tau/2)$	−0.006	0.175	0.173	0.95	0.021	0.179	0.177	0.95	0.010	0.175	0.177	0.96
$A_2(3\tau/4)$	−0.011	0.274	0.282	0.96	0.036	0.291	0.288	0.95	0.014	0.293	0.289	0.95
β_1	0.010	0.283	0.283	0.95	0.016	0.280	0.280	0.95	0.005	0.283	0.290	0.96
β_2	0.002	0.245	0.244	0.95	0.009	0.241	0.241	0.95	0.006	0.256	0.250	0.94
$\Lambda(\tau/4)$	0.004	0.173	0.172	0.96	0.003	0.167	0.169	0.96	0.010	0.178	0.177	0.96
$\Lambda(\tau/2)$	0.037	0.324	0.323	0.96	0.038	0.322	0.319	0.95	0.058	0.348	0.338	0.95
$\Lambda(3\tau/4)$	0.170	0.899	0.904	0.96	0.162	0.857	0.884	0.96	0.214	0.951	0.941	0.96
ϕ_1	−0.008	0.506	0.548	0.97	−0.006	0.508	0.512	0.98	0.003	0.545	0.565	0.98
ϕ_2	0.003	0.518	0.567	0.96	0.029	0.531	0.521	0.97	0.039	0.519	0.566	0.98
σ_{11}^2	−0.007	0.134	0.128	0.96	−0.006	0.125	0.129	0.96	−0.010	0.132	0.131	0.97
σ_{12}^2	−0.014	0.093	0.092	0.95	−0.015	0.092	0.091	0.93	−0.008	0.093	0.094	0.95
σ_{22}^2	−0.010	0.133	0.129	0.96	−0.014	0.123	0.128	0.96	−0.005	0.131	0.132	0.97

See the note in Table I.

Table V. Summary statistics for the simulation studies under $H(x)=x$ and $G(x)=\log(x+1)$ for $n=400$.

	$\phi=[-0.3, 0.5]$				$\phi=[0,0]$				$\phi=[0.3, 0.5]$			
Parameters	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
α_1	−0.007	0.089	0.091	0.96	0.003	0.093	0.091	0.95	−0.004	0.091	0.091	0.94
α_2	0.003	0.081	0.079	0.94	−0.004	0.080	0.079	0.94	0.001	0.080	0.079	0.95
$A_1(\tau/4)$	0.004	0.089	0.091	0.95	−0.002	0.102	0.091	0.95	0.004	0.095	0.092	0.94
$A_1(\tau/2)$	0.005	0.123	0.124	0.95	−0.004	0.137	0.123	0.95	0.005	0.127	0.125	0.94
$A_1(3\tau/4)$	0.010	0.199	0.200	0.95	−0.004	0.217	0.200	0.95	0.011	0.205	0.204	0.95
$A_2(\tau/4)$	0.006	0.093	0.091	0.94	−0.001	0.101	0.091	0.95	0.002	0.093	0.092	0.96
$A_2(\tau/2)$	0.009	0.125	0.124	0.94	−0.002	0.138	0.124	0.95	0.000	0.126	0.125	0.95
$A_2(3\tau/4)$	0.016	0.206	0.202	0.95	−0.007	0.218	0.200	0.94	−0.001	0.203	0.203	0.95
β_1	0.010	0.189	0.197	0.95	0.008	0.201	0.193	0.94	−0.001	0.205	0.201	0.95
β_2	0.006	0.171	0.169	0.95	0.013	0.170	0.167	0.95	0.010	0.167	0.173	0.95
$\Lambda(\tau/4)$	−0.004	0.117	0.118	0.95	−0.005	0.117	0.116	0.95	0.005	0.128	0.122	0.94
$\Lambda(\tau/2)$	0.008	0.221	0.219	0.95	0.010	0.219	0.215	0.96	0.021	0.233	0.226	0.94
$\Lambda(3\tau/4)$	0.075	0.606	0.603	0.96	0.069	0.588	0.585	0.96	0.082	0.623	0.613	0.95
ϕ_1	−0.007	0.353	0.344	0.96	0.001	0.328	0.330	0.95	0.008	0.349	0.353	0.96
ϕ_2	0.002	0.353	0.359	0.95	−0.012	0.328	0.331	0.96	−0.009	0.358	0.360	0.95
σ_{11}^2	−0.004	0.089	0.090	0.96	0.000	0.093	0.095	0.944	−0.001	0.089	0.092	0.95
σ_{12}^2	−0.011	0.063	0.065	0.95	−0.009	0.065	0.064	0.95	−0.005	0.065	0.066	0.95
σ_{22}^2	−0.005	0.092	0.091	0.96	−0.002	0.093	0.091	0.95	0.002	0.091	0.093	0.95

See the note in Table I.

Table VI. Summary statistics for the simulation studies under $H_1(x)=\log(x+1)$, $H_2(x)=x$ and $G(x)=\log(x+1)$ for $n=200$.

	$\phi=[-0.3, 0.5]$				$\phi=[0,0]$				$\phi=[0.3, 0.5]$			
Parameters	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
α_1	0.003	0.149	0.143	0.94	0.008	0.140	0.143	0.95	−0.003	0.143	0.142	0.95
α_2	0.001	0.130	0.124	0.94	−0.008	0.124	0.123	0.95	−0.001	0.127	0.124	0.94
$A_1(\tau/4)$	−0.019	1.237	1.232	0.94	0.010	1.214	1.232	0.95	0.030	1.225	1.235	0.96
$A_1(\tau/2)$	0.000	1.756	1.761	0.95	0.033	1.731	1.762	0.95	0.032	1.763	1.756	0.95
$A_1(3\tau/4)$	0.068	3.213	3.111	0.95	0.071	3.082	3.108	0.95	−0.045	3.030	3.094	0.95
$A_2(\tau/4)$	0.003	0.134	0.132	0.94	0.008	0.131	0.132	0.95	0.002	0.131	0.132	0.96
$A_2(\tau/2)$	−0.001	0.178	0.178	0.95	0.010	0.177	0.179	0.95	0.000	0.176	0.178	0.96
$A_2(3\tau/4)$	0.001	0.286	0.286	0.94	0.019	0.282	0.287	0.96	0.003	0.285	0.288	0.96
β_1	0.001	0.291	0.287	0.94	0.018	0.275	0.284	0.95	−0.008	0.296	0.293	0.95
β_2	0.004	0.236	0.247	0.95	0.001	0.244	0.245	0.95	0.007	0.251	0.253	0.95
$\Lambda(\tau/4)$	0.010	0.163	0.158	0.96	0.000	0.158	0.154	0.94	−0.001	0.165	0.162	0.95
$\Lambda(\tau/2)$	0.046	0.309	0.295	0.95	0.042	0.301	0.290	0.95	0.023	0.298	0.299	0.95
$\Lambda(3\tau/4)$	0.195	0.852	0.834	0.96	0.185	0.820	0.813	0.95	0.210	0.867	0.861	0.96
ϕ_1	0.014	0.671	0.837	0.99	−0.015	0.713	0.838	0.99	0.040	0.644	0.798	0.99
ϕ_2	0.027	0.582	0.666	0.98	0.005	0.578	0.639	0.98	0.037	0.546	0.641	0.98
σ_{11}^2	0.023	0.239	0.300	0.95	0.025	0.251	0.299	0.94	0.032	0.236	0.299	0.94
σ_{12}^2	−0.017	0.126	0.138	0.95	−0.023	0.129	0.138	0.96	−0.030	0.124	0.138	0.96
σ_{22}^2	−0.008	0.128	0.128	0.96	−0.014	0.126	0.126	0.97	−0.020	0.127	0.127	0.96

See the note in Table I.

4. Applications

In this section, we apply the methodology proposed in the previous sections to the CCSS described previously. As mentioned before, the study followed the children with cancer and surviving at least 5 years after the cancer treatment from 26 collaborating institutions. The diagnoses include leukemia, central nervous system (CNS) malignancy, Hodgkin lymphoma, non-Hodgkin lymphoma, Wilm's tumor, neuroblastoma, soft-tissue sarcoma, and bone tumors. During the follow-up, detailed information was

Table VII. Summary statistics for the simulation studies under $H_1(x) = \log(x+1)$, $H_2(x) = x$ and $G(x) = \log(x+1)$ for $n = 400$.

	$\phi = [-0.3, 0.5]$				$\phi = [0, 0]$				$\phi = [0.3, 0.5]$			
Parameters	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
α_1	0.003	0.097	0.101	0.95	−0.003	0.100	0.101	0.95	0.009	0.099	0.101	0.95
α_2	0.006	0.083	0.087	0.95	0.003	0.088	0.087	0.94	−0.001	0.088	0.087	0.95
$A_1(\tau/4)$	0.025	0.845	0.872	0.95	0.038	0.871	0.874	0.96	−0.006	0.893	0.868	0.95
$A_1(\tau/2)$	0.019	1.182	1.241	0.95	0.044	1.239	1.245	0.95	−0.003	1.263	1.236	0.94
$A_1(3\tau/4)$	0.086	2.095	2.187	0.95	0.082	2.188	2.193	0.95	−0.023	2.205	2.159	0.95
$A_2(\tau/4)$	0.001	0.097	0.093	0.94	0.004	0.091	0.093	0.96	0.008	0.098	0.094	0.93
$A_2(\tau/2)$	−0.008	0.123	0.125	0.95	0.006	0.122	0.126	0.96	0.013	0.131	0.127	0.94
$A_2(3\tau/4)$	−0.023	0.195	0.199	0.96	0.001	0.194	0.201	0.96	0.006	0.206	0.203	0.93
β_1	0.006	0.205	0.199	0.95	0.010	0.189	0.195	0.96	0.013	0.204	0.204	0.95
β_2	0.005	0.176	0.171	0.95	0.011	0.170	0.168	0.94	0.002	0.175	0.175	0.96
$\Lambda(\tau/4)$	−0.004	0.104	0.108	0.96	0.000	0.107	0.106	0.95	−0.004	0.107	0.110	0.95
$\Lambda(\tau/2)$	0.014	0.191	0.198	0.96	0.023	0.197	0.196	0.96	0.009	0.204	0.203	0.94
$\Lambda(3\tau/4)$	0.087	0.545	0.547	0.95	0.082	0.523	0.528	0.95	0.049	0.537	0.546	0.94
ϕ_1	0.001	0.509	0.547	0.98	0.007	0.452	0.488	0.98	−0.010	0.482	0.521	0.98
ϕ_2	−0.004	0.361	0.420	0.97	0.001	0.385	0.389	0.97	0.039	0.393	0.411	0.97
σ_{11}^2	0.004	0.175	0.212	0.95	0.016	0.164	0.213	0.95	0.016	0.176	0.212	0.95
σ_{12}^2	−0.008	0.092	0.098	0.95	−0.011	0.092	0.098	0.96	−0.012	0.096	0.097	0.95
σ_{22}^2	0.005	0.097	0.090	0.93	−0.005	0.089	0.089	0.95	−0.008	0.088	0.090	0.96

See the note in Table I.

collected via questionnaires on occurrence of the original or new cancer diagnoses more than 5 years after the primary diagnosis. If the survivor died during the study, the date of death was obtained from the death certificate. One objective of this study is to understand the risk factors for recurrences and new subsequent cancers, which facilitates surveillance and early prevention strategies.

In the following, we will consider two subgroups, leukemia survivors ($n = 4033$) and CNS malignancy survivors ($n = 1562$). For both cases, the covariates considered include age at diagnosis (0, < 10 years old; 1, $\geqslant 10$ years old), sex (0, female; 1, male), radiation exposure (0, no; 1, yes), total dose of alkylating agents [25] (score 0, alk1 = 0, alk2 = 0; score 1–2, alk1 = 1, alk2 = 0; score 2–4, alk1 = 0, alk2 = 1), and family history of cancer (0, no; 1, yes). The recurrence of original cancers will be treated as type I ($k = 1$) recurrent event, and the occurrence of new cancers will be treated as type II ($k = 2$) recurrent event. To find appropriate models for recurrent event processes and the terminal or death process, following [15], we will take $H_k(x)$ and $G(x)$ to be $r^{-1}\log(1+rx)$, the logarithmic transformation, or $\rho^{-1}\{(1+x)^{\rho}-1\}$, the Box–Cox transformation. We restrict the values of the rs and ρ between 0 and 1 with a grid of 0.1 and select the best models based on the log-likelihood value.

First, we applied the proposed methodology to the leukemia survivors. For the patients in this subgroup, the range of the number of recurrences of original cancers is between 0 and 2 per person with the average being 0.07, whereas the range of the number of new cancers is between 0 and 18 per person with an average of 0.13. By applying the model selection procedure described previously, the best models were given by models (1) and (2). Table VIII shows the estimated effects for the six covariates of interest. They suggest that the recurrence of the original cancer appears to be positively significantly related to all covariates except the age at diagnosis. For the occurrence of new cancers, all covariates except the gender had significant effects, whereas death was significantly associated with all covariates. In particular, radiation exposure, a high dose of alkylating agents, and family history of cancer appeared to increase the risk of the recurrences of original cancer, new cancers, and death. Patients who were older at diagnosis tend to have a higher risk for both new cancers and death than those who were younger at diagnosis. Male leukemia survivors were more likely to experience the recurrences of the original cancer and to have shorter life than female leukemia survivors.

To check the goodness-of-fit of model (1) to the observed data on leukemia survivors, we obtained the estimated cumulative rate functions given in Section 3 for the occurrences of both original cancers and

Table VIII. Estimation results for the Childhood Cancer Survivor Study.

		Leukemia survivors			CNS malignancy survivors		
Event	Covariate	Estimate	Standard error	*p*-value	Estimate	Standard error	*p*-value
Recurrences of original cancer	Age@dx	0.166	0.172	0.336	0.073	0.201	0.716
	Sex	0.311	0.143	0.029	−0.031	0.190	0.870
	RT	1.160	0.204	0.000	0.682	0.242	0.005
	Alk1	1.036	0.164	0.000	0.684	0.277	0.014
	Alk2	2.314	0.242	0.000	1.102	0.339	0.001
	FH	0.484	0.199	0.015	−0.102	0.271	0.706
New cancer	Age@dx	0.691	0.156	0.000	0.663	0.212	0.002
	Sex	−0.068	0.134	0.612	−0.247	0.207	0.233
	RT	1.892	0.226	0.000	1.604	0.315	0.000
	Alk1	0.469	0.149	0.002	0.755	0.282	0.007
	Alk2	1.405	0.268	0.000	0.872	0.437	0.046
	FH	0.752	0.174	0.000	0.408	0.253	0.107
Death	Age@dx	1.264	0.409	0.002	0.445	0.290	0.124
	Sex	1.062	0.341	0.002	0.311	0.244	0.203
	RT	1.997	0.563	0.000	1.961	0.552	0.000
	Alk1	2.484	0.575	0.000	1.253	0.540	0.020
	Alk2	5.770	1.177	0.000	2.634	0.734	0.000
	FH	1.148	0.504	0.023	−0.276	0.329	0.401

Age@dx, age at diagnosis; Alk, alkylating agent; FH, family history; RT, radiation therapy; CNS, central nervous system.

new cancers and plotted them in Figure 1 (top half). Also included in the same plots are the empirical estimates of the two cumulative rate functions given by

$$\frac{\sum_{i=1}^{n} I(Y_i \geqslant t)\, N_{ik}(t)}{\sum_{i=1}^{n} I(Y_i \geqslant t)}.$$

To check model (2), the bottom half of Figure 1 displays the estimated survival function of the death time and the Kaplan–Meier estimate of the same survival function. Figure 1 suggests that both models (1) and (2) fit the data well.

For the patients in the CNS malignancy survivor group, the number of recurrences of original cancers ranged from 0 to 5 with an average being 0.12, whereas the number of occurrences of new cancers ranged from 0 to 14 with an average of 0.15. For this set of the observed data, the model selection procedure described previously resulted in $H_1(x) = 0.2^{-1}\log(1+0.2x)$, $H_2(x) = x$, and $G(x) = x$, and the estimated covariate effects are also presented in Table VIII. The results indicate that for the CNS malignancy survivors, the recurrence of original cancers were significantly related to radiation exposure and the dose of alkylating agents but not to age, gender, and family history of cancer. The new cancers were related to the covariates similarly except that the occurrence of new cancers seems to be significantly affected by age at the diagnosis. In other words, the CNS malignancy survivors older at diagnosis seem to experience significantly higher occurrence rate of new cancers than those who were younger at diagnosis, which was not the case for recurrence rates of original cancers. In terms of the death rate for the CNS malignancy survivors, its relationship with all covariates is similar to that between the recurrence of original cancers and the covariates.

As for the leukemia survivors, we also obtained the estimated cumulative rate functions for the occurrences of both original cancers and new cancers in CNS malignancy survivors and plotted them in Figure 2 (top half) along with the corresponding empirical estimates. The bottom half of Figure 2 shows the estimated survival functions under models (6) and (7) along with the Kaplan–Meier estimates, suggesting that these models fit the data on the CNS malignancy survivors well.

For both examples, we provide the estimated cumulative rate and survival functions based on the naive approach mentioned previously. As we can see, the naive estimates do not fit the data as well as the proposed method, especially for the recurrence of original cancer.

Statistics
in Medicine

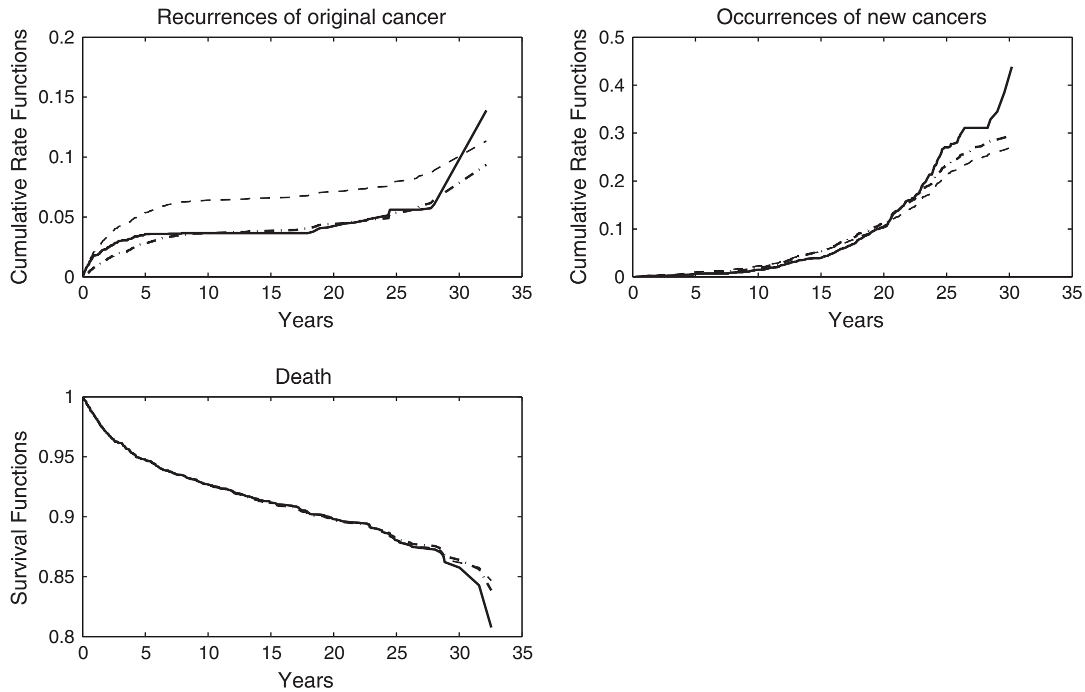

Figure 1. Estimates from the leukemia group in the Childhood Cancer Survivor Study. The upper panel corresponds to cumulative rate functions for two types of cancers, and the lower panel corresponds to survival functions. The solid and dash dot curves pertain to the nonparametric and model-based estimates, respectively. The dash curves pertain to the naive approach-based estimates.

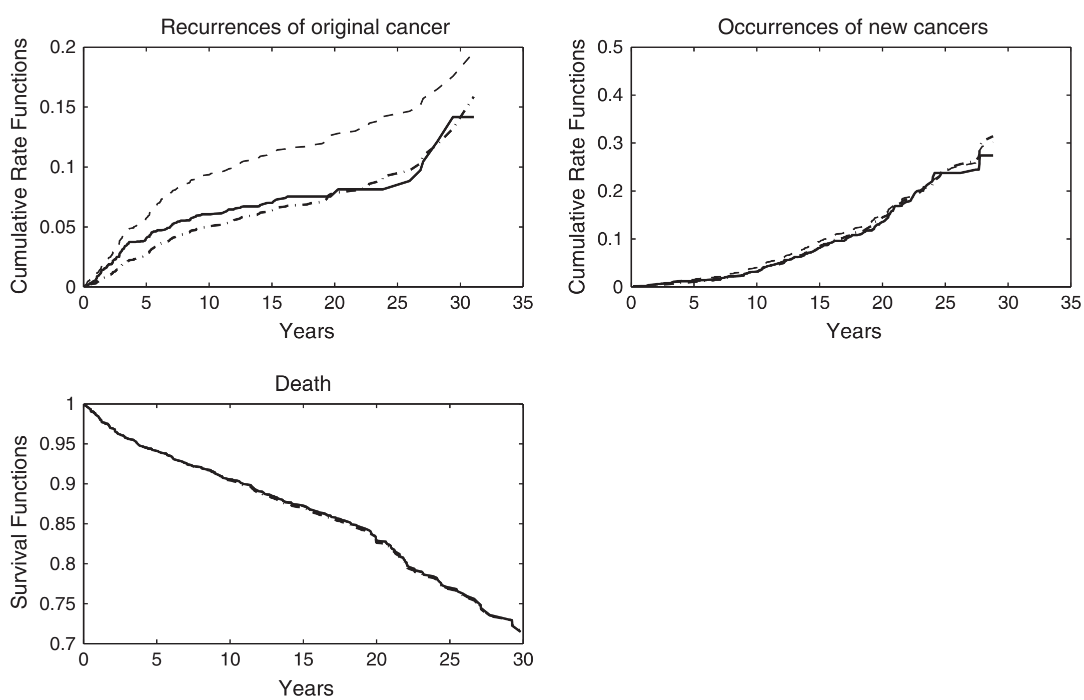

Figure 2. Estimates from the central nervous system malignancy group in the Childhood Cancer Survivor Study. The upper panel corresponds to cumulative rate functions for two types of cancers, and the lower panel corresponds to survival functions. The solid and dash dot curves pertain to the nonparametric and model-based estimates, respectively. The dash curves pertain to the naive approach-based estimates.

5. Discussion

In the previous sections, we discussed regression analysis of multivariate recurrent event data in the presence of terminal events. For the problem, we developed a joint modeling approach by following that proposed in [15] for univariate recurrent event data. The approach directly deals with the correlation between different types of recurrent events and that between recurrent and terminal events simultaneously. Simulation studies indicated that the proposed method seems to work well for practical situations, although a relatively larger sample size may be needed for better estimation of parameter ϕ under transformation models. We applied the proposed method to two real data sets from the CCSS, and the graphical tools used suggested that the the proposed models fitted the data well.

As discussed previously, for the problem considered here, we took the joint modeling approach. A main advantage of this method is that it allows one to directly model and estimate the correlation structure between different types of recurrent events as well as the correlation structure between recurrent events and terminal events. A possible drawback is that sometimes it may be very difficult to verify the assumed models. An alternative is to apply the marginal model approach [17].

Appendix A. Proof for the asymptotic efficiency of maximum likelihood estimator

In this section, we show that the maximum likelihood estimates $\left(\hat{\alpha}_1^T, ..., \hat{\alpha}_K^T, \hat{\beta}^T, \hat{\phi}^T, \hat{\gamma}^T\right)^T$ are asymptotically efficient and that the inverse of the observed information matrix is a consistent estimator of the limiting covariance matrix. For this, we need the following conditions:

(C1) The parameter value $(\alpha_{01}^T, ..., \alpha_{0K}^T, \beta_0^T, \phi_0^T, \gamma_0^T)^T$ belongs to the interior of a compact set Θ in a finite dimensional Euclidean space, and $A_{0k}'(t) > 0$ and $\Lambda_0'(t) > 0$ for all $t \in [0, \tau]$.

(C2) With probability 1, $X_i(t)$ is left-continuous with uniformly bounded left and right derivatives in $[0, \tau]$.

(C3) With probability 1, $P(C_i \geqslant \tau | X_i) > \delta_0 > 0$ for some constant δ_0.

(C4) With probability 1, $\sup_k E[N_{ik}(\tau)] < \infty$.

(C5) For any constant $c_0 > 0$ and for $k = 1, \cdots, K$, $\sup_\gamma E \int_b \exp\{c0(N_{ik}(\tau) + 1)|b|\} f(b; \gamma_0)\mathrm{d}b < \infty$; and $\left|\frac{\partial f(b;\gamma)/\partial\gamma}{f(b;\gamma)}\right| + \left|\frac{\partial^2 f(b;\gamma)/\partial^2\gamma}{f(b;\gamma)}\right| + \left|\frac{\partial^3 f(b;\gamma)/\partial^3\gamma}{f(b;\gamma)}\right| \leqslant \exp\{c0(1 + |b|)\}$;

(C6) For any fixed $k = 1, \cdots, K$, if there exists $c(t)$ and v such that $c(t) + v^T X_{ik}(t) = 0$ with probability 1, then $c(t) = 0$ and $v = 0$. In addition, there exists some $t \in [0, \tau]$ such that $\tilde{X}_k(t)$ spans the whole space of b_k.

(C7) $f(b; \gamma) = f(b; \gamma_0)$ almost surely if and only if $\gamma = \gamma_0$. In addition, if $v^T f'(b; \gamma_0) = 0$ almost surely, then $v = 0$.

One can note that our conditions are almost similar to those in [15] except their condition (D5), which naturally holds in our context.

Following [26] and [15], it suffices to prove the two identifiability conditions (C5) and (C7) in [26]. First, we verify that if $L(O; \theta) = L(O; \theta_0)$ almost surely, then $\theta = \theta_0$. Following the verification of (C5) in [15], we perform operations (1) and (2) in their paper on the likelihood function over the recurrent events for type k, $k = 1, \cdots, K$, we obtain that

$$\int_b \exp\left\{\sum_{k=1}^{K}\sum_{j=1}^{m_k} i\omega_{jk} Q_{1k}(t_{1kj}; b)\right\} \exp\{-Q_2(t_2; b)\} f(b; \gamma)\mathrm{d}b$$
$$= \int_b \exp\left\{\sum_{k=1}^{K}\sum_{j=1}^{m_k} i\omega_{jk} Q_{01k}(t_{1kj}; b)\right\} \exp\{-Q_{02}(t_2; b)\} f(b; \gamma_0)\mathrm{d}b,$$

where $Q_{1k}(t; b) = \int_0^t \exp\left(\alpha_k^T X(s)\ (+b_k^T \tilde{X}(s)\right) \mathrm{d}A_k(s)$ and $Q_2(t; b) = \int_0^t \exp\left(\beta^T X(s)\right) + \sum_k (\phi \circ b_k)^T \tilde{X}(s)\big) \mathrm{d}\Lambda(s)$, Q_{01k} and Q_{02} are Q_{1k} and Q_2 evaluated at the true θ_0. Here, i is the imaginary unit satisfying $i^2 = -1$.

Letting $t_2 = 0$ and for $k = 1, \cdots, K$, letting $t_{1k'j} = 0$ for $k' \neq k$, one obtains that $Q_{1k}(t_{1kj}; b_1)(k = 1, \cdots, K, j = 1, \cdots m_k)$ as a function of $b_1 \sim f(b, \gamma)$ has the same distribution as $Q_{01k}(t_{1kj}; b)(k =$

$1, \cdots, K, j = 1, \cdots m_k$) with $b \sim f(b; \gamma_0)$, respectively. Then, following the proof in [15], one can prove that $\theta = \theta_0$.

It follows similarly that (C7) in [26] also holds in this context from the similar arguments for the proof of (C5) and the verification of (C7) in [15].

References

1. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer-Verlag: New York, 1993.
2. Cook RJ, Lawless JF. Interim monitoring of longitudinal comparative studies with recurrent event responses. *Biometrics* 1996; **52**:1311–1323.
3. Cook RJ, Lawless JF. *The Analysis of Recurrent Event Data*. Springer-Verlag: New York, 2007.
4. Cook RJ, Lawless JF, Nadeau JC. Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* 1996; **52**:557–571.
5. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* 1982; **10**:1100–1120.
6. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika* 1981; **68**:373–379.
7. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995; **37**:158–168.
8. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate function of recurrent events. *Journal of Royal Statistical Society, Series B* 2000; **69**:711–730.
9. Lin DY, Wei LJ, Ying Z. Semiparametric transformation models for point processes. *Journal of the American Statistical Association* 2001; **96**:620–628.
10. Robison L, Mertens A, Boice J, et al. Study design and cohort characteristics of the childhood cancer survivor study: a multi-institutional collaborative project. *Medical and Pediatric Oncology* 2002; **38**(4):229–239.
11. Robison L, Armstrong G, Boice J, et al. The Childhood Cancer Survivor Study: a National Cancer Institute-supported resource for outcome and intervention research. *Journal of Clinical Oncology* 2009; **27**(14):2308–2318.
12. Wang MC, Qin J, Chiang CT. Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* 2001; **96**:1057–1065.
13. Huang CY, Wang MC. Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* 2004; **99**:1153–1165.
14. Ye Y, Kalbfleisch JD, Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* 2007; **63**:78–87.
15. Zeng D, Lin DY. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics* 2009; **65**:746–752.
16. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics* 2000; **56**:554–562.
17. Ghosh D, Lin DY. Marginal regression models for recurrent and terminal events. *Statistica Sinica* 2002; **12**:663–688.
18. Cai J, Schaubel DE. Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis* 2004; **10**:121–138.
19. Clegg LX, Cai J, Sen PK. A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics* 1999; **55**:805–812.
20. Schaubel DE, Cai J. Rate/mean regression for multiple-sequence recurrent event data with missing event category. *Scandinavian Journal of Statistics* 2006; **33**:191–207.
21. Spiekerman CF, Lin DY. Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* 1998; **93**:1164–1175.
22. Louis TA. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* 1982; **44**:226–233.
23. Cheng SC, Wei LJ, Ying Z. Analysis of transformation models with censored data. *Biometrika* 1995; **83**:835–846.
24. Zhang Z, Sun L, Zhao X, Sun J. Regression analysis of interval-censored failure time data with linear transformation models. *The Canadian Journal of Statistics* 2005; **33**:1–10.
25. Tucker MA, D'Angio GJ, Boice JD, et al. Bone sarcomas linked to radiotherapy and chemotherapy in children. *New England Journal of Medicine* 1987; **317**:588–593.
26. Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *Journal of the Royal Statistical Society, Series B* 2007; **69**:507–564.