

# Analysis of bivariate recurrent event data with zero inflation

Taeun Kim<sup>a</sup>, Yang-Jin Kim<sup>1,a</sup>

<sup>a</sup>Department of Statistics, Sookmyung Women's University, Korea

---

## Abstract

Recurrent event data frequently occur in clinical studies, demography, engineering reliability and so on (Cook and Lawless, *The Statistical Analysis of Recurrent Events*, Springer, 2007). Sometimes, two or more different but related type of recurrent events may occur simultaneously. In this study, our interest is to estimate the covariate effect on bivariate recurrent event times with zero inflations. Such zero inflation can be related with susceptibility. In the context of bivariate recurrent event data, furthermore, such susceptibilities may be different according to the type of event. We propose a joint model including both two intensity functions and two cure rate functions. Bivariate frailty effects are adopted to model the correlation between recurrent events. Parameter estimates are obtained by maximizing the likelihood derived under a piecewise constant hazard assumption. According to simulation results, the proposed method brings unbiased estimates while the model ignoring cure rate models gives underestimated covariate effects and overestimated variance estimates. We apply the proposed method to a set of bivariate recurrent infection data in a study of child patients with leukemia.

Keywords: bivariate recurrent event, cure rate model, frailty effects, joint model, zero inflation

---

## 1. Introduction

Recurrent event data are frequently encountered in clinical and observational studies. Examples include tumor recurrences, recurrent heart attacks, repeated hospitalization and insurance claims and so on. Such data can provide a history of the symptom of interest and be utilized to evaluate the long term effects of covariates. To analyze recurrent event data, various methods have been proposed for either the intensity function (Andersen and Gill, 1982) and the marginal rate function (Wei *et al.*, 1989). Cook and Lawless (2007) provided a comprehensive review of the analysis about recurrent event data. A bivariate recurrent event arises when two different types of events repeatedly occur. Like other types of multivariate failure times, a frailty effect approach (Cook *et al.*, 2010) and a marginal approach (Cai and Schaubel, 2004) are applied to analyze bivariate recurrent event data.

Sometimes, a portion of subjects would not experience recurrent events of interest and these subjects might be regarded as cured or non-susceptible. A similar situation is discussed in a context of a failure time data. A cure rate model assumes there exists a risk-free group explaining the Kaplan-Meier survival curve with a long and stable plateau at the tail. An introduction of cure rate model makes it possible to estimate the effect of covariate on both the incidence as well as the latency of the event. Sy and Taylor (2000) applied a proportional hazard model and adopted an EM algorithm to recover the unknown susceptibility of censored subjects. Peng and Dear (2000) considered a frailty

---

<sup>1</sup> Corresponding author: Department of Statistics, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul 01910, Korea. E-mail: [yjin@sookmyung.ac.kr](mailto:yjin@sookmyung.ac.kr)

effect for a cure probability and Kim and Jhun (2008) also applied a frailty effect to interval censored failure time with cure proportion and Kim (2017) extended to bivariate interval censored data. There have been several recent studies using joint models that simultaneously consider both zero-inflated recurrent events and terminal events. For example, Liu *et al.* (2016) considered two different models according to the dependency structure between a recurrent event and a terminal event and Zhu *et al.* (2018) suggested a Bayesian approach for a joint model. However, we could not find studies about bivariate recurrent event data with zero inflation.

The motivation for this study is infection record data obtained from 210 child patients with acute myeloid leukemia (AML). During the chemotherapy course, patients had been repeatedly infected by bacterial, viral or fungal infection. The causes are classified into two types, the bacterial infection and the fungal or viral, as examples of a bivariate recurrent event. The bacterial infection more often occurred and 156 patients (74%) had experienced more than one infection while only 73 patients (35%) had been infected with fungal or viral. In order to reflect such different occurrence rates, separate cure rate models are implemented for bivariate recurrent event data. We consider each event has a separate susceptibility model for more general extension of univariate recurrent event with zero inflation to bivariate case.

In this article, we present a joint model for bivariate recurrent event data with zero inflation. The remainder of the article is organized as follows. After introducing notations and assumptions, a joint model is presented in Section 2 and an estimation procedure is described in Section 3. In Section 4, a simulation study is performed for assessing the finite sample properties of the proposed method. In Section 5, the proposed method is applied to a leukemia dataset. Finally, some concluding remarks are provided in Section 6.

## 2. Model

Suppose there are  $n$  subjects who may experience two different types of recurrent events. Denote  $T_{i1j}$  and  $T_{i2l}$  as the  $j^{\text{th}}$  and  $l^{\text{th}}$  recurrent event of two types of events of subject  $i$ , where  $i = 1, \dots, n$ ,  $j = 1, \dots, n_{i1}$ , and  $l = 1, \dots, n_{i2}$ . Let  $N_{i1}(t) = \sum_j I(T_{i1j} \leq t)$  denote the cumulative number of the first event of subject  $i$  with  $dN_{i1}(t) = 1$  indicating subject  $i$  has experienced the first event at time  $t$ . Similar definitions are applied to  $N_{i2}(t)$  and  $dN_{i2}(t)$  for the second recurrent event. A censoring time  $C_i$  is assumed to be independent of  $N_{i1}$  and  $N_{i2}$ , respectively.

Let  $W_i = (W_{i1}, W_{i2})$  be a latent random vector that indicates susceptibilities.  $W_{ik} = 1$ ,  $k = 1, 2$  means a subject has experienced the  $k$ th type of event and  $W_{ik} = 0$  indicates a risk free from the  $k$ th type of event. However, these values are unobservable at the subjects with no event who either would have experienced events at future time or would remain risk free.

Logistic regression models are used to model the susceptible probabilities,  $\pi_{ik} = \Pr(W_{ik} = 1|Z_{ik})$ ,  $k = 1, 2$ , where  $Z_{ik} = (1, X'_{ik})'$

$$\Pr(W_{i1} = 1|Z_{i1}) = \frac{\exp(\gamma'_1 Z_{i1})}{1 + \exp(\gamma'_1 Z_{i1})} = \pi_{i1},$$

$$\Pr(W_{i2} = 1|Z_{i2}) = \frac{\exp(\gamma'_2 Z_{i2})}{1 + \exp(\gamma'_2 Z_{i2})} = \pi_{i2}.$$

A bivariate frailty effect  $\nu_i = (\nu_{i1}, \nu_{i2})$  is introduced to represent the associations between recurrent events and assumed to follow a bivariate normal distribution with mean zeros and a covariance matrix

$\Sigma$  which is composed of  $\text{Var}(v_{i1}) = \sigma_1^2$ ,  $\text{Var}(v_{i2}) = \sigma_2^2$ , and  $\text{Cov}(v_{i1}, v_{i2}) = \sigma_{12}$ . Given frailty effects  $v_i = (v_{i1}, v_{i2})$  and covariate vectors  $X_{i1}$  and  $X_{i2}$ , the proportional intensity models of recurrent events are assumed

$$\begin{aligned}\lambda_{i1}(t|v_{i1}, W_{i1} = 1) &= \lambda_{01}(t)\exp(\beta_1'X_{i1} + v_{i1}), \\ \lambda_{i2}(t|v_{i2}, W_{i2} = 1) &= \lambda_{02}(t)\exp(\beta_2'X_{i2} + v_{i2}),\end{aligned}$$

where  $\lambda_{01}(t)$  and  $\lambda_{02}(t)$  are arbitrary baseline intensity functions with cumulative functions,  $\Lambda_{01}(t)$  and  $\Lambda_{02}(t)$  and  $\beta_1$  and  $\beta_2$  are regression coefficient vectors, respectively. Then the corresponding survival functions are

$$\begin{aligned}S_{i1}(t|v_{i1}, W_{i1} = 1) &= \exp\left(-\Lambda_{01}(t)e^{\beta_1'X_{i1} + v_{i1}}\right), \\ S_{i2}(t|v_{i2}, W_{i2} = 1) &= \exp\left(-\Lambda_{02}(t)e^{\beta_2'X_{i2} + v_{i2}}\right).\end{aligned}$$

Denote  $\theta_0$  as the vector of parameters of interest. The conditional likelihood is then

$$L(\theta_0|v) = \prod_{i=1}^n \prod_{k=1}^2 L_{ik,a}^{I(n_{ik}>0)} L_{ik,b}^{I(n_{ik}=0)}, \quad (2.1)$$

where

$$\begin{aligned}L_{ik,a} &= \pi_{ik} S_{ik}(c_i|v_{ik}, W_{ik} = 1) \prod_{j=1}^{n_{ik}} \lambda_{ik}(t_{ij}|v_{ik}, W_{ik} = 1), \\ L_{ik,b} &= 1 - \pi_{ik} + \pi_{ik} S_{ik}(c_i|v_{ik}, W_{ik} = 1).\end{aligned}$$

### 3. Estimation

For easy implementation of computation, a piecewise constant rate is adopted as a baseline intensity function of recurrent event. Lawless and Zhan (1998), Liu *et al.* (2016) showed that the pieces defined with appropriately selected cutpoints perform well at several situations. Let  $0 = a_{k0} < a_{k1} < a_{k2} < \dots < a_{kq_k} = \tau_k$  be the cutpoints which were determined from the equally spaced points of the ordered times of the  $k$ th type of recurrent event. Here,  $\tau_k$  denotes the largest event time of the  $k$ th recurrent event. Then the baseline intensity functions for bivariate recurrent events are redefined as

$$\rho_{10}(t_1; \lambda_{01}) = \sum_{l_1=1}^{q_1} \lambda_{01l_1} I(a_{1l_1-1} < t_1 \leq a_{1l_1}), \quad (3.1)$$

$$\rho_{20}(t_2; \lambda_{02}) = \sum_{l_2=1}^{q_2} \lambda_{02l_2} I(a_{2l_2-1} < t_2 \leq a_{2l_2}), \quad (3.2)$$

where  $\lambda_{01} = (\lambda_{011}, \dots, \lambda_{01q_1})$  and  $\lambda_{02} = (\lambda_{021}, \dots, \lambda_{02q_2})$  denote the parameters of the baseline rates and the cumulative baseline intensities are defined as

$$\begin{aligned}\tilde{\Lambda}_{10}(t_1; \lambda_{01}) &= \sum_{l_1=1}^{q_1} \lambda_{01l_1} \max(0, \min(a_{1l_1} - a_{1l_1-1}, t_1 - a_{1l_1-1})), \\ \tilde{\Lambda}_{20}(t_2; \lambda_{02}) &= \sum_{l_2=1}^{q_2} \lambda_{02l_2} \max(0, \min(a_{2l_2} - a_{2l_2-1}, t_2 - a_{2l_2-1})).\end{aligned}$$

Let  $\theta = (\lambda_{01}, \lambda_{02}, \beta_1, \beta_2, \gamma_1, \gamma_2, \Sigma)$  be the parameters of interest. Denote  $e_{1il}(t_{1ij}) = I(a_{1l-1} < t_{1ij} \leq a_{1l})$ ,  $l = 1, \dots, q_1$  as the  $j^{\text{th}}$  occurrence of the first-type of event is observed at the  $l^{\text{th}}$  interval and  $m_{1l} = \sum_{ij} e_{1il}(t_{1ij})$  a total number of the first-type of event occurring at the  $l^{\text{th}}$  interval. For the second-type of event,  $e_{2il}$  and  $m_{2l}$  are similarly defined. Then given a bivariate frailty effect  $v_i = (v_{i1}, v_{i2})$ , a conditional likelihood (2.1) is rewritten as

$$L(\theta|v) = \prod_{i=1}^n \prod_{k=1}^2 L_{ik}(\theta|v_{ik}),$$

where

$$L_{ik}(\theta|v) = \left\{ \prod_{j=1}^{n_{ik}} \prod_{l=1}^{q_k} \left[ \lambda_{0kl} e^{\beta'_k X_{ik} + v_{ik}} \right]^{e_{kil}(t_{1ij})} \times \exp \left[ -\tilde{\Lambda}_{k0}(t_{in_{ik}}) e^{\beta'_k X_{ik} + v_{ik}} \right] \right\}^{I(n_{ik} > 0)} \\ \times \left\{ \exp \left[ -\tilde{\Lambda}_{k0}(c_i) e^{\beta'_k X_{ik} + v_{ik}} \right] \pi_{ik} + (1 - \pi_{ik}) \right\}^{I(n_{ik} = 0)}. \quad (3.3)$$

The marginal log likelihood is obtained by integrating (3.3) with respect to a bivariate frailty density  $f(v_i|\Sigma) = f(v_{i1}, v_{i2}|\Sigma)$ ,

$$L_c(\theta) = \prod_{i=1}^n \iint \prod_{k=1}^2 L_{ik}(\theta|v_{ik}) f(v_i|\Sigma) dv_{ik}.$$

An EM algorithm is applied to recover unknown quantities  $v_i = (v_{i1}, v_{i2})$ . In the E-step, since the complete log likelihood has no closed form, a Gauss-Hermite integration or Markov chain Monte Carlo (MCMC) algorithm can be adopted. Once completing E-step, the vector of parameters is updated by applying a Newton-Raphson algorithm at the M-step using the score function and the hessian matrix derived from likelihood. The standard error is estimated by the observed Fisher information updated with final estimates of  $\theta$ . In our study, the estimation procedure in PROC NLMIXED of SAS is used.

#### 4. Simulation

We present the results from a simulation study to investigate the performance of the proposed estimation procedures. Five hundred replicates are generated and each one has a sample size of 100. A single binary covariate  $Z_i$  taking value 0 or 1 with same probability is considered. In this simulation, four settings are provided according to the sign of correlation between bivariate frailties and the proportions of zero inflation for each event. The following logistic regression model is applied for generating  $W_{ik}$ ,

$$\text{logit}P(W_{ik} = 1|Z_i) = \log \frac{\pi_{ik}}{1 - \pi_{ik}} = \gamma_{k0} + \gamma_{k1}Z_i, \quad k = 1, 2,$$

For the values of  $\gamma$ 's controlling the susceptibilities, (i)  $(\gamma_{10}, \gamma_{11}) = (\gamma_{20}, \gamma_{21}) = (-1.0, 0.1)$  at the same proportions of zero inflation and (ii)  $(\gamma_{10}, \gamma_{11}) = (0.1, 0.1)$ ,  $(\gamma_{20}, \gamma_{21}) = (-1.0, 0.1)$  at different proportions of zero inflation, respectively. A bivariate frailty effect is assumed to follow a bivariate normal distribution  $v_i = (v_{i1}, v_{i2}) \sim \text{BVN}(0, \Sigma)$ , where  $\sigma_1^2 = 0.5$ ,  $\sigma_2^2 = 0.5$  and two different covariance

Table 1: Simulation 1: Same zero rates at two recurrent events with positive correlation  $\sigma_{12} = 0.3$ 

	Parameter	Bivariate frailty				No zero inflation			
		Est	SSE	SEM	CP	Est	SSE	SEM	CP
1st event	$\beta_1(0.5)$	0.496	0.196	0.184	0.940	0.375	0.345	0.336	0.926
	$\gamma_{01}(-1.0)$	-1.031	0.350	0.342	0.954				
	$\gamma_{11}(0.1)$	0.124	0.462	0.489	0.964				
2nd event	$\beta_2(0.5)$	0.497	0.185	0.185	0.954	0.380	0.349	0.339	0.926
	$\gamma_{02}(-1.0)$	-1.021	0.369	0.339	0.950				
	$\gamma_{12}(0.1)$	0.120	0.490	0.467	0.956				
Covariance	$\sigma_1^2(0.5)$	0.486	0.115	0.115	0.930	2.372	0.600	0.520	0.000
	$\sigma_2^2(0.5)$	0.493	0.113	0.112	0.924	2.404	0.564	0.520	0.000
	$\sigma_{12}(0.3)$	0.397	0.098	0.087	0.942	0.220	0.321	0.286	0.922

Est = estimate; SEE = standard error; SEM = empirical values; CP = coverage probabilities.

values  $\sigma_{12} = -0.3$  and  $\sigma_{12} = 0.3$ , respectively. Now, given  $W_{ik} = 1$ , recurrent gap times  $s_{ikl}$  are generated from the following intensity,

$$\tilde{\lambda}_{0k}(t)\exp(\beta_k Z_i + \nu_{ik}), \quad k = 1, 2,$$

where  $\tilde{\lambda}_{0k}$  is assumed to follow a Weibull distribution. The parameters are set as  $\beta_1 = \beta_2 = 0.5$ . Then set  $t_{ikl} = s_{ikl} + t_{ikl-1}$  with  $t_{ik0} = 0$  and continue until  $t_{ikr} > C_i$  where two recurrent events are subject to the same censoring time  $C_i = 1 + U(0, 1)$ . Then the number of the  $k^{th}$  recurrent event is defined as  $n_{ik} = r - 1$  where  $t_{ikr-1} < C_i < t_{ikr}$ . Therefore, if  $t_{ik1} > C_i$ , subject  $i$  experiences no events and then  $n_{ik} = 0$ . For subject  $i$  with  $W_{ik} = 0$ , generate  $C_i$  and set  $n_{ik} = 0$  which means a cure group.

At the first and second settings, both events are assumed to have same susceptible rates. About 30% of subjects are unsusceptible(cure) and another 15% of susceptible subjects had no recurrent events. All estimation procedures are performed using Proc NLMIXED in SAS under a piecewise baseline intensity assumption. The results of Tables 1 and 2 with the same proportions of two recurrent events indicate that all parameter estimates have small biases and standard error (SEE) and empirical values (SEM) are almost same. Coverage probabilities (CP) are also close to 95%. To evaluate the effect of zero inflation model, the simulation results ignoring a cure rate model are presented and show  $\beta$  and correlation coefficients  $\sigma_{12}$  are underestimated but two variances  $\sigma_1^2$  and  $\sigma_2^2$  are overestimated. The third and fourth setting is configured to mimic the leukemia recurrent infection data with different susceptible rates. The first recurrent event has a 60% susceptible rate and the second recurrent event has 40% susceptibility rate, respectively. Tables 3 and 4 show the simulation results under two different correlations where the biases become more severe than those at the same susceptible rates. In particular, the biases of the variance deteriorate at the first type of event which has a higher susceptible rate.

## 5. Data analysis

In this section, the proposed method was applied to a leukemia clinical trial data in which the chemotherapy was administered to 210 patients with AML from October 2002 to October 2008. During this period, patients experienced repeated bacteria, viral or fungal related infections. In this study, we investigated the effects of some covariates on two kinds of infections. The first infection is caused by bacteria and the second one is related to viral or fungal ones. The bacterial infection occurred more often and 156 patients had more than one infections while only 73 patients had been infected with fungal and viral. The range of each infection is [1, 6] and [1, 4] and the average recurrent numbers were 2.08 and 1.21, respectively.

Table 2: Simulation 2: Same zero rates at two recurrent events with negative correlation  $\sigma_{12} = -0.3$ 

	Parameter	Bivariate frailty				No zero inflation			
		Est	SSE	SEE	CP	Est	SSE	SEE	CP
1st event	$\beta_1(0.5)$	0.505	0.191	0.187	0.940	0.394	0.324	0.340	0.916
	$\gamma_{01}(-1.0)$	-0.993	0.336	0.337	0.954				
	$\gamma_{11}(0.1)$	0.102	0.459	0.415	0.932				
2nd event	$\beta_2(0.5)$	0.503	0.190	0.184	0.940	0.381	0.357	0.334	0.916
	$\gamma_{02}(-1.0)$	-1.044	0.363	0.343	0.932				
	$\gamma_{12}(0.1)$	0.129	0.515	0.491	0.940				
Covariance	$\sigma_1^2(0.5)$	0.479	0.104	0.108	0.920	2.420	0.584	0.531	0.000
	$\sigma_2^2(0.5)$	0.484	0.119	0.109	0.896	2.336	0.542	0.510	0.000
	$\sigma_{12}(-0.3)$	-0.295	0.090	0.087	0.942	-0.236	0.277	0.283	0.954

Est = estimate; SEE = standard error; SEM = empirical values; CP = coverage probabilities.

Table 3: Simulation 3: Different zero rates at two recurrent events with positive correlation  $\sigma_{12} = 0.3$ 

	Parameter	Bivariate frailty				No zero inflation			
		Est	SSE	SEE	CP	Est	SSE	SEE	CP
1st event	$\beta_1(0.5)$	0.511	0.241	0.223	0.920	0.258	0.635	0.594	0.910
	$\gamma_{01}(0.1)$	0.103	0.297	0.294	0.960				
	$\gamma_{11}(0.1)$	0.088	0.422	0.411	0.944				
2nd event	$\beta_2(0.5)$	0.495	0.194	0.189	0.938	0.401	0.351	0.336	0.938
	$\gamma_{02}(-1.0)$	-1.012	0.346	0.341	0.960				
	$\gamma_{12}(0.1)$	0.098	0.490	0.471	0.962				
Covariance	$\sigma_1^2(0.5)$	0.481	0.137	0.136	0.924	6.441	1.133	1.608	0.000
	$\sigma_2^2(0.5)$	0.488	0.106	0.110	0.930	2.372	0.587	0.519	0.000
	$\sigma_{12}(0.3)$	0.295	0.104	0.104	0.946	0.254	0.525	0.504	0.948

Est = estimate; SEE = standard error; SEM = empirical values; CP = coverage probabilities.

Table 4: Simulation 4: Different zero rates at two recurrent events with negative correlation  $\sigma_{12} = -0.3$ 

	Parameter	Bivariate frailty				No zero inflation			
		Est	SSE	SEE	CP	Est	SSE	SEE	CP
1st event	$\beta_1(0.5)$	0.487	0.221	0.222	0.948	0.248	0.639	0.593	0.912
	$\gamma_{01}(0.1)$	0.105	0.315	0.294	0.942				
	$\gamma_{11}(0.1)$	0.082	0.432	0.414	0.942				
2nd event	$\beta_2(0.5)$	0.489	0.192	0.185	0.930	0.380	0.315	0.336	0.944
	$\gamma_{02}(-1.0)$	-1.022	0.345	0.341	0.970				
	$\gamma_{12}(0.1)$	0.095	0.464	0.470	0.970				
Covariance	$\sigma_1^2(0.5)$	0.468	0.132	0.133	0.902	6.383	1.168	1.593	0.000
	$\sigma_2^2(0.5)$	0.484	0.114	0.109	0.902	2.371	0.588	0.519	0.000
	$\sigma_{12}(-0.3)$	-0.289	0.103	0.103	0.940	-0.168	0.513	0.497	0.954

Est = estimate; SEE = standard error; SEM = empirical values; CP = coverage probabilities.

To reflect this discrepancy, a joint model with a separate zero inflation rate model was implemented. Three covariates (age at the diagnosis of AML, the leukemia risk level and the dose of cytarabine given for the first course of chemotherapy) are included in the logistic models and hazard models. The average of age was 8.65 (year). The risk has three classes (low = 0 ( $n = 68$ ); standard = 1 ( $n = 89$ ) and; high = 2 ( $n = 62$ )) and there are two classes for dose variable (standard = 0 ( $n = 109$ ) and high = 1 ( $n = 105$ )). Here, two dummy variables for risk variables were used (Risk 1 = 1 if standard; Risk 2 = 1 if high).

Using piecewise constants baseline hazards for  $\lambda_{01}$  and  $\lambda_{02}$  with five intervals, two models(a bivariate frailty effect with cure rates and a reduced model without cure rates) were compared. Table

Table 5: Leukemia patients' infection data

	Parameter	Bivariate frailty			No zero inflation			Independence			
		Est	SE	<i>p</i> -value	Est	SE	<i>p</i> -value	Est	SE	<i>p</i> -value	
Bacterial infection	logistic model	intercept	-2.562	7.384	0.729				-2.010	1.751	0.242
		age	-5.630	9.346	0.547				-1.240	4.216	0.764
		Risk 1	-0.115	6.418	0.987				-1.333	2.390	0.564
		Risk 2	0.941	4.527	0.835				0.897	1.897	0.636
	recurrent event	dose	0.230	3.844	0.952				0.896	1.611	0.579
		age	-0.025	0.009	0.012	-0.024	0.009	0.015	-0.025	0.009	0.006
		Risk 1	-0.244	0.140	0.084	-0.041	0.144	0.772	-0.244	0.138	0.078
		Risk 2	-0.435	0.161	0.008	-0.229	0.162	0.157	-0.413	0.160	0.010
		dose	0.075	0.121	0.534	0.077	0.120	0.520	0.083	0.121	0.488
Fungal or viral infection	logistic model	intercept	-1.589	4.717	0.736				-0.283	3.286	0.931
		age	-4.625	6.426	0.473				-1.872	2.668	0.484
		Risk 1	0.278	5.045	0.956				-1.362	2.941	0.644
		Risk 2	1.086	1.540	0.481				-1.089	1.358	0.423
	recurrent event	dose	0.412	2.384	0.862				0.374	0.574	0.516
		age	-0.010	0.019	0.609	-0.004	0.023	0.836	-0.015	0.019	0.430
		Risk 1	-0.183	0.290	0.530	0.338	0.412	0.413	-0.140	0.269	0.602
		Risk 2	-0.170	0.324	0.606	0.406	0.491	0.408	-0.213	0.297	0.474
		dose	0.311	0.252	0.217	0.214	0.301	0.475	0.231	0.226	0.307
covariance	$\sigma_1^2$	0.111	0.081	0.174	0.096	0.066	0.152	0.080	0.071	0.260	
	$\sigma_2^2$	0.599	0.339	0.078	1.186	1.009	0.239	0.257	0.229	0.262	
	$\sigma_{12}$	0.239	0.113	0.035	0.338	0.147	0.022				

Est = cure rates; SE = standard errors.

5 shows the covariate effects on both event intensities and cure rates (Est), the standard errors (SE) and *p*-values three models. Using the bivariate frailty effect model, bacterial infection showed no significant covariates in the cure logistic model, however, age and risk level were significant covariates for intensity of infection such as younger children and those with lower level risk experienced more bacterial infections. However, there were no significant effect on the logistic model and infection intensity for fungal and viral infection. For a bivariate frailty effect, the variance of fungal and viral infection, the variance estimate of bacterial infection is smaller than fungal and viral infection. The estimated covariance of bivariate frailty effect is a positive value ( $\hat{\sigma}_{12} = 0.239$ , *p*-value = 0.035) indicating that these two infection types were significantly correlated. A bivariate recurrent event model without a cure model was also applied to the same dataset to compare the suggested methods. Some covariates (age and dose) on intensity function of bacteria infection demonstrated similar results in both models which can be explained by higher incidence rates so as not affected by the ignorance of cure rate model. However, at viral and fungal infection, even though covariates have insignificant effects, the estimates seem to have larger values even at the variance estimates. Under an independence frailty effect model, the significance of covariates does not change.

## 6. Concluding remarks

In the preceding sections, we presented a joint model for regression analysis of bivariate recurrent events in the presence of zero inflation. Bivariate frailty effect is also applied to model the correlation between two recurrent events. In particular, bivariate recurrent events are assumed to have different susceptibilities for each event. The simulation results indicated that the model ignoring zero inflation brings biased estimates of covariate effects and variance. Such modelling can be extended to connect the recurrent event and susceptibility. For example, as a referee pointed out, one type of infection can be applied with an ordinary intensity model without a cure model because of a more frequent

occurrence.

Another extension is to include a terminal event process. A similar model was studied in the context of univariate recurrent event data with zero inflation as well as a terminal event by Liu *et al.* (2016). Under the proportional hazard model, a terminal event has the following hazard model,

$$\lambda^D(t|Z, v_1, v_2) = \lambda_0^D(t)\exp(\eta_1 Z + \tau_1 v_1 + \tau_2 v_2),$$

where  $\tau_1$  and  $\tau_2$  measure the associations between the first event and death as well as the second event and death, respectively. That is,  $\tau_1 > 0$  indicates that a subject with more frequent events tends to die more rapidly than a subject with fewer events. The relation between recurrent event and a terminal event becomes more clear by adopting this information simultaneously. The proposed estimation procedure is based on the piecewise baseline intensity function with pre specified time points and numbers. Simulation shows that this method still works well; however, more general rate functions based on marginal model and conditional model should be discussed as alternatives.

### Acknowledgements

Authors thank Drs. Rubnitz, Stanley Pounds and Liang Zhu for providing the dataset. This work was supported by a Korea Research Grant (NRF-2017R1D1A1B03030578).

### Appendix A: SAS code

```
proc nlmixed data=spptotal
parms   h1=5 h2=5 h3=5 h4=5 h5=5  beta11=0.5 gam01=-1.0 gam11=0.1
        g1=5 g2=5 g3=5 g4=5 g5=5  beta21=0.5 gam02=-1.0 gam12=0.1
        theta1=0.5 theta=0.5;
bounds  theta theta1>0;

if $event_type$=1 then do; /** for the first event **/
eta1=gam01+gam11*x1; /*** cure fraction **/
prob1=1/(1+exp(-eta1));

basehaz1=h1*e1+h2*e2+h3*e3+h4*e4+h5*e5;
cumhaz1=h1*d1+h2*d2+h3*d3+h4*d4+h5*d5;
mu1=x1*beta11+nu1;
loglik01=-exp(mu1)*cumhaz1;

if delta=0 & rep=0 then loglik1=log(prob1+(1-prob1)*exp(loglik01));
/** log likelihood for subjects without the 1st recurrent event **/
if delta=1 then loglik1=log(basehaz1)+mu1;
/** log likelihood for the 1st recurrent event **/
if delta=0 & rep>0 then loglik1=log(1-prob1)+loglik01;
/*** log likelihood for subjects with 1st recurrent event and last
observation **/
end;

if $event_type$=2 then do; /** for the second event **/
```



```

eta2=gam02+gam12*x1; /*** cure fraction**/
prob2=1/(1+exp(-eta2));

basehaz2=g1*e1+g2*e2+g3*e3+g4*e4+g5*e5;
cumhaz2=g1*d1+g2*d2+g3*d3+g4*d4+g5*d5;
mu2=x1*beta21+nu2;

loglik02=-exp(mu2)*cumhaz2;

if delta=0 & rep=0 then loglik1=log(prob2+(1-prob2)*exp(loglik02));
  /** log likelihood for subjects without the 2nd recurrent event **/
if delta=1 then loglik1=log(basehaz2)+mu2;
  /**log likelihood for the 2nd recurrent event ***/
if delta=0 & rep>0 then loglik1=log(1-prob2)+loglik02;
  /***log likelihood for subjects with 2nd recurrent event and last
  observation **/
end;

loglik=loglik1;
model time~general(loglik);
random nu1 nu2 ~ normal([0,0],[ theta, theta12, theta1]) subject=id;
run;

```

## References

- Andersen PK and Gill RD (1982). Cox's regression model for counting processes: a large sample study, *Annals of Statistics*, **10**, 1100–1120.
- Cai J and Schaubel DE (2004). Marginal means/rates models for multiple type recurrent event types, *Lifetime Data Analysis*, **10**, 121–138.
- Cook RJ and Lawless JF (2007). *The Statistical Analysis of Recurrent Events*, Springer, New York.
- Cook RJ, Lawless JF, and Lee KA (2010). A copula-based mixed Poisson model for bivariate recurrent events under event-dependent censoring, *Statistics in Medicine*, **29**, 694–707.
- Cook RJ, Zeng L, and Lee KA (2008). A multistate model for bivariate interval-censored failure time data, *Biometrics*, **64**, 1100–1109.
- Kim YJ and Jhun MS (2008). Cure rate model with interval censored data, *Statistics in Medicine*, **27**, 3–14.
- Kim YJ (2017). Cure rate model with bivariate interval censored data, *Communications in Statistics, Computation and Simulation*, **46**, 7114–7124.
- Lawless JF and Zhan M (1998). Analysis of interval-grouped recurrent event data using piecewise constant rate functions, *Canadian Journal Statistics*, **26**, 549–565.
- Liu L, Huang X, Yaroshinsky A, and Cormier JN (2016). Joint frailty models for zero-inflated recurrent events in the presence of a terminal event, *Biometrics*, **72**, 204–214.
- Liu L and Huang X (2008). The use of Gaussian quadrature for estimation in frailty proportional hazards models, *Statistics in Medicine*, **27**, 2665–2683.
- Peng Y and Dear KBG (2000). A nonparametric mixture model for cure rate estimation, *Biometrics*, **56**, 237–243.

- Rondeau V, Schaffner E, Corbiere F, Gonzalez GR, and Mathoulin-Plissier S (2013). Cure frailty models for survival data: Application to recurrences for breast cancer and to hospital readmission for colorectal cancer, *Statistical Methods in Medical Research*, **22**, 243–260.
- Sy JP and Taylor JMG (2000). Estimation in a Cox proportional hazards cure model, *Biometrics*, **56**, 227–236.
- Wei LJ, Lin DY, and Weissfeld L (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, *Journal of the American Statistical Association*, **84**, 1065–1073.
- Zhao XB, Zhou X, and Wang JL (2012). Semiparametric model for recurrent event data with excess zeros and informative censoring, *Journal of Statistical Planning and Inference*, **142**, 289–300.
- Zhu H, DeSantis SM, and Luo S (2018). Joint modeling of longitudinal zero-inflated count and time-to-event data: a Bayesian perspective, *Statistical methods in Medical Research*, **27**, 1258–1270.
- Zhu L, Sun J, Tong X, and Srivastava DK (2010). Regression analysis of multivariate recurrent event data with a dependent terminal event, *Lifetime Data Analysis*, **16**, 478–490.

*Received June 15, 2019; Revised August 16, 2019; Accepted December 11, 2019*