BIOMETRIC METHODOLOGY

Biometrics WILEY
A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY

Penalized survival models for the analysis of alternating recurrent event data

Lili Wang[1] | **Kevin He**[1] | **Douglas E. Schaubel**[2]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan

[2]Department of Biostatistics, Epidemiology, and Informatics, University of Pennslyvania, Philadelphia, Pennslyvania

Correspondence
Douglas E. Schaubel, Department of Biostatistics, Epidemiology, and Informatics, University of Pennslyvania, Philadelphia 19104, PA.
Email: douglas.schaubel@pennmedicine.upenn.edu

Abstract

Recurrent event data are widely encountered in clinical and observational studies. Most methods for recurrent events treat the outcome as a point process and, as such, neglect any associated event duration. This generally leads to a less informative and potentially biased analysis. We propose a joint model for the recurrent event rate (of incidence) and duration. The two processes are linked through a bivariate normal frailty. For example, when the event is hospitalization, we can treat the time to admission and length-of-stay as two alternating recurrent events. In our method, the regression parameters are estimated through a penalized partial likelihood, and the variance-covariance matrix of the frailty is estimated through a recursive estimating formula. Moreover, we develop a likelihood ratio test to assess the dependence between the incidence and duration processes. Simulation results demonstrate that our method provides accurate parameter estimation, with a relatively fast computation time. We illustrate the methods through an analysis of hospitalizations among end-stage renal disease patients.

KEYWORDS

alternating recurrent events, correlated frailty model, end-stage renal disease, penalized partial likelihood

1 | INTRODUCTION

Recurrent event data are commonly encountered in clinical experiments and observational studies. Examples include repeated hospitalizations, recurrent opportunistic infections for HIV-infected patients, and recurrent tumors for cancer patients. Various methods have been developed, most based on either intensity functions (Andersen and Gill, 1982) or rate functions (Lin *et al.*, 2000). Most such methods treat the recurrent event sequence as a point process and hence assume (at least implicitly) that event durations are negligible. In cases where event duration is variable and not negligible, information and accuracy are sacrificed by not incorporating event duration. For example, hospitalizations are frequently analyzed as recurrent events and each admission has a corresponding length-of-stay. An analysis which ignores length-of-stay results in biased regression coefficients, in accordance with the degree of association between the length-of-stay and admission rate.

A natural way to accommodate recurrent events with nonnegligible duration is to cast the data structure as an alternating gap time sequence. Very few methods have been proposed along these lines. A nonparametric approach was developed by Huang and Wang (2005). Recently, Lee *et al.* (2018) developed an estimating equation approach for accelerated failure time model for an alternating recurrent event data. In their proposed estimating procedure, the distribution of the possibly correlated random variables for two processes was left unspecified. Like Huang and Wang (2005), Lee *et al.* (2018) does not provide information on the correlations within or between the two recurrent event sequences.

Alternatively, correlated frailty models can be developed to accommodate the alternating recurrent event setting by adapting bivariate frailty models for clustered multivariate failure time data (Yashin *et al.*, 1995; Xue and Brookmeyer, 1996). The marginal likelihood (integrating out the unobserved frailties) can be obtained through the expectation-maximization (EM) algorithm, Gaussian-Hermite quadrature or a Laplace approximation (Vaida and Xu, 2000; Ripatti *et al.*, 2002; Huang and Liu, 2007; Liu and Huang, 2008). Ripatti and Palmgren (2000) developed a Laplace approximation–based approach to estimate a penalized partial-likelihood (PPL) for a multivariate frailty model, which was shown to converge much faster than EM (Therneau *et al.*, 2003).

Ripatti and Palmgren (2000) developed a frailty model applicable to alternating gap time data. The methods restrict the sign of the correlation between the gap time sequences to be positive. Dharmarajan *et al.* (2018) developed related techniques but required the correlation to be negative. Although suitable for many practical settings, the need to prespecify the correlation limits the implementation of such approaches in the alternating recurrent event setting. It could be difficult to accurately prespecify the direction of the correlation.

In this report, our objective is to develop methods for analyzing alternating recurrent event data that are flexible, informative, computationally efficient, and implementable in very large data sets. We propose a novel PPL estimating approach for alternating recurrent event data using correlated log-normal frailties. This is equivalent to modeling two recurrent event processes jointly, incorporating a bivariate random intercept (with correlated elements) to represent between- and within-process correlations.

The key difference between our proposed methods and Ripatti and Palmgren (2000) lies in how the variance/covariance parameters are estimated. Our estimating equation for the variance components is obtained by differentiating the approximate marginal likelihood with respect to its inverse variance matrix; in contrast to Ripatti and Palmgren (2000), for which differentiation was with respect to the scalar elements. We provide a detailed description of the technical differences between the two approaches in the Supporting Information document. In addition, we propose a likelihood ratio test (LRT) to assess whether or not the two alternating recurrent event sequences are mutually independent.

The remainder of this report is organized as follows. We introduce our model in the setting of alternating recurrent events in Section 2. In Section 3, our proposed methods are described. Simulations to evaluate the proposed methods on finite sample sizes are summarized in Section 4. We apply the proposed method to an analysis of end-stage renal disease patients from the Dialysis Outcomes and Practice Patterns Study (DOPPS) in Section 5. Some concluding remarks are provided in Section 6.

2 | MODEL SPECIFICATION

A total of n-independent subjects are followed over time and experience two alternating events indexed by k; that is, $k = 1$ and $k = 2$ represent the first and second event types, respectively. For subject i, the total event time of the jth occurrence of event type k is denoted by T^*_{ijk}. The ordered event times are then given by $0 < T^*_{i11} < T^*_{i12} < T^*_{i21} < T^*_{i22} < \cdots$. The gap times of interest for subject i are then given by $\tilde{T}^*_{ij1} = T^*_{ij1} - T^*_{i(j-1)2}$ for $k = 1$ and $\tilde{T}^*_{ij2} = T^*_{ij2} - T^*_{ij1}$ for $k = 2$. This setup is depicted in Figure 1. For example, the two alternating recurrent events could be time-to-readmission to a hospital (measured from the most recent discharge) and time-to-discharge (ie, length of stay).

We set the covariate for gap time $\tilde{T}_{ijk}$ to $\boldsymbol{Z}_{ijk}$, such that all elements are set to their values at their respective gap time origins; that is, $\boldsymbol{Z}_{ij1} = \boldsymbol{Z}_{ij}(T_{i,j-1,2})$ and $\boldsymbol{Z}_{ij2} = \boldsymbol{Z}_{ij}(T_{ij1})$. This is for notational convenience, as the proposed methods can accommodate continuously changing covariates. The covariates comprising the two recurrent event types, $\boldsymbol{Z}_{ij1}$ and $\boldsymbol{Z}_{ij2}$, can be different. Note that any time-dependent elements are restricted to be external (Kalbfleisch and Prentice, 2002). The alternating recurrent event process is right-censored by C_i. Let m_i be the number of observed complete event pairs from subject i. Correspondingly, we denote the observed covariate history by $\boldsymbol{Z}_i = \{\boldsymbol{Z}_{ijk}, j = 1, \ldots, m_i + 1, k = 1, 2\}$, and assume that C_i is independent of the event times given $\boldsymbol{Z}_i$. The event indicators and total observation times are defined as $\delta_{ijk} = I(T^*_{ijk} < C_i)$ and $T_{ijk} = T^*_{ijk} \wedge C_i$, respectively, where $I(\cdot)$ is a 0/1 indicator function and $a \wedge b = min\{a, b\}$. Note that it is always true that for $j = 1, \ldots, m_i$, we have $\delta_{ijk} = 1$ and $T_{ijk} = T^*_{ijk}$, while $\delta_{i(m_i+1)2} = 0$ and $T_{i(m_i+1)2} = C_i$. Event gap times $\tilde{T}^*_{ijk}$ are subject to censoring $\tilde{C}_{ijk}$, where $\tilde{C}_{ij1} = C_i - T_{i(j-1)2}$ and
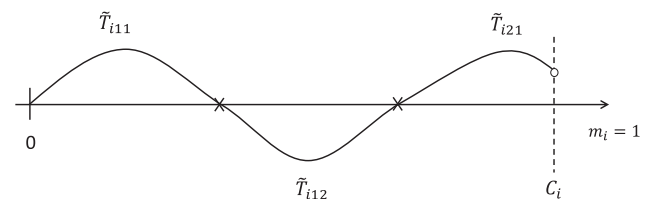

FIGURE 1 Alternating recurrent events under right censoring. $\tilde{T}_{i11}$ is the observed awaiting time to the first hospitalization of subject i. $\tilde{T}_{i12}$ is the length of stay before discharge. $\tilde{T}_{i21}$, the observed time to the second hospitalization, is censored. Therefore, we only observe one complete event pair ($m_i = 1$)

$\tilde{C}_{ij2} = C_i - T_{ij1}$. Note that $\tilde{T}_{i0k}^* = 0$ and $\tilde{T}_{i0k} = 0$ for $k = 1, 2$. Consequently, the observed gap times are $\tilde{T}_{ijk} = \tilde{T}_{ijk}^* \wedge \tilde{C}_{ijk}$. For each event gap, we introduce an at-risk indicator $Y_{ijk}(t) = I(t \leqslant \tilde{T}_{ijk})$.

The assumed hazard model for $\tilde{T}_{ijk}$ is given by

$$\lambda_{ijk}(t \mid \mathbf{Z_i}, \boldsymbol{\gamma}_i) = \lambda_{0k}(t)\exp(\boldsymbol{\beta}_k' \mathbf{Z}_{ijk} + \gamma_{ik}), \quad (1)$$

where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2})'$ are independent draws from a mean-zero bivariate normal distribution, BV $N(\mathbf{0}_2, \boldsymbol{D}_{2 \times 2})$. We assume that the $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots$ are mutually independent, and that each $\boldsymbol{\gamma}_i$ is independent of $\mathbf{Z}_i$. Moreover, for $(j, k) \neq (p, q)$, $\tilde{T}_{ijk}^*$ and $\tilde{T}_{ipq}^*$ are independent given $\{\boldsymbol{\gamma}_i, \mathbf{Z}_{ijk}, \mathbf{Z}_{ipq}\}$. Consequently, the $\tilde{T}_{ijk}^*$ are subject to censoring via $\tilde{C}_{ijk}$, which is conditionally independent given $\boldsymbol{\gamma}_i$ and $\mathbf{Z}_i$.

We stack $\boldsymbol{\gamma}_i$ from the n subjects into a vector $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \ldots, \boldsymbol{\gamma}_n')'$, which then follows a mean-zero multivariate normal distribution $MVN(\mathbf{0}_{2n}, \boldsymbol{\Sigma}_{2n \times 2n})$. Note that $\boldsymbol{\Sigma} = \boldsymbol{D} \otimes \boldsymbol{I}_{n \times n}$ is a block-diagonal matrix, where $\otimes$ is a Kronecker product and $\boldsymbol{I}_{n \times n}$ is an n by n identity matrix. We also include frailty design vectors $\boldsymbol{R}_{ik} = (0_{(1)}, \ldots, 1_{(2i-2+k)}, \ldots, 0_{(2n)})$ to indicate γ_{ik} in the $(2i - 2 + k)$th entry of $\boldsymbol{\gamma}$ is present. We only account for random intercepts here, though it is possible for the models to be extended such that various covariates have random effects. The proposed event-specific intensity in (1) will become

$$\lambda_{ijk}(t \mid \mathbf{Z}_i, \boldsymbol{R}_{ik}, \boldsymbol{\gamma}) = \lambda_{0k}(t)\exp(\boldsymbol{\beta}_k' \mathbf{Z}_{ijk} + \boldsymbol{\gamma}' \boldsymbol{R}_{ik}). \quad (2)$$

The likelihood for subject i conditional on $\boldsymbol{\gamma}$ and $\mathbf{Z}_i$ is given by

$$L_i(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{01}(\cdot), \lambda_{02}(\cdot) \mid \boldsymbol{\gamma}, \mathbf{Z}_i)$$
$$= \prod_{j=1}^{m_i+1} \prod_{k=1}^{2} \{\lambda_{0k}(\tilde{T}_{ijk})\exp(\eta_{ijk})\}^{\delta_{ijk}}$$
$$\times \exp\{-\Lambda_{0k}(\tilde{T}_{ijk})\exp(\eta_{ijk})\}, \quad (3)$$

where the cumulative intensity of event process k is $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(s)ds$, and $\eta_{ijk} = \boldsymbol{\beta}_k' \mathbf{Z}_{ijk} + \boldsymbol{\gamma}' \boldsymbol{R}_{ik}$. It follows from (3) that the marginal likelihood is

$$L_m = L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{01}(\cdot), \lambda_{02}(\cdot))$$
$$= \int_{\boldsymbol{\gamma}} \prod_{i=1}^{n} L_i(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{01}(\cdot), \lambda_{02}(\cdot) \mid \boldsymbol{\gamma}, \mathbf{Z}_i)$$
$$\times \exp\left(-\frac{1}{2}\boldsymbol{\gamma}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}\right) |\boldsymbol{\Sigma}|^{-1/2} d\boldsymbol{\gamma}. \quad (4)$$

Note that the marginal likelihood in (4) is not in closed form. A PPL-based estimation procedure is developed for the proposed model.

3 | PARAMETER ESTIMATION

3.1 | Approximate likelihood

We derive an approximate marginal likelihood for the log-likelihood from (3). The joint likelihood function from the n subjects and their frailties can be represented as

$$L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{01}(\cdot), \lambda_{02}(\cdot), \boldsymbol{\gamma}) = |\boldsymbol{\Sigma}|^{-1/2}\exp(-\boldsymbol{K}(\boldsymbol{\gamma})), \quad (5)$$

where we define

$$K(\boldsymbol{\gamma}) = \frac{1}{2}\boldsymbol{\gamma}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma} + \sum_{i=1}^{n} \sum_{j=1}^{m_i+1} \sum_{k=1}^{2} \Lambda_{0k}(\tilde{T}_{ijk})\exp(\eta_{ijk})$$
$$- \delta_{ijk}\{\eta_{ijk} + \log(\lambda_{0k}(\tilde{T}_{ijk}))\}. \quad (6)$$

Through a Taylor expansion, $\boldsymbol{K}(\boldsymbol{\gamma})$ is approximated by

$$\boldsymbol{K}(\boldsymbol{\gamma}) \approx \hat{\boldsymbol{K}}(\boldsymbol{\gamma}) = \boldsymbol{K}(\tilde{\boldsymbol{\gamma}}) + \frac{1}{2}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}})' \boldsymbol{K}_2(\tilde{\boldsymbol{\gamma}})(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}), \quad (7)$$

where $\tilde{\boldsymbol{\gamma}}$ is the solution of $\boldsymbol{K}_1(\boldsymbol{\gamma}) = 0$, with

$$\boldsymbol{K}_1(\boldsymbol{\gamma}) = \frac{\partial \boldsymbol{K}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}$$
$$+ \sum_{i=1}^{n} \sum_{j=1}^{m_i+1} \sum_{k=1}^{2} \{\Lambda_{0k}(t_{ijk})\exp(\eta_{ijk}) - \delta_{ijk}\}\boldsymbol{R}_{ik}. \quad (8)$$

The corresponding second derivative is given by

$$\boldsymbol{K}_2(\boldsymbol{\gamma}) = \frac{\partial^2 \boldsymbol{K}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}$$
$$= \boldsymbol{\Sigma}^{-1} + \sum_{i=1}^{n} \sum_{j=1}^{m_i+1} \sum_{k=1}^{2} \Lambda_{0k}(t_{ijk})\exp(\eta_{ijk})\boldsymbol{R}_{ik}^{\otimes 2}, \quad (9)$$

where we define $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}'$. Note that $\boldsymbol{K}_2(\boldsymbol{\gamma})$ in (9) is block-diagonal.

Through a Laplace approximation, we plug $\hat{\boldsymbol{K}}(\boldsymbol{\gamma})$ into (5) and integrate it to obtain an approximate marginal log-likelihood,

$$l_m = \log(L_m) \approx -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \boldsymbol{K}(\tilde{\boldsymbol{\gamma}}) - \frac{1}{2}\log|\boldsymbol{K}_2(\tilde{\boldsymbol{\gamma}})|. \quad (10)$$

The Laplace approximation is believed to perform better for larger cluster sizes, and when the variance components are small (Breslow and Lin, 1995).

Note that the function $K(\gamma)$ is actually the negative penalized joint log-likelihood which can be decomposed as follows:

$$K(\gamma) = -\text{PPLL} - h(\lambda_{01}(\cdot), \lambda_{02}(\cdot), \beta_1, \beta_2, \gamma), \quad (11)$$

where PPLL represents the penalized partial log-likelihood (Section 3.2) and $h(\cdot)$ is defined in Appendix 1. In practice, inference on the regression parameters is often simplified by solely focusing on the PPLL term. Ripatti and Palmgren (2000) adopted a similar simplification, and demonstrated in their simulation studies that the information loss due to neglecting $h(\cdot)$ was negligible. Ignoring the $h(\cdot)$ and estimate the regression parameters β and the frailties γ merely based on the PPLL term largely speeds up the computations by avoiding tedious calculations of the baselines in $h(\cdot)$. Moreover, we are going to show in Section 3.2 that PPLL and $K(\gamma)$ share identical score functions if we replace the baselines with Breslow estimators.

3.2 | Penalized partial likelihood estimation

To estimate the regression coefficients and variance components, we used two iterating steps, an inner and an outer loop. In the inner loop, a Newton-Raphson algorithm is conducted based on PPLL, treating both γ and β_k as parameters, and treating $\hat{D}$ as known from the previous outer loop. The outer loop is grounded in an approximate marginal profile log-likelihood (MPLL), fixing $\hat{\theta}$ from the recent inner loop, where $\theta = (\beta_1', \beta_2', \gamma')'$. We outline the proposed algorithm below.

3.2.1 | Inner loop

Given the variance matrix $\hat{\Sigma}$ (or $\hat{D}$), the PPLL is expressd as

$$\text{PPLL} = -\frac{1}{2}\gamma'\hat{\Sigma}^{-1}\gamma + \sum_{i=1}^{n}\sum_{j=1}^{m_i+1}\sum_{k=1}^{2}\delta_{ijk}$$
$$\times \left\{\eta_{ijk} - \log\left(\sum_{l=1}^{n}\sum_{p=1}^{m_i+1}Y_{lpk}(t_{ijk})\exp(\eta_{lpk})\right)\right\}. \quad (12)$$

We define the following, for $d \in \{0, 1, 2\}$,

$$S_{Z_k}^{(d)}(t) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i+1}Y_{ijk}(t)\exp(\eta_{ijk})Z_{ijk}^{\otimes d},$$

$$S_{R_k}^{(d)}(t) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i+1}Y_{ijk}(t)\exp(\eta_{ijk})R_{ik}^{\otimes d}, \quad (13)$$

$$S_{ZR_k}(t) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i+1}Y_{ijk}(t)\exp(\eta_{ijk})Z_{ijk}R_{ik}'.$$

Let $\bar{Z}_k(t) = S_{Z_k}^{(1)}(t)/S_{Z_k}^{(0)}(t)$, $\bar{R}_k(t) = S_{R_k}^{(1)}(t)/S_{R_k}^{(0)}(t)$, $V_{Z_k}(t) = S_{Z_k}^{(2)}(t)/S_{Z_k}^{(0)}(t)$, $V_{R_k}(t) = S_{R_k}^{(2)}(t)/S_{R_k}^{(0)}(t)$ and $V_{ZR_k}(t) = S_{ZR_k}(t)/S_{Z_k}^{(0)}(t) = S_{ZR_k}(t)/S_{R_k}^{(0)}(t)$. Taking first and second partial derivatives of PPLL with respect to β_k and γ, we obtain corresponding score functions

$$\frac{\partial PPLL}{\partial \beta_k} = \sum_{i=1}^{n}\sum_{j=1}^{m_i+1}\delta_{ijk}\{Z_{ijk} - \bar{Z}_k(\tilde{T}_{ijk})\}, \quad (14)$$

$$\frac{\partial PPLL}{\partial \gamma} = \sum_{i=1}^{n}\sum_{j=1}^{m_i+1}\sum_{k=1}^{2}\delta_{ijk}\{R_{ik} - \bar{R}_k(\tilde{T}_{ijk})\} - \hat{\Sigma}^{-1}\gamma, \quad (15)$$

and information matrices

$$-\frac{\partial^2 PPLL}{\partial \beta_k \partial \beta_k'} = \sum_{i=1}^{n}\sum_{j=1}^{m_i+1}\delta_{ijk}\{V_{Z_k}(\tilde{T}_{ijk}) - \bar{Z}_k(\tilde{T}_{ijk})^{\otimes 2}\} \quad (16)$$

$$-\frac{\partial^2 PPLL}{\partial \gamma \partial \gamma'} = \sum_{i=1}^{n}\sum_{j=1}^{m_i+1}\sum_{k=1}^{2}\delta_{ijk}\{V_{R_k}(\tilde{T}_{ijk}) - \bar{R}_k(\tilde{T}_{ijk})^{\otimes 2}\} + \hat{\Sigma}^{-1} \quad (17)$$

$$-\frac{\partial^2 PPLL}{\partial \beta_k \partial \gamma'} = \sum_{i=1}^{n}\sum_{j=1}^{m_i+1}\delta_{ijk}\{V_{ZR_k}(\tilde{T}_{ijk}) - \bar{Z}_k(\tilde{T}_{ijk})\bar{R}_k(\tilde{T}_{ijk})'\}. \quad (18)$$

If we plug in Breslow estimators for $\Lambda_{0k}(t)$, $K_1(\gamma)$ is equal to $\partial PPLL/\partial\gamma$, and $\partial K(\gamma)/\partial\beta_k$ is equal to $\partial PPLL/\partial\beta_k$. These identities further justify that estimating parameters solely on $PPLL$ the inner loop is not only convenient but also plausible to some extent. It is trivial to show that $\partial^2 \text{PPLL}/\partial\beta_1\partial\beta_2'$ equals 0.

The Hessian matrix $H(\theta)$ is given by

$$H(\theta) = I(\theta) + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \hat{\Sigma}^{-1} \end{bmatrix}, \quad (19)$$

where, letting PLL be the partial log-likelihood without a penalty term, we have

$$
I(\theta) = -\frac{\partial^2 \text{PLL}}{\partial \theta \partial \theta'} = -\begin{bmatrix} \dfrac{\partial^2 \text{PLL}}{\partial \beta_1 \partial \beta_1'} & \mathbf{0} & \dfrac{\partial^2 \text{PLL}}{\partial \beta_1 \partial \gamma'} \\[2ex] \mathbf{0}' & \dfrac{\partial^2 \text{PLL}}{\partial \beta_2 \partial \beta_2'} & \dfrac{\partial^2 \text{PLL}}{\partial \beta_2 \partial \gamma'} \\[2ex] \dfrac{\partial^2 \text{PLL}}{\partial \gamma \partial \beta_1'} & \dfrac{\partial^2 \text{PLL}}{\partial \gamma \partial \beta_2'} & \dfrac{\partial^2 \text{PLL}}{\partial \gamma \partial \gamma'} \end{bmatrix}.
$$

$$(20)$$

There are two options for the asymptotic covariance estimate of $\hat\theta$, $H(\hat\theta)^{-1} I(\hat\theta) H(\hat\theta)^{-1}$ (Gray, 1992) and $H(\hat\theta)^{-1}$ (Verweij and Van Houwelingen, 1994). Since $H(\theta)^{-1}$ has been demonstrated in related contexts to be more conservative (Therneau *et al.*, 2003), $H(\theta)^{-1}$ is employed here. Note that, when the sample size is large (eg, >500 or >1000, dependent on the computer's capacity), one may sparsen the Hessian matrix before calculating the inverse. We found that the off-2 × 2-block-diagonal parts of $K_{PPL2}(\gamma)$ are dominated by the block-diagonal parts, implying that little accuracy should be lost by treating the off-2 × 2-block-diagonal part of $K_{PPL2}(\gamma)$ to be 0. This simplification results in considerable computational savings and, according to our simulation results, little reduction in accuracy.

3.2.2 | Outer loop

Fixing $\hat\theta$ from the previous inner loop, D can be estimated in the outer loop through an approximate marginal profile log-likelihood by dropping irrelevant terms from (10) (eg, not including D), such that

$$
\begin{aligned}
l_m &\approx -\frac{1}{2}\log|\Sigma| - \frac{1}{2}\log|K_2(\hat\gamma)| - K(\hat\gamma) \\
&\propto -\frac{1}{2}\log|\Sigma| - \frac{1}{2}\log|K_2(\hat\gamma)| - \frac{1}{2}\hat\gamma'\Sigma^{-1}\hat\gamma. \quad (21)
\end{aligned}
$$

We now derive an estimator for the entire variance-covariance matrix of the frailties. Given the baselines fixed, $K_2(\hat\gamma)$ is a block-diagonal matrix with each block defined as $K_2(\hat\gamma)_{ii}$, thus the marginal likelihood from (21) can be rearranged as a function of D^{-1},

$$
\frac{n}{2}\log|D^{-1}| - \frac{1}{2}\sum_{i=1}^{n}\log|K_2(\hat\gamma)_{ii}| - \frac{1}{2}\sum_{i=1}^{n}\hat\gamma_i' D^{-1}\hat\gamma_i,
$$

$$(22)$$

where we have

$$
K_2(\hat\gamma)_{ii} = D^{-1} + M_i(\hat\theta),
$$
$$
M_i(\hat\theta) = \text{diag}\left[\sum_{j=1}^{m_i+1}\Lambda_{01}(\tilde{T}_{ij1})\exp(\hat\eta_{ij1}), \sum_{j=1}^{m_i+1}\Lambda_{02}(\tilde{T}_{ij2})\exp(\hat\eta_{ij2})\right],
$$

and when the diag($\cdot$) function maps a vector to a diagonal matrix. Taking the first derivative of approximate marginal likelihood with respect to D^{-1}, we obtain the estimating equation

$$
\frac{n}{2}D - \frac{1}{2}\sum_{i=1}^{n}K_2(\hat\gamma)_{ii}^{-1} - \frac{1}{2}\sum_{i=1}^{n}\hat\gamma_i\hat\gamma_i' = 0, \quad (23)
$$

and its solution

$$
D = \frac{1}{n}\sum_{i=1}^{n}[K_2(\hat\gamma)_{ii}^{-1} + \hat\gamma_i\hat\gamma_i']. \quad (24)
$$

Let $\hat{D}_t$ denote the variance-covariance matrix estimate from the *i*th outer loop. The $(t+1)$th estimator can thus be expressed as

$$
\hat{D}_{t+1} = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\hat{D}_t^{-1} + M_i(\hat\theta)\right)^{-1} + \hat\gamma_i\hat\gamma_i'\right]. \quad (25)
$$

The variance-covariance estimator (25) is analogous to the recursive estimating formula for logistic regression models derived through a Laplace approximation (Demidenko, 2004, chapter 7.7.2). The convergence of this type recursive estimator is verified by the Fixed Point Theorem (Zamfirescu, 1972). We suggest initializing D with a diagonal matrix; for example, identity matrices were employed in our simulations. Standard errors for the variance components, if of interest, could be obtained by bootstrapping.

We found that directly replacing $\Lambda_{0k}(t)$ with its corresponding Breslow estimator tends to result in overestimation of the diagonal entries for D. Moreover, the baseline calculation could be intensive, especially when the sample size is large. Both $K_2(\gamma)$ and $K_{PPL2}(\gamma)$ are summations involving a Σ^{-1}, and their numerical difference can be captured by the second derivative of $h(\cdot)$ with respect to γ. Ripatti and Palmgren (2000) suggested to replace the $K_2(\gamma)$ with $K_{PPL2}(\gamma)$ for a more accurate estimation on the variance components, and to avoid computing the baselines for each updating step. In a similar vein, we substitute the 2 × 2 matrices located on the block-diagonal of $[K_{PPL2}(\hat\gamma)]^{-1}$, which is denoted as $[K_{PPL2}(\hat\gamma)^{-1}]_{blk_i}$, for $K_2(\hat\gamma)_{ii}^{-1}$ in (24). Subsequently, we have a new estimator

$$
\hat{D}^{\#} = \frac{1}{n}\sum_{i=1}^{n}\{[K_{PPL2}(\hat\gamma)^{-1}]_{blk_i} + \hat\gamma_i\hat\gamma_i'\}, \quad (26)
$$

which is shown to be positive-definite in Web Appendix.

3.3 | Likelihood ratio test

In order to test whether the frailties from the two alternating gap time processes are mutually independent, we propose a LRT. We evaluate an approximate MPPL,

$$\text{MPPL} = -\frac{n}{2} \log |\boldsymbol{D}| + \frac{1}{2} \sum_{i=1}^{n} |[\boldsymbol{K}_{PPL2}(\hat{\boldsymbol{\gamma}})^{-1}]_{blk_i}| + \text{PPLL}(\hat{\boldsymbol{\gamma}}), \qquad (27)$$

under the null (restricting $\boldsymbol{D}$ to be diagonal) and alternative hypotheses. Notice that, under the null, the two recurrent processes can either be fitted separately, or be fitted together while restricting the off-diagonal entries of $\boldsymbol{D}^{\#}$ to be 0 in each outer step. Let $\hat{\boldsymbol{D}}_0^{\#}$ be the variance estimator under the independence assumption,

$$\hat{\boldsymbol{D}}_0^{\#} = \text{extdiag}\left[\frac{1}{n} \sum_{i=1}^{N} \left\{ \left[\boldsymbol{K}_{PPL2}(\hat{\boldsymbol{\gamma}}_0)^{-1} \right]_{blk_i} + \hat{\boldsymbol{\gamma}}_{0i} \hat{\boldsymbol{\gamma}}_{0i}' \right\} \right], \qquad (28)$$

where extdiag($\cdot$) is a function to extract diagonal part of the matrix, distinct from diag($\cdot$) defined previously. Correspondingly, if the parameters subscripted with 1 are from the unrestricted alternative, and those subscripted with 0 are from the null, the test statistic is then given by

$$\text{LRT} = 2 \left[\text{MPPL}\left(\hat{\theta}_1, \hat{\boldsymbol{\Sigma}}_1^{\#} \right) - \text{MPPL}\left(\hat{\theta}_0, \hat{\boldsymbol{\Sigma}}_0^{\#} \right) \right]. \quad (29)$$

One would reject the null hypothesis of independence if $\text{LRT} > \chi_{(1)\alpha}^2$, where α is the prespecified type 1 error and $\chi_{(1)}^2$ is χ^2 with degree of freedom 1.

4 | SIMULATION STUDIES

Simulations under different settings were carried out to evaluate the proposed method in reasonable sample sizes. To begin, the $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2})'$ were drawn independently from a mean-zero bivariate normal distribution with various specifications for $\boldsymbol{D}$. Given $\boldsymbol{\gamma}$, event times were generated in alternating turns, added up and recorded until the censoring time $C_i = 10$. The covariates were drawn from independent standard normal distributions. The intensity function with respect to the jth occurrence of the event type k is given by $\lambda_{0k} \exp(\boldsymbol{\beta}_k' \boldsymbol{Z}_{ijk} + \gamma_{ik})$. We denote the (a, b)th entry of the variance matrix $\boldsymbol{D}$ by $\boldsymbol{D}[a, b]$. When $\boldsymbol{D}[1, 2] = 0$, the two alternating sequences are independent. We generated 500 samples and set the convergence tolerance to be 10^{-6} for each replicate.

Table 1 provides results for samples with different sizes and baseline intensities, $\lambda_{0k} = 1.5$ ($k = 1, 2$). The median number of uncensored complete recurrent event pairs was ≈ 4. We varied sample sizes from $n = 100$ to 1000. Estimated regression parameters were approximately unbiased, with asymptotic standard error (ASE) generally close to the empirical standard deviation (ESD). The results are similar across different sample sizes, implying that the estimation is not affected much by sample size if the cluster sizes or recurrent event numbers are fixed. By comparing their results, we noticed that the negatively correlated setting would experience less bias in their $\boldsymbol{D}$ matrix estimation than the positively correlated setting. For data with more recurrent events (greater m_i), the bias in estimating the $\boldsymbol{D}$ matrix decreased dramatically (Web Table 1). Note that, the ASEs for regression parameters were calculated from the inverse Hessian matrix, but the ASEs for the entries of $\boldsymbol{D}$ were obtained via bootstrapping, for which we tried different bootstrapping sample size $B = 50$, 100, and 200. We notice that using $B = 50$ provides sufficiently accurate estimates for the standard errors while takes relatively less computation time. For large n, we turned to m-out-of-n bootstrapping: resampling m observations out of n subjects with replacement (Bickel *et al.*, 1997; Bickel and Sakov, 2008). By comparing different $\sqrt{n}$ ESE of all the estimated parameters with $n = 50$, 100, 200, ..., 1000, we noticed that when $n = 200$, it started to provide stable size adjusted standard errors ($\sqrt{n}$ ESE). Thus we set $m = 200$ for data sets with $n > 200$, and the estimated ASEs from m-bootstraps were adjusted by the size difference by multiplying $\sqrt{m/n}$. The resulting coverage probabilities (CP) were acceptable for the regression parameters and variance components under various sample sizes.

The dimension of the Hessian matrix is $(2n + p) \times (2n + p)$, where n is the sample size, and p is the total regression parameters we are estimating from two event processes. Increasing the sample size would largely inflate the dimension of the matrix and consequently the computational burden. It was found that frailty part of the Hessian matrix (19) is quite sparse, thus we set its off-block-diagonal part to be 0 in order to improve the computation speed and reduce the memory usage without causing much information loss. By comparing results with vs without sparsening (Tables 1 and 2, respectively), though the coverage probabilities of the estimation from sparse computation are slightly less than those from the regular estimation procedure, the differences are negligible, especially when the sample size is larger. However, the saving of the computation burden and memory cost was dramatic. For example, a data set with sample size n will have its space needed for

TABLE 1 Simulation results: estimating regression coefficients and variance components based on 500 replicates, with $\lambda_{0k} = 1.5$ for $k = 1, 2$ and a median m_i of ≈ 5

		D_{w+}					D_{w-}			
	True value	Bias	ESD	ASE	CP	True value	Bias	ESD	ASE	CP
$n = 100$	cr: 3.81%					cr: 3.64%				
β_1	1	−0.005	0.062	0.056	0.934	1	−0.006	0.064	0.059	0.920
β_2	−1	0.003	0.062	0.059	0.940	−1	0.001	0.064	0.062	0.934
$D[1, 1]$	0.25	−0.010	0.067	0.066	0.880	0.25	−0.007	0.075	0.073	0.922
$D[2, 2]$	0.25	−0.012	0.068	0.068	0.902	0.25	−0.003	0.081	0.078	0.912
$D[1, 2]$	0.125	−0.014	0.047	0.047	0.928	−0.125	−0.008	0.053	0.052	0.934
$n = 500$	cr: 3.84%					cr: 3.70%				
β_1	1	−0.004	0.025	0.025	0.950	1	−0.004	0.028	0.026	0.922
β_2	−1	0.003	0.028	0.026	0.930	−1	0.005	0.029	0.028	0.934
$D[1, 1]$	0.25	−0.012	0.028	0.030	0.932	0.25	−0.006	0.032	0.033	0.942
$D[2, 2]$	0.25	−0.012	0.032	0.031	0.900	0.25	−0.006	0.034	0.035	0.938
$D[1, 2]$	0.125	−0.013	0.022	0.021	0.868	−0.125	−0.005	0.023	0.024	0.954
$n = 1000$	cr: 3.86%					cr: 3.68%				
β_1	1	−0.003	0.018	0.018	0.952	1	−0.004	0.019	0.019	0.924
β_2	−1	0.003	0.020	0.019	0.934	−1	0.004	0.021	0.020	0.934
$D[1, 1]$	0.25	−0.010	0.020	0.021	0.914	0.25	−0.006	0.022	0.023	0.944
$D[2, 2]$	0.25	−0.011	0.022	0.022	0.894	0.25	−0.004	0.025	0.025	0.930
$D[1, 2]$	0.125	−0.013	0.015	0.015	0.836	−0.125	−0.005	0.017	0.017	0.938

Note: Sample sizes vary from $n = 100$ to 1000 and "cr" denotes censoring rate, or the proportion of subjects with $m_i = 0$. Note that the Hessian matrix was not sparsened for these runs.

Abbreviations: ASE, asymptotic standard error; CP, coverage probabilities; cr, censoring rate; ESD, empirical standard deviation.

TABLE 2 Simulation results: Estimating regression coefficients and variance components based on 500 replicates, with $\lambda_{0k} = 1.5$ for $k = 1, 2$ and a median m_i of ≈ 5

		$D[1, 2] > 0$					$D[1, 2] < 0$			
	True value	Bias	ESD	ASE	CP	True value	Bias	ESD	ASE	CP
$n = 100$	cr: 3.81%					cr: 3.64%				
β_1	1	−0.006	0.062	0.056	0.932	1	−0.006	0.064	0.059	0.922
β_2	−1	0.004	0.062	0.059	0.940	−1	−0.000	0.064	0.062	0.932
$D[1, 1]$	0.25	−0.014	0.066	0.065	0.880	0.25	−0.010	0.074	0.072	0.916
$D[2, 2]$	0.25	−0.016	0.067	0.067	0.892	0.25	−0.007	0.079	0.077	0.900
$D[1, 2]$	0.125	−0.015	0.047	0.047	0.924	−0.125	−0.005	0.052	0.052	0.930
$n = 500$	cr: 3.84%					cr: 3.70%				
β_1	1	−0.004	0.025	0.025	0.950	1	−0.004	0.028	0.026	0.920
β_2	−1	0.003	0.028	0.026	0.928	−1	0.005	0.029	0.028	0.934
$D[1, 1]$	0.25	−0.012	0.028	0.029	0.924	0.25	−0.007	0.032	0.033	0.936
$D[2, 2]$	0.25	−0.012	0.031	0.031	0.894	0.25	−0.007	0.034	0.034	0.936
$D[1, 2]$	0.125	−0.014	0.022	0.021	0.862	−0.125	−0.004	0.023	0.023	0.950
$n = 1000$	cr: 3.86%					cr: 3.68%				
β_1	1	−0.003	0.018	0.018	0.952	1	−0.004	0.019	0.019	0.922
β_2	−1	0.003	0.020	0.019	0.934	−1	0.004	0.021	0.019	0.934
$D[1, 1]$	0.25	−0.010	0.020	0.021	0.908	0.25	−0.007	0.022	0.023	0.934
$D[2, 2]$	0.25	−0.011	0.022	0.022	0.892	0.25	−0.005	0.025	0.025	0.928
$D[1, 2]$	0.125	−0.013	0.015	0.015	0.832	−0.125	−0.005	0.017	0.017	0.938

Note: Sample sizes vary from $n = 100$ to 1000 and "cr" denotes censoring rate, or the proportion of subjects with $m_i = 0$. Note that the Hessian matrix was not sparsened for these runs.

Abbreviations: ASE, asymptotic standard error; CP, coverage probabilities; cr, censoring rate; ESD, empirical standard deviation.

a Hessian matrix of $\boldsymbol{\gamma}$ reduced from $(2n)^2 \times 8/10^6$ to $2^2 n \times 8/10^6$ MB. Thus when the sample size is large we recommend sparsening the Hessian matrix. In addition to including the same cases from Table 1, we also incorporated the cases with a large sample size $n = 6000$, a high censoring rate (cr: 75%–80%) and many covariates in Web Appendix D and Web Figures 1 to 4 to show that our method can handle data sets with high censoring rates when the sample size is about as large as the data we analyzed in Section 5.

We carried out simulations to compare the computational speed and estimation performance of the proposed method with the R package coxme (2018) implementation of Ripatti and Palmgren (2000) under different settings. For each sample, there were n individuals generated, each with a median of $\widetilde{m}_i$ complete event pairs. The regression parameters were estimated comparably well with both methods, and coxme has less bias than the proposed method with respect to the variance matrix estimator (presumably owing to having correctly pre-specified the direction of covariance component).

Examining the computation times (in seconds) listed in Table 3, our method appears to be much faster than coxme. This comparison is not completely fair considering, for example, features such as coxme's front-end error-checking capabilities. Our goal was not to compare sets of code, but to demonstrate that the improved flexibility of the proposed methods does not come at the expense of computational efficiency.

The proposed LRT was tested under different sizes, event frequencies, and covariance matrices in Table 4. We fixed the regression parameters to be $\beta_1 = 1$ and $\beta_2 = -1$ for simplicity. We focuses on a big family of variance matrices whose variance components $\boldsymbol{D}[1, 1] = \boldsymbol{D}[2, 2] = 0.5$ and the covariance component varies: $\boldsymbol{D}[1, 2] = \pm 0.25$ or ± 0.125 as representatives of strong ($|\rho| = 0.5$) or weak correlated ($|\rho| = 0.25$) event pairs for power calculations, and $\boldsymbol{D}[1, 2] = 0$ for type 1 error (T1E) evaluations. The proposed LRT test starts to performs well when the sample size and the event frequencies are not too small. In addition, if the magnitude of the correlation coefficient (ρ) is low, the

TABLE 3 Simulation results: Comparing the proposed method and coxme with respect to computational speed via average run time based on 500 replicates

		Proposed method		coxme	
	True value	**Bias**	**ESE**	**Bias**	**ESE**
$n = 50, \widetilde{m}_i = 9$	Time cost:	6.37 s		77.2 s	
β_1	1	−0.006	0.061	−0.006	0.061
β_2	−1	−0.003	0.069	0.002	0.069
$\boldsymbol{D}[1, 1]$	0.25	−0.011	0.076	−0.010	0.077
$\boldsymbol{D}[2, 2]$	0.25	−0.000	0.081	0.001	0.081
$\boldsymbol{D}[1, 2]$	−0.125	−0.006	0.057	−0.002	0.057
$n = 50, \widetilde{m}_i = 17$	Time cost:	5.31 s		136 s	
β_1	1	−0.002	0.046	−0.001	0.046
β_2	−1	−0.001	0.044	−0.001	0.044
$\boldsymbol{D}[1, 1]$	0.25	−0.000	0.069	0.002	0.070
$\boldsymbol{D}[2, 2]$	0.25	−0.001	0.067	0.001	0.068
$\boldsymbol{D}[1, 2]$	−0.125	−0.004	0.052	−0.002	0.053
$n = 100, \widetilde{m}_i = 10$	Time cost:	69.3 s		745 s	
$\beta_1[1]$	−1	−0.004	0.040	−0.004	0.040
$\beta_1[2]$	−0.5	−0.003	0.032	−0.003	0.032
$\beta_1[3]$	0	−0.001	0.031	−0.001	0.031
$\beta_1[4]$	0.5	−0.000	0.034	−0.000	0.034
$\beta_1[5]$	1	0.005	0.042	0.005	0.042
$\beta_2[1]$	−1	−0.001	0.041	−0.001	0.041
$\beta_2[2]$	−0.5	−0.004	0.034	−0.004	0.031
$\beta_2[3]$	0	0.000	0.031	0.000	0.034
$\beta_2[4]$	0.5	0.002	0.034	0.002	0.034
$\beta_2[5]$	1	0.001	0.041	0.001	0.041
$\boldsymbol{D}[1, 1]$	0.25	−0.004	0.051	−0.001	0.053
$\boldsymbol{D}[2, 2]$	0.25	−0.008	0.051	−0.003	0.053
$\boldsymbol{D}[1, 2]$	0.125	−0.007	0.038	0.001	0.039

Abbreviation: ESD, empirical standard deviation.

TABLE 4 Power and type I error of the proposed likelihood ratio test, with $\lambda_{01} = \lambda_{02}$, and correlation coefficients $\rho = \pm 0.5$ corresponding to $D[1, 2] \pm 0.25$ or ± 0.125 for $D[1, 1] = D[2, 2] = 0.5$ or 0.25, and $\rho = 0.25$ corresponding to $D[1, 2] = \pm 0.125$ for $D[1, 1] = D[2, 2] = 0.5$

			T1E	Power			
$D[1, 1] = D[2, 2]$	n	$\tilde{m}$	$\rho = 0$	$\rho = 0.5$	$\rho = -0.5$	$\rho = 0.25$	$\rho = -0.25$
0.5	50	4	0.066	0.656	0.646	0.236	0.174
	100	4	0.038	0.902	0.918	0.394	0.366
	100	9	0.042	0.976	0.982	0.516	0.482
0.25	50	5	0.058	0.498	0.436		
	100	5	0.054	0.754	0.758		
	100	9	0.060	0.942	0.930		

proposed LRT is less likely to correctly detect the existence of a non-zero covariance; but when ρ is identical, a larger covariance entry improved the detection of the dependence (compared with $D[1, 1] = D[2, 2] = 0.25$ but $|\rho| = 0.5$ or $D[1, 2] = \pm 0.125$). According to our simulations, T1Es were be close to the nominal value ($\alpha = 0.05$) in most settings.

5 | APPLICATION

The DOPPS is a well-known prospective, longitudinal, international study of hemodialysis patients. This study aims to improve the understanding of dialysis practices that are associated with better outcomes for end-stage renal disease patients. Details regarding the DOPPS study can be found in several reports (Young *et al.*, 2000; Pisoni *et al.*, 2004; Robinson *et al.*, 2012). Mortality, hospital admission and inpatient stay are important indicators of quality of life, and morbidity related outcomes have arguably been under-utilized in the DOPPS and other studies of ESRD patients.

We applied our proposed methods to jointly analyze the time-to-readmission and time-to-discharge (from admission) alternating gap time sequence. Our objective was to determine the important predictors for each recurrent event process, and to quantify the correlation between the two processes. Our study population ($n = 6032$) included DOPPS Phase-5 adult patients (age $\geqslant 18$) who entered the DOPPS within 3 months of initiating hemodialysis. Each member of the study cohort was followed for a maximum of 3 years, with the database closing on 31 December 2015. The study population included patients from 11 different countries, including Belgium, Canada, China, Gulf Coast Consortium, Germany, Italy, Japan, Spain, Sweden, the United Kingdom, and the United States. The median age among DOPPS patients was 67, with 39.5% being female.

Our primary goal was to compare the hospital admission and the discharge event rates among dialysis patients by country. In particular, Belgium, Canada, China, Gulf Coast Consortium, Germany, Italy, Japan, Spain, Sweden, the United Kingdom, Asian-American and African-American are compared to the US Caucasians (reference). Adjustment covariates included age, sex, height, vascular access (arteriovenous [AV] graft, central venous catheter, with AV fistula as the reference), and the following comorbid condition indicators: coronary artery disease [CAD], cancer, cerebral vascular disease (CVD), congestive heart failure symptoms (CHF), chronic obstructive pulmonary disease (COPD), peripheral vascular disease (PVD), stroke, diabetes, hypertension, neurological disorder, psychological disorder, and cellulitis. Table 5 lists results based on our model (2). DOPPS patients from Belgium ($e^{0.386} = 1.47$), Germany ($e^{0.98} = 2.66$), Italy ($e^{0.360} = 1.43$), Japan ($e^{0.842} = 2.32$), Sweden ($e^{0.507} = 1.66$), and the United Kingdom ($e^{0.534} = 1.71$) had significantly higher covariate-adjusted hospital admission rates than US Caucasians. In contrast, the hospital admission rates for patients in China ($e^{-0.621} = 0.537$) was approximately half that of US Caucasians. With respect to length of hospital stay, patients from Canada ($e^{-0.807} = 0.446$), China ($e^{-1.198} = 0.302$), Germany ($e^{-0.433} = 0.649$), Italy ($e^{-0.667} = 0.513$), Japan ($e^{-0.624} = 0.526$), Spain ($e^{-0.456} = 0.634$), and the United Kingdom ($e^{-0.470} = 0.625$) had lower discharge rates (implying longer hospital stay) than US Caucasians. We did not observe significant differences in the US among races for either hospital admission or discharge rates.

Comorbid conditions were generally positively associated with hospital admission and negatively associated with hospital discharge. Common significant predictors for both episodes include cancer and neurological disorder. CAD, CVD, stroke, COPD, and psychological disorder were associated with significantly increased hospital admission rates, while cellulitis was significantly associated with increased discharge rates. Note that CAD ($P = .056$) was marginally significantly associated with discharges. The impact of age was found to be negatively associated with both the admission and discharge,

TABLE 5 Application of the proposed method to DOPPS data: estimated regression parameters (bolded when $P < .05$)

	Admission			Discharge		
	Estimate	SÊ	*P* value	Estimate	SÊ	*P* value
Age (per 5 yr)	**−0.026**	0.010	.007	**−0.050**	0.010	<.001
Height (per 5 cm)	−0.028	0.017	.103	−0.005	0.018	.804
Female	0.019	0.069	.782	−0.065	0.073	.373
Vascular access						
Arteriovenous graft	0.524	0.156	.001	−0.034	0.149	.822
Central venous catheter	0.783	0.059	<.001	0.033	0.060	.583
Comorbid conditions						
CAD	0.447	0.069	<.001	−0.130	0.068	.056
Cancer	0.214	0.082	.009	**−0.208**	0.082	.011
CVD	0.177	0.076	.020	−0.070	0.075	.345
Stroke	0.190	0.090	.034	−0.004	0.088	.968
CHF	0.078	0.068	.251	0.028	0.070	.692
Diabetes	0.053	0.056	.347	−0.072	0.060	.228
Hypertension	0.017	0.068	.802	0.111	0.076	.143
COPD	0.264	0.090	.003	−0.075	0.087	.387
Neurological disorder	0.373	0.101	<.001	**−0.341**	0.096	<.001
Psychological disorder	0.293	0.090	.001	−0.076	0.088	.389
PVD	0.111	0.078	.156	0.124	0.079	.115
Cellulitis	0.169	0.131	.198	**−0.454**	0.126	<.001
Countries						
Belgium	0.386	0.127	.002	−0.041	0.125	.741
Canada	0.234	0.125	.060	**−0.807**	0.124	<.001
China	**−0.621**	0.231	.007	**−1.198**	0.253	<.001
Gulf	−0.069	0.131	.596	−0.054	0.138	.693
Germany	0.980	0.100	<.001	**−0.433**	0.097	<.001
Italy	0.360	0.127	.005	**−0.667**	0.130	<.001
Japan	0.842	0.100	<.001	**−0.624**	0.107	<.001
Spain	−0.138	0.127	.274	**−0.456**	0.135	.001
Sweden	0.507	0.130	<.001	−0.115	0.132	.382
UK	0.534	0.135	<.001	**−0.470**	0.140	.001
USA: Asian	−0.148	0.305	.628	−0.214	0.354	.546
USA: African–American	−0.035	0.089	.693	−0.020	0.102	.848
USA: Caucasian	0	–	–	0	–	–

though the difference for every 5-year increment was small (3%–5%). Every 5 cm increment in height was associated with a 3% decrease in the hospitalization risk. In comparison to AV fistula (the most commonly adopted vascular access approach) AV graft and central venous catheter increased the rate hospital admission by 1.69 and 2.19 times, respectively. Note that each of the regression parameters should be interpreted as a conditional effect, given the unobserved frailties.

The estimated variance for time to admission, at 0.819, with 95% confidence interval (0.709, 0.929), was larger than the estimated frailty variance for length-of-stay, 0.375 (0.253, 0.497). The estimated covariance of −0.139 (−0.229, −0.049) implies that the two events were negatively correlated with a correlation coefficient $\hat{\rho} = -0.251$. Note that the standard errors for variance components were estimated through standard bootstrap due to the presence of many unbalanced binary variables. The LRT for the dependence between two event processes was 5.41 with a P of .02. These results indicate

that those who had more frequent admission to a hospital would (through a lower discharge rate) tend to have a longer inpatient stay, and the association is significant.

Note that only 21.3% of patients experienced more than one hospitalization. The average length of stay in hospital was 8.8 days (range = 1–331; median = 5). We tested the validity to implement the proposed estimating method on huge data sets with a small proportion of subjects that experienced multiple recurrences. The proposed methods appear to work well in data structures resembling DOPPS with respect to event rates and the variance-covariance matrix (Table 2, Web Appendix D and Web Figures 1–4). The convergence tolerance for parameter estimation is 10^{-6}.

6 | DISCUSSION

In this report, we propose a correlated bivariate frailty model for alternating recurrent event (gap time)

processes. We also derived a variance-covariance estimator of the bivariate frailty in a recursive estimating formula. Through simulations, the methods were demonstrated to work well for both regression parameters and variance components. We also developed a LRT to evaluate whether joint modeling of the two gap time processes is necessary. This is important in practice, since fitting two separate frailty models requires less computation than fitting a joint frailty model. The proposed methods were applied to simultaneously analyze time-to-readmission and length-of-stay among end-stage renal disease patients in the DOPPS study.

The proposed estimating approach does not require prespecifying the sign of the correlation between the two recurrent event processes. In the context of hospitalization data, it is possible that the longer length-of-stay tends to accompany hospitalization for a more severe episode which, in turn, can be associated with shorter time to readmission. In contrast, longer hospital stay could reflect better care and, as such, be associated with longer time to readmission.

To be able to analyze large data sets, part of our program is written in the R package RcppArmadillo (2018). Besides, the proposed computing algorithm yields very reasonable computation times. When the sample size is large, we recommend using the "sparse" Hessian matrix option to reduce the memory cost. The current version has been tested for data sets with sample sizes up to 100 000.

The methods in this report could be extended to other settings; for example, clustered survival outcomes. Technically, the method can be extended to fit more than two event types. Furthermore, although the frailties we consider primarily in this report represent subject-specific intercepts, one could also include frailties corresponding to one or more covariates.

ACKNOWLEDGMENTS

ORCID

Lili Wang http://orcid.org/0000-0003-4276-3930
Kevin He http://orcid.org/0000-0002-8354-426X
Douglas E. Schaubel http://orcid.org/0000-0002-9792-4474

REFERENCES

Andersen, P.K. and Gill, R.D. (1982) Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10, 1100–1120.

Bickel, P., Götze, F. and vanZwet, W. (1997) Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica*, 7, 1–31.

Bickel, P.J. and Sakov, A. (2008) On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*, 18, 967–985.

Breslow, N.E. and Lin, X. (1995) Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, 82, 81–91.

Demidenko, E. (2004) *Mixed Models: Theory and Applications*. New York, NY: Wiley-Interscience.

Dharmarajan, S.H., Schaubel, D.E. and Saran, R. (2018) Evaluating center performance in the competing risks setting: application to outcomes of wait-listed end-stage renal disease patients. *Biometrics*, 74, 289–299.

Gray, R.J. (1992) Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87, 942–951.

Huang, C.-Y. and Wang, M.-C. (2005) Nonparametric estimation of the bivariate recurrence time distribution. *Biometrics*, 61, 392–402.

Huang, X. and Liu, L. (2007) A joint frailty model for survival and gap times between recurrent events. *Biometrics*, 63, 389–397.

Kalbfleisch, J.D. and Prentice, R.L. (2002) *The Statistical Analysis of Failure Time Data*. New York, NY: Wiley.

Lee, C.H., Huang, C.-Y., Xu, G. and Luo, X. (2018) Semiparametric regression analysis for alternating recurrent event data. *Statistics in Medicine*, 37, 996–1008.

Lin, D., Wei, L., Yang, I. and Ying, Z. (2000) Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62, 711–730.

Liu, L. and Huang, X. (2008) The use of gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine*, 27, 2665–2683.

Pisoni, R.L., Gillespie, B.W., Dickinson, D.M., Chen, K., Kutner, M.H. and Wolfe, R.A. (2004) The dialysis outcomes and practice patterns study: design, data elements, and methodology. *American Journal of Kidney Diseases*, 44, 7–15.

R Package coxme. (2018) coxme: mixed effects Cox models, version: 2.2-5 to 2.2-10. Available at: https://cran.r-project.org/web/packages/coxme/. [Accessed 20 February 2019].

R Package RcppArmadillo. (2018) RcppArmadillo: 'Rcpp' integration for the 'Armadillo' templated linear algebra library, version: 0.7.960.1.2 to 0.9.200.4.0. Available at: https://cran.r-project.org/web/packages/RcppArmadillo/. [Accessed 20 February 2019].

Ripatti, S., Larsen, K. and Palmgren, J. (2002) Maximum likelihood inference for multivariate frailty models using an automated monte carlo em algorithm. *Lifetime Data Analysis*, 8, 349–360.

Ripatti, S. and Palmgren, J. (2000) Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56, 1016–1022.

Robinson, B.M., Bieber, B., Pisoni, R.L. and Port, F.K. (2012) Dialysis outcomes and practice patterns study: its strengths, limitations, and role in informing practices and policies. *Clinical Journal of the American Society of Nephrology*, 7, 1897–1905.

Therneau, T.M., Grambsch, P.M. and Pankratz, V.S. (2003) Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, 12, 156–175.

Vaida, F. and Xu, R. (2000) Proportional hazards model with random effects. *Statistics in Medicine*, 19, 3309–3324.

Verweij, P.J. and VanHouwelingen, H.C. (1994) Penalized likelihood in cox regression. *Statistics in Medicine*, 13, 2427–2436.

Xue, X. and Brookmeyer, R. (1996) Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis*, 2, 277–289.

Yashin, A.I., Vaupel, J.W. and Iachine, I.A. (1995) Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, 5, 145–159.

Young, E.W., Goodkin, D.A., Mapes, D.L., Port, F.K., Keen, M.L., Chen, K. *et al.* (2000) The dialysis outcomes and practice patterns study: an international hemodialysis study. *Kidney International*, 57, S74–S81.

Zamfirescu, T. (1972) Fix point theorems in metric spaces. *Archiv der Mathematik*, 23, 292–298.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 3, 4, and 5 are available with this paper at the Biometrics website on Wiley Online Library. Code and an example data set are also available. The R package to implement the proposed method is available at https://github.com/lilywang1988/BivPPL.