Published in final edited form as:

Stat Med. 2018 August 15; 37(18): 2771–2786. doi:10.1002/sim.7682.

Joint modeling of recurrent events and a terminal event adjusted for zero inflation and a matched design

Cong Xu, Vernon M. Chinchilli, Ming Wang

Division of Biostatistics and Bioinformatics, Department of Public Health Sciences, Pennsylvania State Hershey Medical Center, Hershey, PA 17033, USA

Abstract

In longitudinal studies, matched designs are often employed to control the potential confounding effects in the field of biomedical research and public health. Because of clinical interest, recurrent time-to-event data are captured during the follow-up. Meanwhile, the terminal event of death is always encountered, which should be taken into account for valid inference because of informative censoring. In some scenarios, a certain large portion of subjects may not have any recurrent events during the study period due to nonsusceptibility to events or censoring; thus, the zero-inflated nature of data should be considered in analysis. In this paper, a joint frailty model with recurrent events and death is proposed to adjust for zero inflation and matched designs. We incorporate 2 frailties to measure the dependency between subjects within a matched pair and that among recurrent events within each individual. By sharing the random effects, 2 event processes of recurrent events and death are dependent with each other. The maximum likelihood based approach is applied for parameter estimation, where the Monte Carlo expectation-maximization algorithm is adopted, and the corresponding R program is developed and available for public usage. In addition, alternative estimation methods such as Gaussian quadrature (PROC NLMIXED) and a Bayesian approach (PROC MCMC) are also considered for comparison to show our method's superiority. Extensive simulations are conducted, and a real data application on acute ischemic studies is provided in the end.

Keywords

death; frailty models; joint modeling; Monte Carlo expectation-maximization algorithm; recurrent events; zero inflation

1 | INTRODUCTION

Recurrent events occur frequently during the follow-up period in longitudinal clinical studies. For example, patients may experience relapsed tumors, recurrent strokes, or hypoglycemia events on multiple occasions. In many instances, the terminal event of death is commonly encountered during follow-up, which prevents the observations and even the occurrence of any further recurrent events, but not vice versa.¹ Thus, the common

assumption of independent censoring for recurrent events is violated because of the competing risk of death because these 2 event processes are often correlated. For instance, if recurrent events (eg, heart attacks) have a substantially negative effect on health condition, then the hazard for death could be diminished. The association between recurrent events and death has attracted increasing interest recently, and a joint analysis taking their correlation into account is needed for valid inference.²

Regarding the joint analysis of recurrent events and death, 2 general approaches can be adopted to accommodate the dependent censoring, namely, marginal models and frailty models. Marginal models regard the terminating event of death as a censoring event for each recurrent event, or estimate the marginal mean of the cumulative number of recurrent events over time.³⁻⁵ However, the dependence between recurrent events and death cannot be specified by a marginal model.⁶ Moreover, copula models could be considered because of their convenience, but undesirable issues still exist, such as strong parametric assumptions on the association structure, and misspecification of the copula may lead to unreliable estimates.⁷ Frailty models or shared random effects models, not only capture the correlation among recurrent events but also incorporate the dependence between 2 event processes of recurrent events and death. Lancaster and Intrator² initially considered a correlation between recurrent events and death via a person-specific frailty term by proposing a joint parametric modeling of the repeated inpatient episodes (via Poisson process) and survival time in a human immunodeficiency virus study. Huang and Wolfe⁸ proposed a joint frailty model for clustered data with informative censoring, where the risk to be censored was affected by the risk of failure by sharing the common log normal frailty. Liu et al⁶ proposed joint models by incorporating a shared gamma frailty that was included in both intensity functions of recurrent events and death to account for their dependency. Zeng and Lin⁹ generalized joint frailty models of recurrent events and death using a broad class of transformation models, rather than Cox proportional hazards (PH) models.

In this paper, our motivation example explores recurrent acute ischemic stroke. As is known, stroke leads to huge economical burden with the estimated direct and indirect cost as 33 billion dollars from 2011 to 2012, and also stroke is the third leading cause of death in the United States.¹⁰ Literatures have shown that recurrent stroke can be one of the major causes of morbidity and mortality among stroke victims, and provided valuable information about the risk of stroke recurrence in stroke survivors and the associated effects of comorbidities.^{11,12} However, most studies adopted the basic Cox PH model to analyze the stroke recurrence without considering the competing risk of death or a matched design because the majority are randomized trials or epidemiological studies. Our motivation data are a matched cohort of patients with and without acute ischemic stroke at baseline from MarketScan research database year 2011 to 2014.¹³ The *International Classification of Diseases, Ninth Revision* code is used to identify the episode of acute ischemic stroke, and the 1:1 exact matching strategy is performed based on several important confounding variables, such as age, gender, admission date, and the follow-up window. Our objective is to investigate the effect of initial stroke detected at baseline and comorbidity (eg, diabetes mellitus, hypertension, and heart disease) on the risk of recurrent stroke occurrence and death during index hospitalization, and also explore the dependency between recurrent stroke events and death.

Regarding this joint analysis, 2 main issues require attention, namely, zero-inflated recurrent events and the matched design. In our study, recurrent stroke events are observed in 597 out of 2122 patients; thus, a substantial portion of patients (71.8%) have no recurrent stroke events during the follow-up period due to nonsusceptibility to events or censoring; thus, the zero-inflated nature of data should be considered in analysis. Zero-inflated Poisson or negative binomial models have been proposed and popularly used for count data to handle the overdispersion.^{14,15} Regarding zero-inflated recurrent events data, the models have been extended from the cure models for single-event data, where the population consists of a mixture of noncured (susceptible to the event) and cured subjects (nonsusceptible to the event). For estimation of the cure rate and survival distribution of the noncured subjects, a logistic regression for mixture proportions has been widely considered.¹⁶⁻¹⁹ Recently, Rondeau et al²⁰ and Liu et al¹⁵ extended the Cox PH cure models to recurrent events by incorporating the frailties. The other issue is the matched design, where the correlation within each matched pair should not be ignored; otherwise, the inference will be invalid. The common strategy is to incorporate random effects (ie, frailties) to capture it.¹⁴ Therefore, a hierarchical or nested structure of correlation will be employed, where the frailties that takes the correlation among recurrent stroke events are nested to the ones that capture matched-pair correlation. In recent years, such kind of clinical studies are commonly encountered indicating the needs of advanced model development, for instance, a child survival study in Northeast Brazil with data being collected according to families and communities,²¹ and the assessment, serial evaluation, and subsequent sequelae of acute kidney injury study in which patients are matched by baseline acute kidney injury status within each stratum according to baseline chronic kidney disease and clinical research center.²² Several studies using nested frailty models to account for the hierarchical clustering of the data have been discussed, but limited work exist for recurrent time-to-event data.^{21,23,24}

In this article, due to specific issues on zero-inflated nature of data and matched designs, we propose a novel joint frailty modeling for recurrent events and a terminal event of death by accommodating a conditional logistic model for subjects with zero recurrent events and nested frailties accounting for the hierarchical correlation structure. In particular, 2 nested frailties are to measure the dependency (1) between stroke and nonstroke subjects within a matched pair and (2) among recurrent stroke events within one individual. By sharing these frailties, death is dependent on recurrent stroke event process. The remainder of this article is organized as follows. In Section 2, we introduce the joint modeling of recurrent events and death in a matched cohort study. We provide the theoretical work on parameter estimation via the Monte Carlo expectation-maximization (MCEM) algorithm in Section 3. Results from simulation and a real data analysis are presented in Sections 4 and 5, respectively. We summarize the conclusions and a discussion of future work in Section 6.

2 | JOINT MODELING

2.1 | Notations

Let i denote the cluster index (ie, matched pairs based on our motivation example), and j denote the subject index, $i = 1, 2, \dots, J$; $j = 1, 2, \dots, J_i$. Because of the simplicity and our motivation data structure, we consider balanced cluster size; thus, $J_i = J$, where J is a

constant number (ie, $J = 2$ based on our motivation example); thus, the total sample size $N = I \times J$; however, the generalization of our approach to accommodate unbalanced cluster size is straightforward. For the ij th subject, let C_{ij} , M_{ij} , and D_{ij} be the follow-up time, drop-out time, and the death time, respectively. Let $T_{ij} = \min(D_{ij}, C_{ij}, M_{ij})$ be the observed follow-up time with $\delta_{ij} = I(D_{ij} = \min(C_{ij}, M_{ij}))$ as the death indicator, where $I(\cdot)$ is the indicator function. Denote $\Psi_{ij}(t) = I(T_{ij} \geq t)$ as the “at-risk” indicator to show whether the subject is still under observation at time t or not. X_{ij} denotes the vector of the observed covariates including baseline risk factors, such as gender, race, and smoking status and so on. Currently, we consider time-independent covariates, but the time-varying covariates can be incorporated in a straightforward manner.

Define $N_{ij}^D(t) = I(D_{ij} \leq t, \delta_{ij} = 1)$ as the observed death process, and $N_{ij}^{D*}(t) = I(D_{ij} \leq t)$ as the actual death process by time point t . In addition, denote $N_{ij}^{R*}(t) = N\{\min(D_{ij}, t)\}$ and $N_{ij}^R(t) = N\{\min(T_{ij}, t)\} = N_{ij}^{R*}(\min(T_{ij}, t))$ as the actual and observed number of recurrent events by time point t separately. N_{ij}^D and N_{ij}^R are both the observed parts of the counting processes N_{ij}^{D*} and N_{ij}^{R*} . The number of recurrent events that occur for the ij th subject over the small interval $[t, t + dt)$ is $dN_{ij}^{R*}(t) = N_{ij}^{R*}((t + dt)^-) - N_{ij}^{R*}(t^-)$ as $dt \rightarrow 0$ and $dN_{ij}^R = \Psi_{ij}(t)dN_{ij}^{R*}$. $R_{ij} = I(N_{ij}^R > 0)$ denotes the observed indicator for whether the ij th subject has at least one recurrent stroke during the follow-up. Let t_{ijk} be the k th recurrent event time and δ_{ijk} denote the indicator of recurrent events at time t_{ijk} , $k = 1, 2, \dots, N_{ij}^R$. Therefore, the data of the ij th individual at time t is then $O_{ij}(t) = \{\Psi_{ij}(u), N_{ij}^D(u), N_{ij}^R(u), 0 < u \leq t\}$. Then the entire set of observed data for individual ij is $O_{ij} = \{O_{ij}(u), 0 < u \leq T_{ij}\}$.

2.2 | The joint frailty model of zero-inflated recurrent events and death

As we mentioned previously, there may be a high proportion patients (ie, $> 50\%$) who didn't experience any recurrent events by the end of follow-up period. Let Y_{ij} denote the indicator that the ij th subject will eventually ($Y_{ij} = 1$) or never ($Y_{ij} = 0$) experience the recurrent events, with probability $p_{ij} = Pr(Y_{ij} = 1)$. Define y_{ij} as the value taken by the random variable Y_{ij} . It follows that, if $R_{ij} = 1$, $y_{ij} = 1$ and if $R_{ij} = 0$, y_{ij} is unobserved, then this individual could fall in either of the 2 groups. A logistic regression model is used to model the probability of susceptible or not cured p_{ij} :

$$p_{ij} = \frac{\exp(\beta_1^T \tilde{X}_{ij} + \mu_i)}{1 + \exp(\beta_1^T \tilde{X}_{ij} + \mu_i)}, \quad (1)$$

where β_1 is the vector of regression coefficients, indicating the effect of potential covariates $\tilde{X}_{ij}$. Note that the intercept term is absorbed into $\tilde{X}_{ij}$, which could be overlapped with X_{ij} .

Let μ_i denote the frailty measuring the dependency between subjects within a matched pair. Y_{ij} from individuals in the same matched pair tend to be correlated and that also can be captured by μ_i . Denote ω_{ij} as the frailty measuring the dependency among the recurrent events within the ij th individual. μ_i and ω_{ij} are shared by the recurrent and terminal events to

induce their dependency. Assume that $(\mu_i, \omega_{ij})^T \sim MVN(0, \Sigma)$. For simplicity, we can assume that $\mu_i \perp \omega_{ij}$, where $\mu_i \sim N(0, \sigma_\mu^2)$ and $\omega_{ij} \sim N(0, \sigma_\omega^2)$. If σ_μ^2 and σ_ω^2 equal 0, then it implies there is no dependency between recurrent events and death, and the heterogeneity in both event processes is solely explained by covariate $\mathbf{X}_{ij}$. Conditional on frailties μ_i , ω_{ij} , and covariates $\mathbf{X}_{ij}$, the zero-inflated model for recurrent events is defined as follows, which is a special case of latent models with 2 classes:

$$\lambda_{ij}^R(t | \mathbf{X}_{ij}, \mu_i, \omega_{ij}) = \begin{cases} \lambda_0^R(t) \exp(\boldsymbol{\beta}_2^T \mathbf{X}_{ij} + \mu_i + \omega_{ij}) & \text{with probability } p_{ij}(Y_{ij} = 1) \\ 0 & \text{with probability } 1 - p_{ij}(Y_{ij} = 0), \end{cases} \quad (2)$$

where $\boldsymbol{\beta}_2$ is the vector of regression coefficients and $\lambda_0^R(t)$ is the baseline intensity function. Conditional on μ_i and ω_{ij} , the recurrent event process and death are independent of each other. The association between death and recurrent events is quantified by the shared frailties μ_i and ω_{ij} through ϕ_μ and ϕ_ω in Equation 3, and a higher intensity of recurrent events is associated with a higher hazard rate for mortality. Thus, the hazard function for death is given by

$$\lambda_{ij}^D(t | \mathbf{X}_{ij}, \mu_i, \omega_{ij}) = \lambda_0^D(t) \exp(\boldsymbol{\beta}_3^T \mathbf{X}_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij}), \quad (3)$$

where $\boldsymbol{\beta}_3$ is the vector of regression coefficients and $\lambda_0^D(t)$ is the baseline hazard function of death. Of note, $\mathbf{X}_{ij}$ is assumed to be the same for λ_{ij}^R and λ_{ij}^D for simplicity, but can be different due to clinical perspectives in data applications.

Combining Equations 2 and 3, we have a joint model of recurrent events and death adjusted for zero inflation and a matched design. Given $\boldsymbol{\theta} = (\lambda_0^R(\cdot), \lambda_0^D(\cdot), \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \phi_\mu, \phi_\omega, \sigma_\mu^2, \sigma_\omega^2)^T$, the marginal likelihood is

$$\begin{aligned} L(\boldsymbol{\theta}) &= \int \mu_i \int \omega_{ij} \prod_{i=1}^I \prod_{j=1}^J L(\boldsymbol{\theta} | \mu_i, \omega_{ij}) f(\mu_i) f(\omega_{ij}) d\mu_i d\omega_{ij} \\ &= \prod_{i=1}^I \int \mu_i \left\{ \prod_{j=1}^J \int \omega_{ij} ((L_{ij}^0)^{1-R_{ij}} (L_{ij}^1)^{R_{ij}} L_{ij}^D f(\omega_{ij}) d\omega_{ij} f(\mu_i) \right\} d\mu_i, \end{aligned} \quad (4)$$

with

$$\begin{aligned} L_{ij}^0 &= 1 - p_{ij} + p_{ij} \exp\left[-\int_0^\infty \Psi_{ij}(t) \lambda_0^R(t) dt \exp(\boldsymbol{\beta}_2^T \mathbf{X}_{ij} + \mu_i + \omega_{ij})\right], \\ L_{ij}^1 &= p_{ij} \prod_k [\lambda_0^R(t) \exp(\boldsymbol{\beta}_2^T \mathbf{X}_{ij} + \mu_i + \omega_{ij})]^{\delta_{ijk}} \exp\left[-\int_0^\infty \Psi_{ij}(t) \lambda_0^R(t) dt \exp(\boldsymbol{\beta}_2^T \mathbf{X}_{ij} + \mu_i + \omega_{ij})\right], \\ L_{ij}^D &= [\lambda_0^D(T_{ij}) \exp(\boldsymbol{\beta}_3^T \mathbf{X}_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij})]^{A_{ij}} \exp\left[-\int_0^\infty \Psi_{ij}(t) \lambda_0^D(t) dt \exp(\boldsymbol{\beta}_3^T \mathbf{X}_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij})\right], \end{aligned}$$

where L_{ij}^0 is the likelihood of observing zero recurrent events ($R_{ij}=0$), which includes those participants who are at risk for recurrent events but censored or died before the first event.

L_{ij}^1 , is the likelihood of observing at least one recurrent event ($R_{ij} = 1$). L_{ij}^D is the likelihood for the terminal event of death. $f(\mu_i)$ and $f(\omega_{ij})$ are normal densities. In addition, the cumulative baseline intensities of recurrent events and terminal event denoted by $\Lambda_0^R(t)$ and $\Lambda_0^D(t)$ are given by $\Lambda_0^R(t) = \int_0^\infty \Psi_{ij}(t) \lambda_0^R(t) dt$ and $\Lambda_0^D(t) = \int_0^\infty \Psi_{ij}(t) \lambda_0^D(t) dt$.

3 | ESTIMATION

The parameter estimation can be achieved by maximizing $L(\theta)$. There are several commonly approaches used for estimation in the literature. The first approach is based on numerical integration techniques, such as Gaussian quadrature, to integrate out the frailties and then maximize the integrated likelihood. This approach can be easily implemented in SAS PROC NLMIXED and has been adopted by many researchers.^{15,20,25} Another approach is a Bayesian framework with Markov chain Monte Carlo (MCMC) technique, which has emerged as a popular tool of joint frailty analysis.²⁶⁻²⁸ The MCMC algorithm is able to draw inferences from a complex posterior distribution on a high-dimensional parameter space. Another alternative approach, the most popular for estimation in joint random effects models, is the EM algorithm.^{6,9,18,29-31} The EM algorithm iteratively computes maximum likelihood estimates from an incomplete data set by treating the frailties as missing data. In some cases, the joint likelihood functions involving multiple integrals do not yield closed-form expressions or analytic solutions for parameter estimation. Therefore, several numerical integration techniques, such as Monte Carlo^{6,18,31} or Laplace approximation,^{23,32} can be used.

In this paper, we adopt the MCEM algorithm for estimation. In the E-step, we find the expectation of the conditional log likelihood, and integrate out the frailties by numerical integration with the Metropolis-Hasting algorithm. We define $q_{1ij} = \exp(\beta_1^T \tilde{X}_{ij} + \mu_i)$, $q_{2ij} = \exp(\beta_2^T X_{ij} + \mu_i + \omega_{ij})$, and $q_{3ij} = \exp(\beta_3^T X_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij})$. In addition, the cumulative baseline intensities of recurrent events and the terminal event, denoted by $\Lambda_0^R(t)$ and $\Lambda_0^D(t)$, are given by $\Lambda_0^R(t) = \int_0^\infty \Psi_{ij}(t) \lambda_0^R(t) dt$ and $\Lambda_0^D(t) = \int_0^\infty \Psi_{ij}(t) \lambda_0^D(t) dt$. Thus given the values of latent variable y , frailties μ , and ω , the complete log-likelihood function denoted by $l(\theta)$ can be split into 4 parts:

$$\begin{aligned} l(\theta) = & \sum_{i=1}^I \sum_{j=1}^J \log \left\{ p_{ij}^{y_{ij}} (1 - p_{ij})^{1 - y_{ij}} \right\} \\ & + \sum_{i=1}^I \sum_{j=1}^J \log \left\{ \left(\prod_k (\lambda_0^R(t_{ijk}) \cdot q_{2ij})^{\delta_{ijk}} \right)^{R_{ij} y_{ij}} (\exp(-\Lambda_0^R(t) \cdot q_{2ij}))^{y_{ij}} \right\} \\ & + \sum_{i=1}^I \sum_{j=1}^J \log \left\{ (\lambda_0^D(T_{ij}) \cdot q_{3ij})^{\Delta_{ij}} \exp(-\Lambda_0^D(t) \cdot q_{3ij}) \right\} \\ & + \sum_{i=1}^I \sum_{j=1}^J \log \left\{ \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{\mu_i^2}{2\sigma_\mu^2}\right) \frac{1}{\sqrt{2\pi\sigma_\omega^2}} \exp\left(-\frac{\omega_{ij}^2}{2\sigma_\omega^2}\right) \right\} \\ = & l_1(\beta_1) + l_2(\beta_2, \lambda_0^R(t)) + l_3(\beta_3, \lambda_0^D(t), \phi_\mu, \phi_\omega) + l_4(\sigma_\mu^2, \sigma_\omega^2). \end{aligned} \quad (5)$$

The expectation of the log-likelihood $l(\boldsymbol{\theta})$ conditional on the observed data and the current parameter estimate $\hat{\boldsymbol{\theta}}^{(k)}$ is

$$\begin{aligned} Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}) &= E[l(\boldsymbol{\theta}) | \hat{\boldsymbol{\theta}}^{(k)}] \\ &= E[l_1(\boldsymbol{\beta}_1) | \hat{\boldsymbol{\theta}}^{(k)}] + E[l_2(\boldsymbol{\beta}_2, \lambda_0^R(t)) | \hat{\boldsymbol{\theta}}^{(k)}] + E[l_3(\boldsymbol{\beta}_3, \lambda_0^D(t), \phi_\mu, \phi_\omega) | \hat{\boldsymbol{\theta}}^{(k)}] + E[l_4(\sigma_\mu^2, \sigma_\omega^2) | \hat{\boldsymbol{\theta}}^{(k)}] \\ &= Q_1(\boldsymbol{\beta}_1 | \hat{\boldsymbol{\theta}}^{(k)}) + Q_2(\boldsymbol{\beta}_2, \lambda_0^R(t) | \hat{\boldsymbol{\theta}}^{(k)}) + Q_3(\boldsymbol{\beta}_3, \lambda_0^D(t), \phi_\mu, \phi_\omega | \hat{\boldsymbol{\theta}}^{(k)}) \\ &\quad + Q_4(\sigma_\mu^2, \sigma_\omega^2 | \hat{\boldsymbol{\theta}}^{(k)}), \end{aligned} \quad (6)$$

where

$$\begin{aligned} Q_1(\boldsymbol{\beta}_1 | \hat{\boldsymbol{\theta}}^{(k)}) &= \sum_{i=1}^I \sum_{j=1}^J \left\{ E[y_{ij} \mu_i | \hat{\boldsymbol{\theta}}^{(k)}] + E[y_{ij} | \hat{\boldsymbol{\theta}}^{(k)}] \boldsymbol{\beta}_1^T \tilde{\mathbf{X}}_{ij} - E[\log(1 + q_{1ij}) | \hat{\boldsymbol{\theta}}^{(k)}] \right\}, \\ Q_2(\boldsymbol{\beta}_2, \lambda_0^R(t) | \hat{\boldsymbol{\theta}}^{(k)}) &= \sum_{i=1}^I \sum_{j=1}^J \left\{ R_{ij} \left(\sum_{m=1}^{N_{ij}^R} (\log(\lambda_0^R(t_{ijk}))) + \boldsymbol{\beta}_2^T \mathbf{X}_{ij} + E[\mu_i | \hat{\boldsymbol{\theta}}^{(k)}] + E[\omega_{ij} | \hat{\boldsymbol{\theta}}^{(k)}] \right) \right. \\ &\quad \left. - \Lambda_0^R(T_{ij}) \exp(\boldsymbol{\beta}_2^T \mathbf{X}_{ij} + \log E[y_{ij} \exp(\mu_i + \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}]) \right\}, \\ Q_3(\boldsymbol{\beta}_3, \lambda_0^D(t), \phi_\mu, \phi_\omega | \hat{\boldsymbol{\theta}}^{(k)}) &= \sum_{i=1}^I \sum_{j=1}^J \left\{ \Delta_{ij} (\log(\lambda_0^D(T_{ij})) + \boldsymbol{\beta}_3^T \mathbf{X}_{ij} + \phi_\mu E[\mu_i | \hat{\boldsymbol{\theta}}^{(k)}] + \phi_\omega E[\omega_{ij} | \hat{\boldsymbol{\theta}}^{(k)}]) \right. \\ &\quad \left. - \Lambda_0^D(T_{ij}) \exp(\boldsymbol{\beta}_3^T \mathbf{X}_{ij}) E[\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}] \right\}, \\ Q_4(\sigma_\mu^2, \sigma_\omega^2 | \hat{\boldsymbol{\theta}}^{(k)}) &= \sum_{i=1}^I \sum_{j=1}^J \left\{ -\frac{1}{2} \left(\log(2\pi) + \log \sigma_\omega^2 + \frac{E[\omega_{ij}^2 | \hat{\boldsymbol{\theta}}^{(k)}]}{\sigma_\omega^2} \right) \right. \\ &\quad \left. - \frac{1}{2} \left(\log(2\pi) + \log \sigma_\mu^2 + \frac{E[\mu_i^2 | \hat{\boldsymbol{\theta}}^{(k)}]}{\sigma_\mu^2} \right) \right\}. \end{aligned}$$

More details on the computation of the conditional expectation above and E-step are given in Appendix A.

In the M-step, we use a Newton-Raphson procedure to maximize $Q_1(\boldsymbol{\beta}_1 | \hat{\boldsymbol{\theta}}^{(k)})$, $Q_2(\boldsymbol{\beta}_2, \lambda_0^R(t) | \hat{\boldsymbol{\theta}}^{(k)})$, $Q_3(\boldsymbol{\beta}_3, \lambda_0^D(t), \phi_\mu, \phi_\omega | \hat{\boldsymbol{\theta}}^{(k)})$, $Q_4(\sigma_\mu^2, \sigma_\omega^2 | \hat{\boldsymbol{\theta}}^{(k)})$ to estimate $\boldsymbol{\theta}$. For the estimation of baseline intensity, we suggest the piecewise constant baseline intensity function, which retains majority model structure, because it provides more flexibility over the a priori choice of a baseline intensity distribution (eg, exponential and Weibull). Partition of M intervals with cutpoints $0 = c_0 < c_1 < \dots < c_M = \infty$ on the time duration can be based on the quantiles of event times, where $c_0 = 0$ or the smallest event time. The baseline intensity function is assumed to be constant within each of the M intervals, so that

$$\tilde{\lambda}_0(t) = \sum_{m=1}^M \lambda_m I(c_{m-1} < t \leq c_m). \quad (7)$$

The cumulative baseline intensity is

$$\tilde{\Lambda}_0(t) = \sum_{m=1}^M \lambda_m \max(0, \min(c_m - c_{m-1}, t - c_{m-1})). \quad (8)$$

The variance of the maximum likelihood estimation (MLE) $\hat{\theta} = (\hat{\lambda}_0^R(t), \hat{\lambda}_0^D(t), \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\phi}_\mu, \hat{\phi}_\omega, \hat{\sigma}_\omega^2, \hat{\sigma}_\mu^2)^T$ in the MCEM algorithm cannot be obtained directly from the algorithm. Louis's formula was used to obtain it.³³ The variance-covariance matrix can be estimated using the inverse of an observed information matrix. The observed information matrix $I(\hat{\theta})$ is given by

$$I(\hat{\theta}) = -E \left[\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \mid \hat{\theta} \right] - E \left[\frac{\partial l(\theta)}{\partial \theta} \frac{\partial l(\theta)}{\partial \theta^T} \mid \hat{\theta} \right] + E \left[\frac{\partial l(\theta)}{\partial \theta} \mid \hat{\theta} \right] + E \left[\frac{\partial l(\theta)}{\partial \theta^T} \mid \hat{\theta} \right]. \quad (9)$$

The last term becomes zero because of the MLE $\hat{\theta}$, and all of these terms are evaluated at the last iteration of the EM algorithm. The details of applying the Newton-Raphson algorithm are given in Appendix B.

4 | SIMULATION

In this section, we conduct simulation studies to evaluate our proposed joint frailty model. For simplicity, we only consider one covariate for X , which takes values of 0 or 1 each with probability 0.5, and also only the intercept term for the logistic regression model. Considering exponential distributions for both recurrent event times and death, the following models are shown below:

$$\begin{aligned} \text{logit} P(Y_{ij} = 1 \mid \mu_i) &= \beta_1 + \mu_i, \\ \lambda_{ij}^R(t \mid \mu_i, \omega_{ij}, Y_{ij} = 1) &= \lambda_0^R(t) \exp(\beta_2 X_{ij} + \mu_i + \omega_{ij}), \\ \lambda_{ij}^D(t \mid \mu_i, \omega_{ij}) &= \lambda_0^D(t) \exp(\beta_3 X_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij}). \end{aligned} \quad (10)$$

The censoring time is taken as $C_{ij} = 6 * \text{Unif}(0, 1)$, where $\text{Unif}(0, 1)$ is a random number generated from a uniform distribution in $[0, 1]$. Note that the number of pairs (ie, clusters) $I = 250$, and the cluster size $J = 2$; thus, the sample size is 500 for each scenario. We assume $\mu_i \sim N(0, \sigma_\mu^2)$, $\omega_{ij} \sim N(0, \sigma_\omega^2)$ and set $\sigma_\mu^2 = 1$, $\sigma_\omega^2 = 0.25$. We set the parameters as $\beta_2 = 1.2$, $\beta_3 = 1$, $\phi_\mu = 1$, $\phi_\omega = 0.5$, but invoke different values for β_1 and recurrent events baseline intensity λ_0^R . There are 2 setups: (1) $\beta_1 = 1$, $\lambda_0^R = 0.5$; 48% subjects are observed with at least one recurrent event; 22% subjects are susceptible but have not yet experienced any events during the follow-up period; (2) $\beta_1 = 0.5$, $\lambda_0^R = 0.25$; 33% subjects are observed with at least one recurrent event; 27% subjects are susceptible but have not yet experienced any events during the follow-up period. In both setups, $\lambda_0^D = 0.1$, and the censoring rate for death is 40%. The major difference between the 2 setups is the percentage of subjects who have at least one recurrent events (48% and 33%), we recognize these 2 percentages are reasonable to consider the zero inflation, and obviously if the percentage is too low, the convergence issue

could be challenging. A total of 500 Monte Carlo replicates are generated for results summary. A nonhomogeneous Poisson process is adopted to simulate recurrent event times, and the detailed procedures are provided in Appendix C. Of note, for comparison, we also consider the estimation approaches based on Gaussian quadrature and a Bayesian approach through SAS PROC MCMC to further evaluate our algorithm.

The piecewise constant approach is adopted to estimate the baseline intensity for recurrent events and baseline hazard for death, where 5 intervals are chosen. The proposed MCEM algorithm is implemented in R software, and the functions are available upon request. Regarding the Gaussian quadrature method, we invoke the adaptive Gaussian quadrature option with 5 quadrature points in SAS PROC NLMIXED. The Bayesian method in SAS PROC MCMC is implemented by the random walk Metropolis algorithm to obtain posterior samples with random walk step size 2.38. A Normal (0, 1000) prior is adopted for $\beta_1, \beta_2, \beta_3, \phi_\mu, \phi_\omega$. An Inverse gamma (0.01, 0.01) prior is adopted for σ_μ^2 and σ_ω^2 . A Gamma (0.01, 0.01) prior is adopted for 5 piecewise baseline intensities for both $\tilde{\lambda}_0^R(t)$ and $\tilde{\lambda}_0^D(t)$. The posterior sample size is 50 000. Also, due to the poor mixing and slow convergence, we thin a chain by keeping every 50th simulated draw from each sequence.

The resulting parameter estimates are shown in Tables 1 and 2. It can be seen that in both setups for the full model, MCEM estimation method yield satisfactory results and all parameter estimates have very small empirical biases. Monte Carlo expectation-maximization has better performance than the Gaussian quadrature and the Bayesian in SAS, especially on the estimation of parameter ϕ_ω , which has the smallest bias. Also, Monte Carlo standard error (MCSE) and MSE estimates from the EM algorithm are smaller than those from PROC NLMIXED and PROC MCMC. Comparing the parameter estimates between settings I and II, the empirical biases tend to increase when the percentage of recurrent events decreases from 48% to 33%. We also compute the mean of the standard error estimate for the full model based on our derivation, which are very close to the MCSE. Therefore, here, we only report MCSE in summary results for simulation.

For comparisons with misspecified models, if we do not consider the correlation μ in the matched pair (denoted by model I), the estimation of σ_ω^2 and ϕ_ω are overestimated. The correlation that was induced by matching is accounted for by recurrent events. This indicates that the hierarchical structure of the data needs to be taken into account to obtain accurate inference. If we ignore the zero inflation from the full model (denoted by model II), this will yield poor estimates in variance and coefficients of both frailties ω and μ . For example, in setting I, the estimate of σ_ω^2 is about 2 times larger than the true value, suggesting much more heterogeneity in recurrent events if not accounting for zero inflation. Therefore, our model is recommended in practice for valid inference, and insufficient models could lead to large biased estimates. In addition, per reviewers' comments, we also investigate the robustness of our approach when the assumption of $\mu_i \perp \omega_{ij}$ or the distribution assumption of normality is violated. Additional simulation studies (not provided here because of limited space) are conducted. The results show that the proposed joint modeling remains applicable even if there exist mild or moderate correlation between nested frailties μ_i and ω_{ij} (ie, $\rho = 0.5$); also, if the misspecification of frailty distribution does not affect the performance of

this joint modeling in particularly with negligible influence on the parameter estimates of the covariates effects on the risk of recurrent event and death.

5 | APPLICATION-STROKE STUDY

We apply the proposed model to a real data application on recurrent acute ischemic stroke. Our real data example is obtained from the MarketScan database between January 2011 and December 2014, including a matched cohort of patients with and without acute ischemic stroke at baseline, who are aged 45 to 54 years with surgical and medical admission for inpatient acute care hospitalization. Thus, $I = 1061$ matched pairs with cluster size $J = 2$ are extracted for analysis with the total sample size is $N = 2122$. Of note, the episodes of acute ischemic stroke are diagnosed by the *International Classification of Diseases, Ninth Revision, Clinical Modification* codes with 434.x and 436.x.³⁴ A recurrent stroke is defined as any recurrent stroke occurring more than 28 days after the incident stroke.³⁵ Figure 1 shows the sample data from random 6 matched pairs with and without stroke at baseline. Both cohorts are tracked for in-hospitalized mortality and pre-existing comorbidity conditions including diabetes mellitus, hypertension, and heart disease during the past 12 months, which could potentially influence outcomes.

Based on preliminary check of the data, the number of recurrent stroke ranges from 0 to 43, and 71.8% of the patients are not observed with recurrent strokes. The wide range of the number of recurrent stroke suggests a large variation across subjects; 139 patients (6.55%) died in the follow-up period, and 405 (20%) withdrew from the study. The preliminary analysis potentially shows that patients with baseline stroke have a higher risk of mortality (57.55%). Thus, censoring because of death is very likely to be informative. Joint frailty models of recurrent events in the presence of death can circumvent this problem by modeling recurrent events and death with shared frailties to capture their dependency. Baseline covariates included in the analysis are stroke status: Stroke (1=yes, 2=no) and Comorbidity (1=yes, 2=no). Comorbidity is defined as one if the patient has any one of diabetes mellitus, hypertension, and heart disease during the past 12 months. Our final model is

$$\begin{aligned} \text{logit}P(Y_{ij} = 1 \mid \mu_i) &= \beta_0 + \beta_1 \text{Stroke}_{ij} + \mu_i, \\ \lambda_{ij}^R(t \mid \mu_i, \omega_{ij}, Y = 1) &= \lambda_0^R(t) \exp(\beta_2 \text{Stroke}_{ij} + \beta_3 \text{Comorbidity}_{ij} + \mu_i + \omega_{ij}), \\ \lambda_{ij}^D(t \mid \mu_i, \omega_{ij}) &= \lambda_0^D(t) \exp(\beta_4 \text{Stroke}_{ij} + \beta_5 \text{Comorbidity}_{ij} + \phi_\mu \mu_i + \phi_\omega \omega_{ij}). \end{aligned} \quad (11)$$

We summarize the results in Table 3. We can see that both the baseline stroke and comorbidity covariates are significant in the proposed model. Patients with stroke at baseline had on odds ratio of 1.493 ($= \exp(0.401)$) to be "susceptible." Baseline stroke and comorbidity also have significant effects on both intensity of recurrent stroke among those "susceptible" and death hazard among all subjects. The intensity ratio of having recurrent stroke for patients with stroke at baseline is 2.942 ($= \exp(1.079)$) compared with those without stroke at baseline. Compared with patients who don't have comorbidity at baseline, the intensity ratio of having recurrent events is 1.104 ($= \exp(0.099)$) for patients who have comorbidity. Finally, the frailty variance estimate of μ is 0.669, suggesting the existence of

heterogeneity in the matched stroke and nonstroke pair. The frailty variance estimate for ω is 0.907, which may due to the wide range of the number of recurrent stroke events among patients. The estimates of ϕ_μ (4.001) and ϕ_ω (0.430) are significant greater than 0, which implies that the recurrent strokes events and death rates are positively associated. Thus, higher frailty (ω) will lead to a higher risk of recurrence and a higher risk of death. Also, higher frailty (μ) will result in a higher risk of recurrence, a higher risk of death, and a higher probability to develop a new stroke event.

We also fit 2 reduced joint frailty models for recurrent stroke, ie, without zero inflation and μ . The results from these 2 models are shown in Table 3. We notice that some parameter estimates are quite different from those in the full model, indicating the necessity to adjust for zero inflation and matched correlation, ie, without considering zero inflation or matched-pair correlation, the effect of comorbidity at baseline on having recurrent events is enlarged (the rate of having recurrent stroke increases from 1.104 to 1.359 ($= \exp(0.307)$) and 1.328 ($= \exp(0.284)$) compared with patients who don't have comorbidity). We also note that the frailty variance estimates are much larger than that in the full model, ie, not accounting for the zero inflation would lead to more heterogeneity for recurrent events in the reduced model (variance estimate of μ increased from 0.669 to 1.551). Ignoring the correlation induced by the matched design result in severely overlooking the importance of the correlation in recurrent events (variance estimate of ω increased from 0.907 to 2.923). Two information criteria for model assessment, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are also calculated, and both AIC and BIC statistics indicate that the proposed full model has the best fit with the smallest AIC = 11923 and BIC = 12003.

6 | DISCUSSION

In this article, we proposed a joint frailty model of recurrent events and a terminal event adjusted for zero inflation and the matched design. Compared with the latest research on similar topics, the major advantage of our model is the ability to quantify the dependency between recurrent events and the terminal event that are correlated in a hierarchical structure (eg, matched case-control study and meta-analysis). Extensive simulations have shown the robustness of our proposal with regards to the violation of several assumptions (ie, $\mu_i \perp \omega_{ij}$ normality). Of note is that we consider the multivariate normal distribution for frailties because of the following reasons: (1) The normal distribution is one of most popularly used distributions for frailties.^{8,9,18,36} (2) The (cluster-level) random effect is also included in the logistic regression for the zero inflation, and in such generalized linear regression setups, normal distribution is usually assumed for random effects.³⁷⁻⁴⁰ (3) Limited work exist for nested frailties with assumed multivariate normal distributions, where the computation for parameter estimation is more challenging because of nonclosed-form expressions or analytic solutions. Our work can fill up this gap by providing strategies for parameter estimation and statistical inference with the details of theoretical derivations. In addition, we have shown by simulation, and a real data application that using a reduced model (ignoring matched-pair correlation) instead of the proposed full model when there are 2 levels of clustering can lead to biased estimates, with an overestimation of the correlation in recurrent events. This calls into question the validity of traditional statistical techniques such as the shared frailty model

in studies with matched design. Another major advantage is the consideration of zero inflation of the recurrent event, which results in more accurate parameter estimation. In general, the price of omitting the feature of zero inflation of the events in the data from the models is biased estimates and overlooking the importance of certain cluster effects. Also, in medical research, investigators can evaluate the treatment effect by estimating the fraction of cured subjects, which is a substantially important question.

We also implemented our joint modeling approach in R software by considering an MCEM algorithm with a piecewise baseline intensity for the MLE of the model parameters. The R functions are available upon request, which can be flexibly modified for model extension and other research purposes. Most of the latest research on similar topics adopt the Gaussian quadrature in SAS PROC NLMIXED for estimation. Although SAS PROC NLMIXED is easy to implement for parameter estimation, there exist several limitations in practice. In our situation, we have 2 levels of nested random effects, which leads to dramatically computationally intense and convergence issues with about 3% failed optimization because of the fact that Hessian matrix is not positive definite. The number of failure optimizations may highly increase with a more sophisticated joint frailty model (eg, more levels of random effects). Because of the large volume of parameters for estimation and complexity of the posterior distribution, the adaptive MCMC algorithm in SAS also requires high computation burden because a large number of posterior samples are required. Sometimes, poor mixing or slow convergence issues emerge because the model parameters may be correlated with each other. Also, it is a little bit unstable because it yields relatively large standard errors on some parameter estimates, especially the parameters related with frailties.

There exist some limitations in our data. In the MarketScan research database, it is not possible to obtain the death information when patients are discharged from the hospital. The study cohort was tracked for in-hospital mortality only. Therefore, the death rate in our data underrepresents the true rate. The dependency between recurrent acute ischemic stroke and death may be underestimated. For the future work, our model can be extended in several directions. First, other functional forms of the frailties can be incorporated in the proposed joint model, such as gamma frailties, which have been applied by many researchers.^{6,23} With only a minor modification, the corresponding likelihood and the estimation algorithm are readily available. Second, time-varying covariates also can be incorporated. Third, we can consider more complex joint model settings. For example, we can jointly model longitudinal biomarkers with the current model.

ACKNOWLEDGEMENTS

The authors thank Dr Lan Kong and Dr W. Brian Reeves for helpful suggestions on statistical models and study designs. We also thank Dr Douglas Leslie, Dr Guodong Liu, and MS Djibril Ba for data preparation and permission to use the data from the MarketScan research database.

The project described was supported in part by research grant U01 DK082183 from the National Institute of Digestive, Diabetes and Kidney Diseases of the National Institutes of Health, US Department of Health and Human Services and by the National Center for Advancing Translational Sciences, grant KL2 TR000126 and TR002015. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding information

The National Institute of Digestive, Diabetes and Kidney Diseases (NIDDK), Grant/Award Number: U01 DK082183; The National Center for Advancing Translational Sciences, Grant/Award Number: KL2 TR002015 and UL1 TR002014; The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Appendix

APPENDIX A: E-STEP OF THE MCEM ALGORITHM

A.1 | Conditional expectation

The computation of the conditional expectation Equation 6 is not trivial. Denote the function of frailties as $h(\gamma_{ij})$ and $\gamma_{ij} = (y_{ij}, \mu_i, \omega_{ij})^T$. Thus, the expectation of complete data likelihood is $E[h(\gamma_{ij} | \hat{\theta}^{(k)})]$. The term $\hat{\theta}^{(k)}$ denotes all current parameters estimated in the M-step. The conditional density of γ_{ij} given the observed data and current estimate of the parameters is

$$f(\gamma_{ij} | \hat{\theta}^{(k)}) = \frac{L(\hat{\theta}^{(k)} | \gamma_{ij})f(\gamma_{ij})}{\int_{-\infty}^{\infty} L(\hat{\theta}^{(k)} | \gamma_{ij})f(\gamma_{ij})d\gamma_{ij}}. \quad (A1)$$

Thus, the conditional expectation for any function h of the random effects is

$$\begin{aligned} E[h(\gamma_{ij} | \hat{\theta}^{(k)})] &= \int_{-\infty}^{\infty} h(\gamma_{ij})f(\gamma_{ij} | \hat{\theta}^{(k)})d\gamma_{ij} \\ &= \frac{\int_{-\infty}^{\infty} h(\gamma_{ij})k(\gamma_{ij} | \hat{\theta}^{(k)})d\gamma_{ij}}{k(\gamma_{ij} | \hat{\theta}^{(k)})d\gamma_{ij}}, \end{aligned} \quad (A2)$$

where $k(\gamma_{ij} | \hat{\theta}^{(k)})$ is the kernel density after removing the parts that do not depend on γ_{ij} .

A.2 | E-step

Expectation terms that involve y_{ij} in Equation 6 in E-step are: $E[y_{ij} | \hat{\theta}^{(k)}]$, $E[y_{ij}\mu_i | \hat{\theta}^{(k)}]$, $E[y_{ij} \exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)}]$. $E[y_{ij}\mu_i | \hat{\theta}^{(k)}]$ is not involved in the M-step, and thus not considered. The expectation terms are evaluated with respect to $p(\mathbf{y}, \mu, \omega)$:

$$\begin{aligned} p(\mathbf{y}, \mu, \omega | \hat{\theta}^{(k)}) &\propto \prod_{i=1}^I \prod_{j=1}^J \left\{ [N_{ij}^R \exp(\mu_i + \omega_{ij})]^{R_{ij}} \exp(-y_{ij}\Lambda_0^R(T_{ij})q_{2ij}^{(k)}) \right. \\ &\quad \times \frac{\exp(y_{ij}(\beta_I^{T(k)} \tilde{\mathbf{X}}_{ij} + \mu_i))}{1 + \exp(\beta_I^{T(k)} \tilde{\mathbf{X}}_{ij} + \mu_i)} (\exp(\phi_\mu^{(k)}\mu_i + \phi_\omega^{(k)}\omega_{ij}))^{\Delta_{ij}} \exp(-\Lambda_0^{D(k)}(T_{ij})q_{3ij}^{(k)}) \\ &\quad \left. \times \frac{1}{\sqrt{2\pi\sigma_\mu^{2(k)}}} \exp\left(-\frac{\mu_i^2}{2\sigma_\mu^{2(k)}}\right) \frac{1}{\sqrt{2\pi\sigma_\omega^{2(k)}}} \exp\left(-\frac{\omega_{ij}^2}{2\sigma_\omega^{2(k)}}\right) \right\} \end{aligned} \quad (A3)$$

where, $q_{1ij}^{(k)} = \exp(\beta_1^{(k)} \tilde{X}_{ij} + \mu_i)$, $q_{2ij}^{(k)} = \exp(\beta_2^{(k)} X_{ij} + \mu_i + \omega_{ij})$,

$$q_{3ij}^{(k)} = \exp(\beta_3^{(k)} X_{ij} + \phi_\mu^{(k)} \mu_i + \phi_\omega^{(k)} \omega_{ij}), \text{ and } p_{ij}^{(k)} = \frac{\exp(\beta_1^{(k)} \tilde{X}_{ij} + \mu_i)}{1 + \exp(\beta_1^{(k)} \tilde{X}_{ij} + \mu_i)}.$$

We know that if $R_{ij} = 1$, then $y_{ij} = 1$. If $R_{ij} = 0$, then y_{ij} follows the binomial distribution below:

$$y_{ij} \mid \mu, \omega \sim \text{bin} \left\{ 1, \pi = \frac{p_{ij}^{(k)} \exp(-\Lambda_0^{R(k)}(T_{ij})q_{2ij}^{(k)})}{1 - p_{ij}^{(k)} + p_{ij}^{(k)} \exp(-\Lambda_0^{R(k)}(T_{ij})q_{2ij}^{(k)})} \right\}. \quad (\text{A4})$$

Therefore, conditional on the observed data and the current parameter estimation $\hat{\theta}^{(k)}$, we can evaluate the following expectations:

$$E[y_{ij} \mid \hat{\theta}^{(k)}] = \begin{cases} E \left[\frac{p_{ij}^{(k)} \exp(-\Lambda_0^{R(k)}(T_{ij})q_{2ij}^{(k)})}{1 - p_{ij}^{(k)} + p_{ij}^{(k)} \exp(-\Lambda_0^{R(k)}(T_{ij})q_{2ij}^{(k)})} \mid \hat{\theta}^{(k)} \right] & \text{if } R_{ij} = 0 \\ 1 & \text{otherwise,} \end{cases} \quad (\text{A5})$$

$$E[y_{ij} \exp(\mu_i + \omega_{ij}) \mid \hat{\theta}^{(k)}] = \begin{cases} E \left[\frac{\exp(\mu_i + \omega_{ij}) p_{ij}^{(k)} \exp(-\Lambda_0^{R(k)}(T_{ij})q_{2ij}^{(k)})}{1 - p_{ij}^{(k)} + p_{ij}^{(k)} \exp(-\Lambda_0^{R(k)}(T_{ij})q_{2ij}^{(k)})} \mid \hat{\theta}^{(k)} \right] & \text{if } R_{ij} = 0 \\ E[\exp(\mu_i + \omega_{ij}) \mid \hat{\theta}^{(k)}] & \text{otherwise,} \end{cases} \quad (\text{A6})$$

where the expectation are taken with respect to the $p(\mu, \omega \mid \hat{\theta}^{(k)})$.

$$\begin{aligned} p(\mu, \omega \mid \hat{\theta}^{(k)}) &\propto \prod_{i=1}^I \prod_{j=1}^J \left\{ 1 - p_{ij}^{(k)} + p_{ij}^{(k)} \right. \\ &\exp \left(- \int_0^\infty \Psi_{ij}(t) \lambda_0^{R(k)}(t) dt q_{2ij}^{(k)} \right) p_{ij}^{(k)} \prod_k (q_{2ij}^{(k)})^{\delta_{ijk}} \exp \left(- \int_0^\infty \Psi_{ij}(t) \lambda_0^{R(k)}(t) dt q_{2ij}^{(k)} \right) \\ &\times (q_{3ij}^{(k)})^{A_{ij}} \exp \left(- \int_0^\infty \Psi_{ij}(t) \lambda_0^{D(k)}(t) dt q_{3ij}^{(k)} \right) \frac{1}{\sqrt{2\pi\sigma_\mu^{2(k)}}} \\ &\left. \exp \left(- \frac{\mu_i^2}{2\sigma_\mu^{2(k)}} \right) \frac{1}{\sqrt{2\pi\sigma_\omega^{2(k)}}} \exp \left(- \frac{\omega_{ij}^2}{2\sigma_\omega^{2(k)}} \right) \right\} \end{aligned} \quad (\text{A7})$$

Since there is no closed-form expression of $p(\mu, \omega \mid \hat{\theta}^{(k)})$, Monte Carlo methods in combination with the Metropolis-Hasting algorithm are used to approximate the posterior distributions of μ_i 's and ω_{ij} 's. Given $\hat{\theta}^{(k)}$, N random samples are generated for $\mu_i^{(m)} (m = 1, \dots, M)$ and $\omega_{ij}^{(m)} (m = 1, \dots, M)$ to estimate the expectation of the sufficient statistics involving

frailties. Thus, $E[f(\mu_i) | \hat{\theta}^{(k)}] = \frac{1}{N} \sum_{n=1}^N f(\mu_i)^{(n)}$ and $E[f(\omega_{ij}) | \hat{\theta}^{(k)}] = \frac{1}{N} \sum_{n=1}^N f(\omega_{ij})^{(n)}$, where $f(\cdot)$ could be any smooth and monotone function.

APPENDIX B : M-STEP OF THE MCEM ALGORITHM

In the M-step, first we have closed-form estimators for $\tilde{\lambda}_0^R(t)$, $\tilde{\lambda}_0^D(t)\sigma_\mu^2$ and σ_ω^2 . For the estimation of $\tilde{\lambda}_0^R(t)$, first we introduce several notations. Assume there are M intervals with cutpoints $0 = c_0^R < c_1^R < \dots < c_M^R = \infty$, which can be quantiles of recurrent event times, where $c_0^R = 0$ or the smallest event time. Denote the indicator function $I_m^R(t) = I(c_{m-1}^R < t \leq c_m^R)$. For $m = 1, 2, \dots, M$,

$$\tilde{\lambda}_0^R(t) = \sum_{m=1}^M \lambda_m^R I(c_{m-1}^R < t \leq c_m^R) = \sum_{m=1}^M \lambda_m^R I_m^R(t). \quad (B1)$$

Therefore, we denote by $\lambda^R = (\lambda_1^R, \lambda_2^R, \dots, \lambda_m^R, \dots, \lambda_M^R)$ the parameter we aim to estimate. Denote $E_{ijk,m}^R = I^R(T_{ij} \geq c_{m-1}^R)(c_m^R \wedge T_{ij} - c_{m-1}^R)$ the total time individual ij is at risk in the m th recurrent events time interval $(c_{m-1}^R, c_m^R]$. $O_{ij,m}^R = \cap \varepsilon \sum_k^{N_{ij}^R} I_m^R(t_{ijk}) \delta_{ijk}$, the number of recurrent events for individual ij th in the m th subinterval. The maximum likelihood estimator has an explicit solution:

$$\hat{\lambda}_m^{R(k+1)} = \frac{\sum_m^M [\sum_{i=1}^I \sum_{j=1}^J O_{ij,m}^R]}{\sum_m^M [\sum_{i=1}^I \sum_{j=1}^J E_{ij,m}^R \exp(X_{ij}\beta_2^k) E(y_{ij} \exp(\mu_i + \omega_{ij}) | \hat{\theta}^{(k)})]}. \quad (B2)$$

To estimate $\tilde{\lambda}_0^D(t)$, similarly, we have the closed form:

$$\hat{\lambda}_m^{D(k+1)} = \frac{\sum_m^M [\sum_{i=1}^I \sum_{j=1}^J O_{ij,m}^D]}{\sum_m^M [\sum_{i=1}^I \sum_{j=1}^J E_{ij,m}^D \exp(X_{ij}\beta_3^k) E(\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\theta}^{(k)})]}. \quad (B3)$$

where $0 \leq c_0^D < c_1^D < \dots < c_M^D = \infty$ is the quantiles of death event times.

$I_m^D(t) = I(c_{m-1}^D < t \leq c_m^D)$. $E_{ijk,m}^D = I^D(T_{ij} \geq c_{m-1}^D)(c_m^D \wedge T_{ij} - c_{m-1}^D)$ is the total time individual ij is at risk in the m th death events time interval $(c_{m-1}^D, c_m^D]$.

$O_{ij,m}^D = \cap \varepsilon I_m^D(T_{ij}) \Delta_{ij}$, the number of death event for individual ij th in the m th interval.

The corresponding score functions and second partial derivatives for σ_μ^2 and σ_ω^2 are

$$S(\sigma_\mu^2) = \frac{\partial Q_4}{\partial \sigma_\mu^2} = \sum_{i=1}^I \left\{ -\frac{1}{2\sigma_\mu^2} + \frac{E(\mu_i^2 | \hat{\boldsymbol{\theta}}^{(k)})}{2\sigma_\mu^4} \right\},$$

$$S(\sigma_\omega^2) = \frac{\partial Q_4}{\partial \sigma_\omega^2} = \sum_{i=1}^I \sum_{j=1}^J \left\{ -\frac{1}{2\sigma_\omega^2} + \frac{E(\omega_{ij}^2 | \hat{\boldsymbol{\theta}}^{(k)})}{2\sigma_\omega^4} \right\}.$$

The estimators for σ_μ^2 and σ_ω^2 are

$$\hat{\sigma}_\mu^{2(k+1)} = \frac{1}{I} \sum_{i=1}^I E(\mu_i^2 | \hat{\boldsymbol{\theta}}^{(k)}), \quad (\text{B4})$$

$$\hat{\sigma}_\omega^{2(k+1)} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J E(\omega_{ij}^2 | \hat{\boldsymbol{\theta}}^{(k)}), \quad (\text{B5})$$

of note that IJ is the total number of subjects.

For the other parameters, which do not have the closed-form estimator, the Newton-Raphson algorithm is used to solve the expected log-likelihood iteratively. Given the k th estimate $\hat{\boldsymbol{\theta}}^{(k)}$, the $(k+1)$ th estimate is obtained by

$$\hat{\boldsymbol{\beta}}_2^{(k+1)} = \hat{\boldsymbol{\beta}}_2^{(k)} + I_{\boldsymbol{\beta}_2}^{-1}(\hat{\boldsymbol{\beta}}_2^{(k)}) S_{\boldsymbol{\beta}_2}(\hat{\boldsymbol{\beta}}_2^{(k)}). \quad (\text{B6})$$

Denote the parameters in $Q_3(\boldsymbol{\beta}_3, \tilde{\lambda}_0^D, \phi_\mu, \phi_\omega)$ as $\boldsymbol{\tau} = (\boldsymbol{\beta}_3, \phi_\mu, \phi_\omega)$. The gradient function $g(\boldsymbol{\tau}) = (\frac{\partial Q_3}{\partial \boldsymbol{\beta}_3}, \frac{\partial Q_3}{\partial \phi_\mu}, \frac{\partial Q_3}{\partial \phi_\omega})$, and the Hessian matrix can be obtained and denoted as H . Then, given current estimate $\hat{\boldsymbol{\theta}}^{(k)}$, the $(k+1)$ th estimate is obtained by

$$\hat{\boldsymbol{\tau}}^{(k+1)} = \hat{\boldsymbol{\tau}}^{(k)} - H_{\boldsymbol{\tau}}^{-1}(\hat{\boldsymbol{\tau}}^{(k)}) g_{\boldsymbol{\tau}}(\hat{\boldsymbol{\tau}}^{(k)}). \quad (\text{B7})$$

Similarly, the MLE of $\boldsymbol{\beta}_1$ can be updated by

$$\hat{\boldsymbol{\beta}}_1^{(k+1)} = \hat{\boldsymbol{\beta}}_1^{(k)} + I_{\boldsymbol{\beta}_1}^{-1}(\hat{\boldsymbol{\beta}}_1^{(k)}) S_{\boldsymbol{\beta}_1}(\hat{\boldsymbol{\beta}}_1^{(k)}), \quad (\text{B8})$$

where $S(\cdot)$ and $I(\cdot)$ are the score function and information matrix, respectively. To simplify the formula, we assume here that X_{ij} only contains one covariate, which is denoted as X_{ij} . Then the corresponding score functions and information matrices of the parameters are

$$\begin{aligned}
 S(\beta_1) &= \frac{\partial Q_1}{\partial \beta_1} = \sum_{i=1}^I \sum_{j=1}^J \left\{ E(y_{ij} | \hat{\boldsymbol{\theta}}^{(k)}) X_{ij} - \exp(\beta_1 X_{ij}) X_{ij} E \left[\frac{\exp(\mu_i)}{1 + q_{1ij}} | \hat{\boldsymbol{\theta}}^{(k)} \right] \right\}, \\
 I(\beta_1) &= - \frac{\partial^2 Q_1}{\partial \beta_1^2} = \sum_{i=1}^I \sum_{j=1}^J \left\{ \exp(\beta_1 X_{ij}) X_{ij}^2 E \left[\frac{\exp(\mu_i)}{(1 + q_{1ij})^2} | \hat{\boldsymbol{\theta}}^{(k)} \right] \right\}, \\
 S(\beta_2) &= \frac{\partial Q_2}{\partial \beta_2} = \sum_{i=1}^I \sum_{j=1}^J \left\{ \sum_k^{N_{ij}^R} \delta_{ijk} X_{ij} - \tilde{\Lambda}_0^R(T_{ij}) X_{ij} \exp(X_{ij} \beta_2) E(y_{ij} \exp(\mu_i + \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 I(\beta_2) &= - \frac{\partial^2 Q_2}{\partial \beta_2^2} = \sum_{i=1}^I \sum_{j=1}^J \left\{ \tilde{\Lambda}_0^R(T_{ij}) X_{ij}^2 \exp(X_{ij} \beta_2) E(y_{ij} \exp(\mu_i + \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 S(\beta_3) &= \frac{\partial Q_3}{\partial \beta_3} = \sum_{i=1}^I \sum_{j=1}^J \left\{ \Delta_{ij} X_{ij} - \tilde{\Lambda}_0^D(T_{ij}) X_{ij} \exp(X_{ij} \beta_3) E(\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 S(\phi_\mu) &= \frac{\partial Q_3}{\partial \phi_\mu} = \sum_{i=1}^I \sum_{j=1}^J \left\{ \Delta_{ij} E(\mu_i | \hat{\boldsymbol{\theta}}^{(k)}) - \tilde{\Lambda}_0^D(T_{ij}) \exp(X_{ij} \beta_3) E(\mu_i \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 S(\phi_\omega) &= \frac{\partial Q_3}{\partial \phi_\omega} = \sum_{i=1}^I \sum_{j=1}^J \left\{ \Delta_{ij} E(\omega_{ij} | \hat{\boldsymbol{\theta}}^{(k)}) - \tilde{\Lambda}_0^D(T_{ij}) \exp(X_{ij} \beta_3) E(\omega_{ij} \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 \frac{\partial^2 Q_3}{\partial \beta_3^2} &= \sum_{i=1}^I \sum_{j=1}^J \left\{ -\tilde{\Lambda}_0^D(T_{ij}) X_{ij}^2 \exp(X_{ij} \beta_3) E(\exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 \frac{\partial^2 Q_3}{\partial \phi_\mu^2} &= \sum_{i=1}^I \sum_{j=1}^J \left\{ -\tilde{\Lambda}_0^D(T_{ij}) \exp(X_{ij} \beta_3) E(\exp(\mu_i^2 \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 \frac{\partial^2 Q_3}{\partial \phi_\omega^2} &= \sum_{i=1}^I \sum_{j=1}^J \left\{ -\tilde{\Lambda}_0^D(T_{ij}) \exp(X_{ij} \beta_3) E(\exp(\omega_{ij}^2 \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 \frac{\partial^2 Q_3}{\partial \beta_3 \phi_\mu} &= \sum_{i=1}^I \sum_{j=1}^J \left\{ -\tilde{\Lambda}_0^D(T_{ij}) X_{ij} \exp(X_{ij} \beta_3) E(\mu_i \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 \frac{\partial^2 Q_3}{\partial \beta_3 \phi_\omega} &= \sum_{i=1}^I \sum_{j=1}^J \left\{ -\tilde{\Lambda}_0^D(T_{ij}) X_{ij} \exp(X_{ij} \beta_3) E(\omega_{ij} \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}, \\
 \frac{\partial^2 Q_3}{\partial \phi_\mu \phi_\omega} &= \sum_{i=1}^I \sum_{j=1}^J \left\{ -\tilde{\Lambda}_0^D(T_{ij}) \exp(X_{ij} \beta_3) E(\mu_i \omega_{ij} \exp(\phi_\mu \mu_i + \phi_\omega \omega_{ij}) | \hat{\boldsymbol{\theta}}^{(k)}) \right\}.
 \end{aligned}$$

More components of the information matrix are given below:

$$\begin{aligned}
 \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_\mu^2} &= - \frac{1}{2} \left[- \frac{I}{\sigma_\mu^4} + \frac{2 \sum_{i=1}^I E(\mu_i^2 | \hat{\boldsymbol{\theta}}^{(k)})}{\sigma_\mu^6} \right], \\
 \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_\omega^2} &= - \frac{1}{2} \left[- \frac{IJ}{\sigma_\omega^4} + \frac{2 \sum_{i=1}^I \sum_{j=1}^J E(\omega_{ij}^2 | \hat{\boldsymbol{\theta}}^{(k)})}{\sigma_\omega^6} \right].
 \end{aligned}$$

All other off-diagonal terms are zero. When $\mathbf{X}_{ij}$ is a covariate vector, corresponding score functions and information matrices of the parameters can be easily adapted to matrix versions.

APPENDIX C : SIMULATION ON RECURRENT EVENT TIMES

In this section, we give a brief description about the simulation procedure. Based on the definition of a Poisson process, for each $0 \leq s < t$, $N(t) - N(s)$ has a Poisson distribution with

mean $m(t) - m(s) = \int_s^t \lambda(x) dx$. Therefore, we need to simulate recurrent event times from a Poisson process with any intensity function $\lambda(t)$. The most common parametric choices for time-to-event simulation are exponential, Weibull, and Gompertz distributions. As noted by Bender et al,⁴¹ among the commonly used distributions for survival times, only these 3 distributions share the assumption of proportional hazards, and the survival time, T , can be generated by $T = \Lambda_0^{-1}[-\log(u) \exp(-\beta^T \mathbf{x})]$, where $u \sim \text{Unif}[0, 1]$.

There are several available methods for recurrent event times simulation, and here we adopt the inversion method, which is often used. Referring to the inversion method by Pénichoux et al,⁴² for a given subject, let T_j denote the time elapsed from the origin to the j th event, $T_0 = 0$, and let $W_j = T_j - T_{j-1}$ denote the j th waiting time between 2 consecutive events. The W_j are referred to as gap times. The number of events observed for a nonhomogeneous Poisson process with intensity $\lambda(\cdot)$ between 2 times, s and t follows a Poisson distribution with parameter $\int_s^t \lambda(v) dv$, which leads to $F_j(w) = 1 - \exp(-\int_{T_{j-1}}^{T_j-1+w} \lambda(v) dv)$. $W_j = F_j^{-1}(u_j)$ and $T_j = T_{j-1} + W_j$. For simplicity, we adopt an exponential distribution with constant baseline constant $\lambda_0(t) = \lambda$, the recurrent event process can be considered equivalently as an homogeneous Poisson process or a sequence of exponentially distributed gaps.⁴² The survival times are generated by $T = -\frac{\log(u)}{\lambda \exp(-\beta^T \mathbf{x})}$,

$$T_j = T_{j-1} + W_j = -\frac{\log(u_j)}{\lambda \exp(-\beta^T \mathbf{x})} + T_{j-1}.$$

REFERENCES

1. Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001;88(4):907–919.
2. Lancaster T, Intrator O. Panel data with survival: hospitalization of hiv-positive patients. *J Am Stat Assoc*. 1998;93(441):46–53.
3. Chen BE, Cook RJ. Tests for multivariate recurrent events in the presence of a terminal event. *Biostatistics*. 2004;5(1):129–143. [PubMed: 14744832]
4. Cook RJ, Lawless JF. Marginal analysis of recurrent events and a terminating event. *Stat Med*. 1997;16(8):911–924. [PubMed: 9160488]
5. Ghosh D, Lin DY. Marginal regression models for recurrent and terminal events. *Stat Sin*. 2002;12:663–688.
6. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics*. 2004;60(3):747–756. [PubMed: 15339298]
7. Li R, Cheng Y. Flexible association modelling and prediction with semi-competing risks data. *Can J Stat*. 2016;44(3):361–374.
8. Huang X, Wolfe RA. A frailty model for informative censoring. *Biometrics*. 2002;58(3):510–520. [PubMed: 12229985]
9. Zeng D, Lin DY. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics*. 2009;65(3):746–752. [PubMed: 18945267]
10. Mozaffarian D, Benjamin EJ, Go AS, et al. Executive summary: heart disease and stroke statistics —2016 update: a report from the American Heart Association. *Circ*. 2016;133(4):447.
11. Azarpazhooh MR, Nicol MB, Donnan GA, et al. Patterns of stroke recurrence according to subtype of first stroke event: the North East Melbourne Stroke Incidence Study (NEMESIS). *Int J Stroke*. 2008;3(3):158–164. [PubMed: 18705891]
12. Feng W, Hendry RM, Adams RJ. Risk of recurrent stroke, myocardial infarction, or death in hospitalized stroke patients. *Neurology*. 2010;74(7):588–593. [PubMed: 20157161]

13. TruvenHealth. Marketscan research databases. 1990 <https://marketscan.truvenhealth.com/marketscanportal/>.
14. DeSantis SM, Lazaridis C, Ji S, Spinale FG. Analyzing propensity matched zero-inflated count outcomes in observational studies. *J Appl Stat*. 2014;41(1):127–141.
15. Liu L, Huang X, Yaroshinsky A, Cormier JN. Joint frailty models for zero-inflated recurrent events in the presence of a terminal event. *Biometrics*. 2016;72(1):204–214. [PubMed: 26295794]
16. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*. 1982;38(4):1041–1046. [PubMed: 7168793]
17. Kuk AYC, Chen CH. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*. 1992;79(3):531–541.
18. Peng Y, Taylor JMG. Mixture cure model with random effects for the analysis of a multi-center tonsil cancer study. *Stat Med*. 2011;30(3):211–223. [PubMed: 21213339]
19. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000;56(1):227–236. [PubMed: 10783800]
20. Rondeau V, Schaffner E, Corbière F, Gonzalez JR, Mathoulin-Pélissier S. Cure frailty models for survival data: application to recurrences for breast cancer and to hospital readmissions for colorectal cancer. *Stat Methods Med Res*. 2013;22(3):243–260. [PubMed: 21632696]
21. Sastry N A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *J Am Stat Assoc*. 1997;92(438):426–435. [PubMed: 12155403]
22. Go AS, Parikh CR, Ikizler TA, et al. The assessment, serial evaluation, and subsequent sequelae of acute kidney injury (ASSESS-AKI) study: design and methods. *BMC Nephrol*. 2010;11(1):1. [PubMed: 20078870]
23. Rondeau V, Filleul L, Joly P. Nested frailty models using maximum penalized likelihood estimation. *Stat Med*. 2006;25(23):4036–4052. [PubMed: 16463308]
24. Yau KKW. Multilevel models for survival analysis with random effects. *Biometrics*. 2001;57(1):96–102. [PubMed: 11252624]
25. Liu L, Huang X. The use of gaussian quadrature for estimation in frailty proportional hazards models. *Stat Med*. 2008;27(14):2665–2683. [PubMed: 17910008]
26. Ibrahim JG, Chen MH, Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Stat Sin*. 2004;14:863–883.
27. Ouyang B, Sinha D, Slate EH, Van Bakel AB. Bayesian analysis of recurrent event with dependent termination: an application to a heart transplant study. *Stat Med*. 2013;32(15):2629–2642. [PubMed: 23280968]
28. Wen S, Huang X, Frankowski RF, Cormier JN, Pisters P. A bayesian multivariate joint frailty model for disease recurrences and survival. *Stat Med*. 2016;35(26):4794–4812. [PubMed: 27383540]
29. Kim S, Zeng D, Chambless L, Li Y. Joint models of longitudinal data and recurrent events with informative terminal event. *Stat Biosci*. 2012;4(2):262–281. [PubMed: 23227131]
30. Klein JP. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*. 1992;48:795–806. [PubMed: 1420842]
31. Vaida F, Xu R. Proportional hazards model with random effects. *Stat Med*. 2000;19(24):3309–3324. [PubMed: 11122497]
32. Król A, Ferrer L, Pignon JP, et al. Joint model for left-censored longitudinal data, recurrent events and terminal event: predictive abilities of tumor burden for cancer evolution with application to the FFCD 2000-05 trial. *Biometrics*. 2016;72(3):907–916. [PubMed: 26890381]
33. Louis TA. Finding the observed information matrix when using the EM algorithm. *J R Stat Soc Ser B Methodol*. 1982;44(2):226–233.
34. Goldstein LB. Accuracy of ICD-9-CM coding for the identification of patients with acute ischemic stroke effect of modifier codes. *Stroke*. 1998;29(8):1602–1604. [PubMed: 9707200]
35. Coull AJ, Rothwell PM. Underestimation of the early risk of recurrent stroke evidence of the need for a standard definition. *Stroke*. 2004;35(8):1925–1929. [PubMed: 15192248]

36. Zeng D, Ibrahim JG, Chen MH, Hu K, Jia C. Multivariate recurrent events in the presence of multivariate informative censoring with applications to bleeding and transfusion events in myelodysplastic syndrome. *J Biopharm Stat.* 2014;24(2):429–442. [PubMed: 24605978]
37. Hall DB, Wang L. Two-component mixtures of generalized linear mixed effects models for cluster correlated data. *Stat Model.* 2005;5(1):21–37.
38. Inan G, Ilk O. A marginalized multilevel model for bivariate longitudinal binary data. *Stat Pap.* 2016 10.1007/S00362-016-0840-1.
39. Neuhaus JM, McCulloch CE. Estimation of covariate effects in generalized linear mixed models with informative cluster sizes. *Biometrika.* 2011;98(1):147–162. [PubMed: 23049125]
40. Parzen M, Ghosh S, Lipsitz S, et al. A generalized linear mixed model for longitudinal binary data with a marginal logit link function. *Ann Appl Stat.* 2011;5(1):449. [PubMed: 21532998]
41. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Stat Med.* 2005;24(11):1713–1723. [PubMed: 15724232]
42. Pénichoux J, Moreau T, Latouche A. Simulating recurrent events that mimic actual data: a review of the literature with emphasis on event-dependence. 2014. HAL Id: hal-01133497.

Figure 1.
Sample data on recurrent acute ischemic stroke

TABLE 1

Simulation results of setting I: $\beta_1 = 1$, $\lambda_0^R = 0.5$, $\lambda_0^D = 0.1:48\%$ subjects are observed with at least one recurrent event^a

Setting I		Model I				Model II			Full Model		
<i>N</i> = 500	Parameters	True	AB	MCSE	MSE	AB	MCSE	MSE	AB	MCSE	MSE
MCEM	β_1	1	0.203	0.140	0.061				0.000	0.195	0.038
	β_2	1.2	0.020	0.088	0.008	0.049	0.096	0.012	0.002	0.090	0.008
	β_3	1	0.014	0.147	0.022	0.013	0.164	0.027	0.020	0.145	0.021
	ϕ_μ	1				0.216	0.109	0.059	0.033	0.129	0.018
	ϕ_ω	0.5	0.538	0.141	0.309	0.372	0.175	0.169	0.002	0.368	0.135
	σ_μ^2	1				0.757	0.278	0.650	0.025	0.156	0.025
	σ_ω^2	0.25	0.926	0.204	0.900	0.360	0.205	0.171	0.001	0.044	0.002
GQ	β_1	1	0.192	0.205	0.079				0.023	0.223	0.050
	β_2	1.2	0.013	0.124	0.016	0.010	0.132	0.018	0.004	0.116	0.013
	β_3	1	0.056	0.180	0.036	0.011	0.163	0.026	0.041	0.169	0.030
	ϕ_μ	1				0.228	0.119	0.066	0.033	0.154	0.025
	ϕ_ω	0.5	0.541	0.157	0.317	0.293	0.196	0.124	0.040	0.469	0.221
	σ_μ^2	1				0.762	0.296	0.668	0.005	0.187	0.035
	σ_ω^2	0.25	0.986	0.241	1.030	0.466	0.184	0.252	0.002	0.093	0.009
MCMC	β_1	1	0.178	0.218	0.079				0.031	0.230	0.054
	β_2	1.2	0.008	0.154	0.024	0.010	0.152	0.024	0.001	0.118	0.014
	β_3	1	0.076	0.204	0.057	0.022	0.194	0.038	0.062	0.177	0.035
	ϕ_μ	1				0.207	0.134	0.061	0.035	0.312	0.099
	ϕ_ω	0.5	0.565	0.175	0.350	0.262	0.296	0.156	0.097	0.705	0.507
	σ_μ^2	1				0.764	0.305	0.677	0.022	0.196	0.039
	σ_ω^2	0.25	1.05	0.238	1.159	0.462	0.205	0.255	0.005	0.108	0.012

Abbreviations: AB, absolute value of the difference between Monte Carlo mean of the parameter estimates (bases on 500 replicates) and the true value; GQ, Gaussian quadrature approach; MCEM, Monte Carlo expectation-maximization algorithm; MCMC, Markov chain Monte Carlo method; MCSE, Monte Carlo standard error; MSE, mean square error.

^a Another 22% subjects are susceptible but have not yet experienced any events during the follow-up period. The censoring rate for death is 40%. Full model denotes the proposed joint model of zero-inflated recurrent events and death event; Model I denotes the model without considering the matched-pair correlation, where matched pair random effect μ is removed; Model II denotes the model without considering zero inflation for recurrent events.

TABLE 2

Simulation results of setting II: $\beta_1 = 0.5$, $\lambda_0^R = 0.25$, $\lambda_0^D = 0.1:33\%$ subjects are observed with at least one recurrent event^a

Setup II		Model I				Model II			Full Model		
N=500	Parameters	True	AB	MCSE	MSE	AB	MCSE	MSE	AB	MCSE	MSE
MCEM	β_1	0.5	0.102	0.128	0.027				0.016	0.196	0.038
	β_2	1.2	0.040	0.116	0.015	0.038	0.143	0.022	0.009	0.137	0.019
	β_3	1	0.042	0.163	0.028	0.026	0.159	0.026	0.012	0.143	0.021
	ϕ_μ	1				0.282	0.111	0.092	0.012	0.133	0.018
	ϕ_ω	0.5	0.643	0.182	0.446	0.328	0.198	0.146	0.021	0.344	0.119
	σ_μ^2	1				0.928	0.355	0.987	0.013	0.185	0.034
	σ_ω^2	0.25	0.935	0.289	0.958	0.300	0.227	0.142	0.002	0.032	0.001
GQ	β_1	0.5	0.080	0.237	0.062				0.032	0.283	0.081
	β_2	1.2	0.014	0.152	0.023	0.002	0.156	0.024	0.012	0.144	0.021
	β_3	1	0.084	0.195	0.045	0.005	0.167	0.028	0.036	0.175	0.032
	ϕ_μ	1				0.258	0.129	0.083	0.058	0.188	0.039
	ϕ_ω	0.5	0.597	0.187	0.391	0.319	0.254	0.166	0.090	0.499	0.257
	σ_μ^2	1				0.920	0.399	1.005	0.002	0.251	0.063
	σ_ω^2	0.25	1.114	0.305	1.334	0.529	0.270	0.363	0.033	0.137	0.020
MCMC	β_1	0.5	0.070	0.292	0.090				0.028	0.304	0.093
	β_2	1.2	0.003	0.177	0.031	0.007	0.181	0.032	0.002	0.148	0.022
	β_3	1	0.114	0.226	0.064	0.025	0.190	0.037	0.060	0.184	0.038
	ϕ_μ	1				0.210	1.281	1.685	0.053	0.748	0.562
	ϕ_ω	0.5	0.652	0.234	0.479	0.305	0.870	0.849	0.098	0.968	0.947
	σ_μ^2	1				0.961	0.453	1.129	0.031	0.275	0.076
	σ_ω^2	0.25	1.169	0.318	1.468	0.506	0.352	0.380	0.004	0.177	0.031

Abbreviations: AB, absolute value of the difference between Monte Carlo mean of the parameter estimates (bases on 500 replicates) and the true value; GQ, Gaussian quadrature approach; MCEM, Monte Carlo expectation-maximization algorithm; MCMC, Markov chain Monte Carlo method; MCSE, Monte Carlo standard error; MSE, mean square error.

^a Another 27% subjects are susceptible but have not yet experienced any events during the follow-up period. The censoring rate for death is 40%. Full model denotes the proposed joint model of zero-inflated recurrent events and death event. Model I denotes the model without considering the matched-pair correlation, where matched pair random effect μ is removed. Model II denotes the model without considering zero inflation for recurrent events.

TABLE 3

Analysis of recurrent ischemic stroke data^a

Covariate	Model I			Model II			Full Model		
	Estimate	SE	P Value	Estimate	SE	P Value	Estimate	SE	P Value
Incidence									
Intercept	-0.725	0.218	<.001				-0.300	0.114	.009
Stroke	2.055	0.244	<.001				0.401	0.134	.003
Recurrent stroke									
Stroke	1.152	0.242	<.001	1.045	0.038	<.001	1.079	0.024	<.001
Comorbidity	0.284	0.127	.025	0.307	0.119	<.001	0.099	0.096	.303
Death									
Stroke	1.812	0.356	<.001	1.315	0.147	<.001	0.703	0.143	<.001
Comorbidity	1.148	0.322	<.001	1.025	0.230	<.001	0.615	0.169	<.001
ϕ_{μ}				3.041	0.234	.042	4.001	0.263	<.001
ϕ_{ω}	2.634	0.216	<.001	0.275	0.191	.009	0.430	0.210	.041
σ_{μ}^2				1.551	0.091	<.001	0.669	0.074	<.001
σ_{ω}^2	2.923	0.349	<.001	2.502	0.374	<.001	0.907	0.230	<.001
AIC		14630			14966			11923	
BIC		14709			15036			12003	

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; SE, standard error of the parameter estimate.

^aFull model denotes the proposed joint model of zero-inflated recurrent events and death event. Model I denotes the model without considering the matched-pair correlation, where matched pair random effect μ is removed. Model II denotes the model without considering zero inflation for recurrent events.