HHS Public Access

Author manuscript

Stat Methods Med Res. Author manuscript; available in PMC 2022 February 01.

Published in final edited form as:

Stat Methods Med Res. 2021 February ; 30(2): 508–522. doi:10.1177/0962280220962522.

Joint Analysis of Recurrence and Termination: A Bayesian Latent Class Approach

Zhixing Xu[1], **Debajyoti Sinha**[1], **Jonathan R. Bradley**[1,*]

[1]Department of Statistics, Florida State University, Tallahassee, Florida 32306, U. S. A.

Summary:

Like many other clinical and economic studies, each subject of our motivating transplant study is at risk of recurrent events of Non-Fatal Tissue Rejections (NFTR) as well as the terminating event of death due to total graft rejection. For such studies, our model and associated Bayesian analysis aim for some practical advantages over competing methods. Our semiparametric latent-class based joint model has coherent interpretation of the covariate (including race and gender) effects on all functions and model quantities that are relevant for understanding the effects of covariates on future event trajectories. Our fully Bayesian method for estimation and prediction using a complete specification of the prior process of the baseline functions. We also derive a practical and theoretically justifiable partial likelihood based semiparametric Bayesian approach to deal with analysis when there is a lack of prior information about the baseline functions. Our model and method can accommodate fixed as well as time-varying covariates. Our Markov Chain Monte Carlo tools for both Bayesian methods are implementable via publicly available software. Our Bayesian analysis of transplant study and simulation study demonstrate the practical advantages and improved performance of our approach.

Keywords

Bayesian analysis; Frailty; Joint Model; Intensity and rate; Recurrent events

1. Introduction

Data on times to recurrent events until termination are common in various studies in cancer, chronic diseases, organ-transplant, repairable systems and economics. For example, in our motivating study for evaluating the covariate effects on each patient after receiving transplant, two types of responses for each transplant patient are: (1) the recurrent events of Non-fatal Tissue Rejections (NFTR) that were treated effectively by drug therapy, and (2) the terminating event of Graft-versus-Host Disease event (GvHD event) resulting in either total graft rejection or death. Although methodologies for recurrent events data have a long history in the literature (Cook and Lawless, 2007), the topic of recurrent events data with informative termination is a relatively new research field.

* jrbradley@fsu.edu.

Either using the naive assumption of non-informative termination (as defined in Cook and Lawless, 2007) or making inference about every recurrence while treating the termination and the remaining events as nuisances (Hougaard, 2000) often leads to seriously biased and even misleading inference (Miloslavsky et al., 2004). Other methods (see review by Miloslavsky et al., 2004) using an extension of the Coarsening-At-Random (CAR) assumption of Heitjan and Rubin (1991) preclude any inference on the termination event. Also, the CAR assumption is not verifiable from observed data and often lacks any practical meaning especially for transplant and studies with terminating event being death (Huang and Wang, 2004; Sinha et al., 2008). All of these approaches fail to coherently explain covariate effects on termination, evaluate the link between the recurrences and the risk of termination, and make prediction about future event processes. In many studies including our transplant study, evaluation of covariate effects and the predictions of future trajectories of both recurrent and terminating events are important analysis and prediction goals. For example, given some previous evidence of racial disparity on recurrent NFTR after rejection (example, Higgins and Fishman (2006)), one of the major goals of the analysis of our transplant study is a comprehensive and coherent evaluation of the race effect on joint trajectories of both NFTR and fatal GvHD events after transplant. To present a coherent and comeprehensive interpretation of the overall effect of race on both types of events, the main challenge of any useful joint model is to present clinically interpretable effects of race on following functions related to the trajectories of events after transplant: (1) the intensity function of recurrence and the hazard of termination, both conditional on recurrence history; and (2) the mean number and the rate of events, both unconditional on history. The former functions represent the dynamic effects of race and recurrence history on future events. Second set of functions express the non-dynamic (marginal) effects of race on future events. For the sake of physical interpretation, it is further desirable that a joint model should ensure similar signs and magnitudes of the race effect on all of these functions.

Since Lancaster and Intrator (1998), the joint modeling literature of such data has been dominated by models that use a patient-specific "frailty" random-effect shared by both recurrence and termination within a patient (e.g., Liu et al., 2004; Ye et al., 2007; Sinha et al., 2008; Zeng and Lin, 2009; Huang et al., 2010; Kalbfleisch et al., 2013; Xu et al., 2017). Except few (Xu et al., 2017; Paulon et al., 2018), these shared-frailty models usually require an assumption of parametric frailty distribution that can not be easily assessed from the observed data. These shared-frailty models usually lack simultaneous physical interpretations of covariate effects on all functions of interest listed above. For example, to obtain a reasonable expression of effect of race on mean and rate of recurrences over time, most of the existing shared-frailty models need the recurrent events given the frailty to be a Poisson process, an assumption considered too restrictive in practice. Whereas, other shared-frailty models with clear interpretations of covariate effects on the mean recurrence and rate (e.g. Xu et al., 2017) lack practical interpretation of dynamic effects of covariates on risks of new recurrence and termination at time t given current history of events. There are some recent interesting works using copula structure (Shih and Louis, 1995) for bivariate and even multivariate frailty random effects to model association among several types of events, recurrences and termination while preserving some desired marginal distribution of each frailty effect (e.g., Lee and Cook (2019) and Cook et al. (2010)). The goal is to use the desired marginal density of a particular frailty effect, say, a marginal Gamma frailty effect on the recurrent events, to obtain a computationally tractable likelihood. In spite of being more flexible than shared frailty models, these approaches also share some of the same difficulties in expressing simultaneous physical interpretation of covariate effects on all functions of interest. Also, models using multivariate frailty with copula are not amenable to Bayesian partial likelihood-based approach.

Recently popular Joint Latent Class Models (JLCM) for joint analysis of survival and longitudinal responses outlined in Proust-Lima et al. (2014), Barrett et al. (2015) and others avoid several several pitfalls of shared random effects models, such as increasing dimension with sample size and lack of estimation of individualized survival risk given past longitudinal outcome trajectory. Our goal is also to develop latent class models for our current problem to replicate the successes and advantages of these methods for joint analysis of longitudinal and survival outcomes.

In Section 2, we present a novel JLCM for recurrent events and termination with several practical advantages including a prediction of future profiles of recurrent and terminal events given covariates. We demonstrate the methodological advantages of the JLCM compared to existing models by showing that the JLCM has a coherent interpretation of the dynamic effects of the covariates on the risks of future events given the history, as well as the covariate effects on the rate and mean number of recurrences, unconditional on history. In Section 3, we present two semiparametric Bayesian methods of analysis using JLCM. These methods include the directions for choosing the priors, and demonstration of the ease of implementing associated Markov Chain Monte Carlo (MCMC) tools. The fully Bayesian method of Section 3 requires prior opinions on baseline functions; however, it is capable of predicting the future event trajectories. A partial likelihood based Bayesian method of Section 3 is useful when there is no available prior opinions about these unknown baseline functions of both events. Our MCMC based practical Bayesian methods are easy to implement via publicly available software such as OpenBUGS and these programs are made available by the corresponding author. In Section 4, our simulation studies show the performances of the JLCM under different priors compared to existing Bayesian methods. In Section 5, we provide an analysis of transplant data to illustrates the clinical interpretation and advantages of our models and associated methods in practice. Section 6 presents the concluding discussion including the extension of our methods and results to studies with time-varying covariates.

2. Joint Latent Class Model

Our JLCM assumes that future event trajectories of patients $i = 1, \cdots, n$ depend on the latent class M_i of one of $K + 1$ latent homogeneous sub-populations $G_0, G_1, \cdots, G_K$. The unknown class membership variables $M_1, \cdots, M_n$ are independent multinomial with

$$M_i \mid (K, \pi) \overset{iid}{\sim} Mult(\pi_0, \ldots, \pi_K), \tag{1}$$

where $\pi_j = P[M_i = j]$ for $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_K)$ is the unknown probability of patient i being from class j, and $(K + 1)$ is the unknown number of latent classes with $\sum_{j=0}^{K} \pi_j = 1$. In some applications, this latent class distribution π may be a function of a set of covariates Z, however, for time being we assume that it does not depend on the observed p-dimensional fixed covariate vector $x_i = (x_{i1}, ⋭, x_{ip})$ that only affects the recurrent and termination events. We will later extend our model and methods to time-varying covariates.

Similar to currently popular JLCM models of longitudinal data (e.g, see Proust-Lima et al., 2014, and the references therein), we incorporate an unknown parameter η_j that models the relationship between the profile/trajectory of cumulative counts of NFTR recurrence $N_i(t)$ and the "point-process of termination" $D_i(t) = 1_{[T_i \leqslant t]}$ of termination time T_i for all patients from latent class G_j (see (2) and (3)). We make a clear distinction between "termination" at T_i due to GvHD (either death or total graft rejection) and the non-informative "censoring" at C_i (Kalbfleisch and Prentice, 2002) due to loss of follow-up, end of study, and other factors. Additionally for our JLCM, η_j is used to accommodate the dynamic effect of the observed history $\mathscr{H}_i(t-)$ on increments $dN_i(t) = N_i(t+dt) - N_i(t-)$ and $dD_i(t) = D_i(t+dt) - D_i(t-)$ of both recurrences and termination over time interval $[t, t+dt)$, where $\mathscr{H}_i(t-)$ up to time $t-$ (and not including t) is defined as the σ-algebra generated by the set $\{N_i(u), D_i(u), A_i(u) : u < t\}$ and $A_i(t) = 1_{[T_i, C_i \geqslant t]}$ is the "at observation process". For this purpose, we assume the intensity function to be

$$\lim_{dt \to 0} \frac{P\left[dN_i(t) > 0 \mid \mathscr{H}_i(t-), x_i, M_i = j, \eta_j\right]}{dt} = \lambda_j\left(t \mid x_i, \mathscr{H}_i(t-), \eta_j\right) = A_i(t)\lambda_0(t)\left[\eta_j k + \theta_i\right], \quad (2)$$

where $\theta_i = exp(\beta' x_i)$ with $\beta' x_i = \sum_{m=1}^{p} x_{im}\beta_m$ is the dynamic effect of covariate x_i, $\beta = (\beta_1, \cdots, \beta_p)$ is the regression parameter, $\lambda_0(t)$ is the baseline intensity function, and $N_i(t-) = k$ is the number of past recurrences at time t (included as part of the history $\mathscr{H}_i(t-)$). For a patient with $\{M_i = j\}$, the parameter η_j in (2), quantifies the dynamic effect of past recurrence history $\mathscr{H}_i(t-)$) on the risk of future recurrence $\{dN_i(t) > 0\}$ because every past recurrence contributes to an additional $\eta_j\lambda_0(t)$ to the risk of $dN_i(t)$ around time t. In particular, the first NFTR event for any latent group G_j has the common hazard function $\lambda_0(t)\exp(\beta x_i)$ with Cox's (1972) relative risk model for the covariate effect. The class G_0 with $\eta_0 = 0$ includes patients for whom future recurrence $dN_i(t)$ does not depend on number of past recurrences $N_i(t-)$.

For the increment $dD_i(t)$ in our JLCM, we assume the relative risk model (Cox, 1972)

$$\lim_{dt \to 0} \frac{P\left[dD_i(t) \mid \mathscr{H}_i(t-); x_i, M_i = j, \eta_j\right]}{dt} = A_i(t)h_j\left(t \mid \eta_j; x_i\right) = A_i(t)h_0(t)e^{\gamma x_i + \alpha\eta_j}, \quad (3)$$

where the unknown $\gamma = (\gamma_1, \cdots, \gamma_p)$ quantifies the dynamic effect of covariate vector x_i on $dD_i(t)$, and the scalar parameter α represents the fixed effect of the class-specific profile parameter η_j on the future risk/hazard of termination T_i when $M_i = j$. The practical assumption in (3) ensures that different latent classes have different risks of termination. Also, assumptions (2) and (3) together ensure that all patients within same class G_j share the same joint regression profile of recurrences and termination characterized by the unknown class-profile parameter η_j of G_j. For longitudinal data, the JLCM is a popular modeling option that allows for practical interpretation of covariate effects, heterogeneity of the population and comparison of various patients' response profiles within and across latent classes while bypassing distributional assumption on random effects. Our novel JLCM for recurrence and termination also aims to achieve all of these above goals.

A major challenge for a joint model is to present a good physical interpretation of the covariates effects on joint process $\{N_i, D_i\}$. Existing joint models use a shared patient-specific frailty random effect W_i to accommodate the dynamic dependence between $dN_i(t)$ and $dD_i(t)$ given the history $\mathscr{H}_i(t-)$ (Huang and Wang, 2004; Liu et al., 2004; Ouyang et al., 2013; Qu et al., 2017). These models even accommodate the effect of history $\mathscr{H}_i(t-)$ on $\{dN_i(t), dD_i(t)\}$ via the shared W_i. Consequently, the dynamic effects of x_i on $\{N_i, D_i\}$ can only be explained conditional on random W_i that varies among patients and cannot be reliably estimated. Furthermore, any direct interpretation of the dynamic effect of x_i on the marginal intensity $\lambda(t \mid \mathscr{H}_i(t-), x_i)$ and on the marginal hazard $h(t \mid \mathscr{H}_i(t-), x_i)$ are lacking because these functions (obtained after integrating the random W_i) do not have any interpretable functional forms. Thus, the profiles of two subjects with different covariate values are difficult to compare without some additional restrictive model assumptions. Unlike them, our JLCM model presents the dynamic effects of x_i on the joint profiles of $\{dN_i(t), dD_i(t)\}$ via (2) and (3) based on finite dimensional and estimable η.

The JCLM also presents a synthesized interpretation of covariate effects on multiple quantities of interest related to both N_i and D_i. This is apparent when we evaluate the covariate effects on important marginal functions such as the mean and the rate functions—both of them are unconditional on the observed history $\mathscr{H}_i(t-)$. We obtain the differential equation $d\mu_j(t|x_i) = E[dN_i(t)|x_i; M_i = j, \eta_j] = d\Lambda_0(t)[\eta_j\mu_j(t|x_i) + \theta_i]$ from (2), where $\mu_j(t|x_i) = E[N_i(t)|A_i(t) = 1; x_i, M_i = j, \eta_j]$ is the mean function (expected number) of recurrences given the patient in class G_j is under risk at time t. Solving this differential equation, we obtain the mean function

$$\mu_j(t \mid x_i) = \frac{\theta_i}{\eta_j}\left[exp\left\{\eta_j \Lambda_0(t)\right\} - 1\right], \tag{4}$$

and corresponding rate function $d\mu_j(t|x_i)/dt = \theta_i\lambda_0(t)exp\{\eta_j\Lambda_0(t)\}$ given that $T_i > t$. The (4) implies that even the population rate function $d\mu(t \mid x_i)/dt = \theta_i\lambda_0(t)\sum_{j=0}^{K}\left[\pi_j \exp\left\{\eta_j\Lambda_0(t)\right\}\right]$ given $\{T_i > t\}$ is proportional in time with interpretable fixed effect $\theta_i = \exp(\beta' x_i)$ of covariate x_i. This shows that unlike previous frailty models of Oakes (1992), Lawless (1995) and Lin et al. (2000) under non-informative termination, our JLCM model produces interpretable fixed effects of covariates and latent class index on the expected and rate of recurrences for a patient not terminated at time t. This property is similar to the property of the frailty model of Xu et al. (2017). However, for our transplant study as well as other practical applications, it is sensible to focus on the mean $\mu_j^*(t \mid x_i)$ of $N_i^*(t) = N_i(Min\{T_i, t\})$, the point-process of number of recurrences only until termination time T_i. Using similar arguments as to what were used for deriving (4), we obtain

$$\mu_j^*(t \mid x_i) = E\left[N^*(t) \mid x_i; M_i = j, \eta_j\right] = \theta_i \int_0^t S_j(u \mid x_i)\lambda_0(u) exp\left\{\eta_j \Lambda_0(u)\right\} du, \tag{5}$$

with the corresponding rate-function $r_j^*(t \mid x_i) \equiv d\mu_j^*(t \mid x_i)/dt = \theta_i S_j(t \mid x_i)\lambda_0(t) exp\left\{\eta_j \Lambda_0(t)\right\}$, where $S_j(t \mid x_i) = S_0(t)^{\exp\left(\gamma x_i + \alpha\eta_j\right)}$ is the survival function of T_i with corresponding class-specific hazard function in (3). Unlike previous shared-frailty models, (4) and (5) guarantee that the covariate effect θ_i on the cumulative mean function $\mu_j^*(t \mid x)$ and the rate function $r_j^*(t \mid x_i)$ (both unconditional on history) is same as the dynamic effect of x_i on the risk function $\lambda_j\left(t \mid x_i, \mathscr{H}_i(t-), \eta_j\right)$ for any subject i in G_j. Unlike expression (5) for the JLCM, the shared-frailty models lack any interpretation of the effects of x_i on the marginal mean $\mu^*(t|x)$ and rate $r^*(t|x)$ (after integrating out frailty) because these models provide no simple expressions for these functions (without some strong and unrealistic additional modeling assumptions). There are also issues regarding the sensitivity of these marginal regression functions, say, $r^*(t|x)$ and $\lambda(t \mid \mathscr{H}(t-); x)$, to the assumed parametric form of the frailty density. Recent shared-frailty models of Xu et al. (2017) focus solely on $E[N_i(t)|X_i]$ without considering termination at T_i, and do not provide the marginal function $r^*(t|x_i)$.

3. Bayesian Analysis of Joint Model

The observed data is the set $\mathbf{Y}_0 = \{x_i, y_i, \delta_i, N_i(t)$ for $0 < t \leqslant y_i : i = 1, \cdots, n\}$, where $y_i = min\{T_i, C_i\}$ is the last observation time and $\delta_i = 1_{[T_i < C_i]}$ is the censoring indicator for patient i. The likelihood under the JLCM in (1)–(3) based the observed data $\mathbf{Y}_0$ is a product of two following parts. Using the contributions from the observed NFTR recurrences $N_i(t)$ in the observation interval $(0, y_i]$, the first part based on the intensity function in (2) is:

$$L_R(\beta, \eta, \Lambda_0, M \mid \mathbf{Y}_0) = \prod_{i=1}^{n} \prod_{q=1}^{Q} \left[\left\{d\Lambda_0\left(t_q\right)\left(N_{iq}W_i^* + \theta_i\right)\right\}^{n_{iq}} \exp \left\{-A_{iq}\Lambda_{0q}\left(N_{iq}W_i^* + \theta_i\right)\right\}\right], \tag{6}$$

where $t_1 < \cdots < t_Q$ are ordered distinct NFTR recurrence and last observation times y_i from $i = 1, \cdots, n$ subjects, $\Lambda_{0q} = \Lambda_0\left(t_q\right) - \Lambda_0\left(t_{q-1}\right)$ is the increment in $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ in interval $I_q = (t_{q-1}, t_q]$ with $t_0 = 0$, A_{iq} is the at-risk indicator $A_i(t_q)$ of subject i at time t_q, $N_{iq} = N_i(t_q-)$ is the number of past NFTR recurrences to subject i before time t_q, $n_{iq} = N_{i,q+1} - N_{iq}$ is the number of NFTR recurrences occurring to subject i at time t_q, and $W_i^* = \sum_{j=0}^{K} \eta_j I(M_i = j)$. Under the hazard function (3), another part of the likelihood based on the observed (y_i, δ_i) is

$$L_S(\gamma, \eta, \alpha, H_0, M \mid \mathbf{Y}_0) = \prod_{i=1}^{n} \exp\left\{-H_0(y_i)\exp\left(\gamma x_i + \alpha W_i^*\right) \right\}\left[dH_0(y_i)\exp\left(\gamma x_i + \alpha W_i^*\right)\right]^{\delta_i}, \tag{7}$$

where $dH_0(t)$ is the increment in baseline cumulative hazard $H_0(t) = \int_0^t h_0(u)du$ in the interval $[t, t + dt)$. Full semiparametric Bayesian analysis (see Ibrahim et al., 2005) is based on the joint posterior distribution given by

$$\begin{aligned} & p(\beta, \gamma, \alpha, \eta, \Lambda_0, H_0, M \mid \mathbf{Y}_0) \propto L_R(\beta, \eta, \Lambda_0, M) \times L_S(\gamma, \alpha, H_0, M) \\ & \times \prod_{i=1}^{n} p_C(M_i \mid K, \pi) \times p_1(\Lambda_0) \times p_2(H_0) \times p_3(\eta \mid K) \times p_4(K, \pi) \times p_5(\beta, \alpha, \gamma), \end{aligned} \tag{8}$$

where $p_C(M_i|K, \boldsymbol{\pi})$ is the multinomial distribution of M_i in (1). The density of $p_4(K, \boldsymbol{\pi})$ is the prior of its parameter $(K, \boldsymbol{\pi})$, $p_1(\Lambda_0)$ and $p_2(H_0)$ are two independent prior processes for non-parametric cumulative functions $\Lambda_0(t)$ and $H_0(t)$ respectively, $p_3(\eta|K)$ is the prior distribution of $\eta = (\eta_1, \cdots, \eta_K)$ given K, and $p_5(\beta, \alpha, \gamma)$ is the joint prior of the regression parameters (β, α, γ). It is reasonable and common practice to assume a priori mutual independence of the regression parameters, baseline functions, and latent class parameters $(\eta, \boldsymbol{\pi}, K)$.

There are several ways to specify a prior $p_4(K, \boldsymbol{\pi})$ for unknown latent class variables $(K, \boldsymbol{\pi})$. Methods using $K+1$ to be known, as used in popular JLCM based joint analysis of survival and longitudinal data (Huang and Wang, 2004; Han et al., 2007; Proust-Lima and Taylor, 2009), usually lead to higher than adequate number of classes in practice. The Dirichlet process mixture (DPM) model (Neal, 2000) for W_i^* in (6) also leads to high computational cost and substantially higher than adequate number of classes. Provided it is supported by the observed data, it is desirable to have a small value of K to ensure that marginal mean, rate and intensity functions in (3) and (5) enable a comprehensive comparison among patients with different covariate values. A JLCM with large value of K is subject to the same criticisms leveled at shared-frailty models because shared-frailty models are in some sense JLCM with different classes for all different patients! So, we use the Mixture of Finite Mixtures (MFM) hierarchical prior (Miller and Harrison, 2016) for $p_4(K, \boldsymbol{\pi})$ in (8). This is presented hierarchically as

$$(\pi_0, \ldots, \pi_K) \mid K \sim Dir_{K+1}(\gamma, \ldots, \gamma) \text{ and } K \mid \zeta \sim Pois(\zeta), \tag{9}$$

where $Dir_m(a_1, \ldots, a_m)$ is the Dirichlet distribution with parameter $(a_1, \ldots, a_m)$, and $Pois(\zeta)$ is the Poisson distribution with mean ζ. A popular choice for the prior process $p_1(\Lambda_0)$ in (8) is the Gamma process (Kalbfleisch, 1978) denoted by $GP(\Lambda^*(t), b_\lambda)$, with a "prior guess" (prior mean) $\Lambda^*(t)$ of $\Lambda_0(t)$ and precision b_λ (assumed known). For example, $\Lambda^*(t) = a_\lambda t$ represents the user-specified $a_\lambda > 0$ being the prior guess for baseline intensity $\lambda_0(t)$. Similarly, we use $p_2(H_0)$ as $GP(H_0^*(t), b_h)$ with prior mean $H_0^*(t) = a_h t$ and precision b_h for some known a_h, $b_h > 0$. Unless there are substantial prior information about functions (Λ_0, H_0), these two Gamma processes with small precision b_λ and b_h can be reasonably approximated by independent Gamma priors for unknown increments $\Lambda_{0q} = \Lambda_0(t_q) - \Lambda_0(t_{q-1})$ and $H_{0q} = H_0(t_q) - H_0(t_{q-1})$ for $q = 1, \cdots, Q$ with prior mean $(t_q - t_{q-1})a_\lambda$ and variance $(t_q - t_{q-1})a_\lambda/b_\lambda$, and prior mean $(t_q - t_{q-1})a_h$ and variance $(t_q - t_{q-1})a_h/b_h$, respectively.

When we have useful prior information about both $\Lambda_0(t)$ and $H_0(t)$, we recommend a full semiparametric Bayesian analysis that is capable of inference as well as prediction using our JLCM in (2), (3) and (9). For such an analysis, we need MCMC samples from the posterior in (8). However, when there is a lack of credible prior information about (λ_0, h_0), we recommend following partial likelihood based semiparametric Bayesian inference.

Bayesian Analysis with Partial Likelihood:

Under the intensity function of (2) for JLCM, the partial likelihood for the recurrent events is

$$PL_R(\beta, \eta, M \mid \mathbf{Y}_0) = \prod_{i=1}^{n} \prod_{q=1}^{Q} \left\{ \frac{W_i^* N_{iq} + \theta_i}{\sum_{s=1}^{n} A_s(t_j)\left(W_s^* N_{sq} + \theta_s\right)} \right\}^{n_{iq}}, \tag{10}$$

where $A_s(t_q)$ is the "at risk" indicator of whether subject s is at observation at time t_q. Similarly, for observed (y_i, δ_i), the partial likelihood under the hazard in (3) is

$$PL_S(\gamma, \alpha, \eta, M \mid \mathbf{Y}_0) = \prod_{i=1}^{n} \left\{ \frac{\exp(\gamma x_i + \alpha W_i^*)}{\sum_{s=1}^{n} A_s(y_i)\exp(\gamma x_s + \alpha W_s^*)} \right\}^{\delta_i}. \tag{11}$$

Following arguments of Ibrahim et al. (2005), we can prove that the joint posterior

$$\begin{aligned} p_{PL}(\beta, \gamma, \alpha, \eta, M \mid \mathbf{Y}_0) &\propto PL_R(\beta, \eta, M \mid \mathbf{Y}_0) \times PL_S(\gamma, \alpha, \eta, M \mid \mathbf{Y}_0) \\ &\times \prod_{i=1}^{n} p_C(M_i \mid K, \pi) \times p_3(\eta \mid K) \times p_4(K, \pi) \times p_5(\beta, \alpha, \gamma) \end{aligned} \tag{12}$$

based on the partial likelihoods of (10) and (11) is always a proper joint density as long as the priors $p_3(\eta|K)$, $p_4(K, \pi)$, and $p_5(\beta, \alpha, \gamma)$ are proper. In Appendix I, we present a proof of the posterior of (12) being an approximation of the marginal posterior obtained via integrating (Λ_0, H_0) from the full posterior of (8) under very "diffuse" Gamma processes for $p_1(\Lambda_0)$ and $p_2(H_0)$. This gives a theoretical justification to use the posterior in (12) when there is no substantial prior opinion available for (Λ_0, H_0). Unlike the full posterior of (8), the posterior of (12) does not involve (Λ_0, H_0) and needs fewer steps within the MCMC while sacrificing the ability to make useful posterior predictions and posterior estimation of number and rate of future events.

The choice of priors for regression and variance component parameters often have substantial influence on Bayesian estimates (Gelman et al., 2006, 2008). For frailty models, the sensitivity of the results of Bayesian analysis to the priors of the frailty parameter is already well documented (Ouyang et al., 2013). Following Gelman et al. (2006), we present Bayesian analysis of JLCM using the ordered uniform distribution of size K as the "non-informative" prior and the ordered half-Cauchy distributions of size K and scale 2.5 as the "weakly-informative" prior for $\eta_1 < \cdots < \eta_K$. We use independent Cauchy density with center 0 and scale 2.5 as the priors for the regression parameters β and γ because these priors for regression parameters often outperform other non-informative and weakly-informative priors, including Gaussian and Laplace priors (Gelman et al., 2008). We use a

Gamma(1, 1) density as the prior for the parameter α associated with the class-effects η_j on termination.

4. Simulation Study

Our first two simulation studies compare the performances of Bayesian estimates of mainly the single regression parameter obtained from 3 methods: (1) JLCM with ordered uniform in (−3, +3) priors for $\eta_1 < \cdots < \eta_K$, (2) JLCM with ordered half-Cauchy prior on $\eta_1 < \cdots < \eta_K$, (3) shared-frailty model of Huang and Wang (2004). We compare the performances of these 3 Bayesian methods at sample sizes $n = 100$ and $n = 400$. To compare performances of the Bayesian estimates from competing methods, these two as well as other simulation studies use 500 replicates of datasets from each simulation model and sample-size to approximate the relative bias (RB), the average posterior standard deviation (SD), and the approximate square-root of mean square error (RMSE) of the Bayesian estimates under different methods. To facilitate fair comparisons among all three models, we present results of only full Bayesian analysis (partial likelihood based Bayesian analysis is not readily available for shared-frailty model) of them. Following conventional choices (Bender et al., 2005), we use independent Cauchy priors with center 0 and scale 2.5 for all regression parameters, $GP(a_\lambda t, b_\lambda)$ and $GP(a_h t, b_h)$ with $b_\lambda = b_h = 0.001$ and $a_\lambda = a_h = 1$ for cumulative baseline functions Λ_0 and H_0 respectively.

All simulation models use the baseline functions $\lambda_0(t) = 1$ and $h_0(t) = 0.5$, and fixed censoring time $C_i = 2$. For Simulation Study 1 and 2, we simulate from JLCM with $\eta = (0, 0.4, 0.8)$ for $K + 1 = 3$, a positive association between recurrence and termination with $\alpha = 0.5$, and independent Bernoulli covariate $x_i \sim Ber(0.5)$. The only difference between two simulation models is that the simulation model of former has same direction of covariate effects on risks of both recurrence and termination with $\beta = \gamma = 1$, whereas in later simulation model these true covariate effects are in opposite directions with $\beta = 1$ and $\gamma = -1$. For, The values of RB, SD and RMSE in Table 1 (for Simulation Study 1) and Table 2 (for Simulation Study 2) indicate that JLCM based Bayesian estimates under uniform priors for η perform the best among competing methods. As expected, the RB and RMSE for smaller sample-size $n = 100$ are slightly larger than corresponding values obtained from larger datasets ($n = 400$), however, the estimates for both sample sizes have very small RB. Especially for the estimating η_1 and η_2, the JLCM performs better while using ordered uniform priors compared to using half-Cauchy priors on η, because the later method substantially underestimates η and over-estimates the number of latent groups K with a large RMSE. The RMSE values of the estimates of regression parameter from both JLCM based methods are smaller than the corresponding RMSE values from the shared-frailty model based estimates. Thus, the JLCM bases methods substantially outperform the shared-frailty method when the data is generated from a JLCM.

Simulation Study 3 tests the robustness of JLCM based Bayesian estimates via comparing these three estimates when the true simulation model is the shared-frailty model of Huang and Wang (2004) with conditional intensity function $\lambda(t \mid x_i, \mathscr{H}_i(t-), W_i) = \lambda_0(t) exp(x_i \beta)(1 + W_i)$ and hazard function $h(t \mid x_i, \mathscr{H}_i(t-), W_i) = h_0(t) exp(x_i \beta)(1 + W_i)$ with $\beta = \gamma = 1$, and the frailty density $W_i \sim$

Gamma(1.5, 1.5). Table 4 shows that the estimated regression parameters from all three competing methods have comparable RB and RMSE when the sample size is small (n = 100). However, as the sample-size increases (n = 400), the RB values of shared-frailty based regression estimates seem to decrease faster than those from JLCM based estimates. Thus, the JLCM with uniform prior for η is preferable for Bayesian estimates unless we are assured about the validity of the shared-frailty assumption and the sample size is large.

Our next three simulation studies now compare the estimates from JLCM with ordered uniform prior for η (since it performs better than Cauchy prior in previous three simulation studies) with those from the shared-frailty model when the simulated datasets have both binary and continuous covariates and the interaction among them. So, each of these simulation studies use 500 replicates of datasets of n = 100 subjects in each with two independent covariates $x_1 \sim Ber(0.5)$ and $x_2 \sim N(0.25, 1)$ and their interaction $x_3 = x_1 \times x_2$. In Simulation Study 4, the simulation model is JLCM with $\beta_1 = 0.5$, $\beta_2 = 0.2$, $\beta_3 = 0.6$, $\gamma_1 = 0.3$, $\gamma_2 = 0.4$ and $\gamma_3 = 0.3$ to ensure that the simulated datasets have approximately the same expected value of $X\beta$ and the same a number of recurrent events until terminationas as in Simulation Study 1. In Simulation Study 5, we simulate from same JLCM except with $\gamma_1 = -0.3$, $\gamma_2 = -0.4$ and $\gamma_3 = -0.3$ to ensure the direction of covariate effects on recurrent events to be different from the effects on termination (unlike in Simulation Study 4).

For Simulation Study 4–5, the values of RB, SD and RMSE of the estimates from two competing methods are in Table 4. These results show that the estimates from JLCM have similar performances to the JLCM based estimates in Simulation Study 1–2 with single binary covariate. However, the estimates from the shared-frailty model are perform worse than the results for JLCM except for the γ_2 corresponding to the effect of continuous covariate on termination. These results emphasize the earlier findings that the JLCM based estimates have substantially better performance than the shared-frailty model when the underlying true model is JLCM. Again, unlike Simulation Study 4 and 5 using simulations from JLCM, the Simulation Study 6 uses simulations from the shared-frailty model to assess the robustness of the estimates from JLCOM. The results in Table 4 show JLCM based estimates have comparable and even smaller RB than the shared-frailty model for some parameters. Values of SD and RMSE from JLCM are sometime little smaller than those from the shared-frailty model to indicate better performance of JLCM here. Overall, JLCM model based estimates have better performances than estimates from the shared-frailty model when there are multiple covariates.

Overall, these simulation studies show that the JLCM with ordered uniform priors for η performs better than JLCM with Cauchy priors, especially for small sample-size. JLCM gives reasonable estimates of regression parameters even when the true model is the shared-frailty model, and the estimates from JLCM performs much better than shared-frailty when the true model is JLCM.

5. Analysis of Heart Transplant Data

We compare (1) JLCM with ordered uniform priors for η and (2) shared-frailty model with gamma frailty using Bayesian analyses of a study of n = 114 cardiac transplant patients treated between 1992–2007 under these two competing models. Each patient is at risk of recurrent Non-Fatal Tissue Rejections (NFTR), usually treated with medication, as well as death due to GvHD (considered termination event). Some patients are censored due to loss of follow-up at their last follow-up times. The maximum number of observed recurrent NFTR events amnong these patients is 7, where the median and maximum of follow-up periods are 3 and 17.8 months. There are two binary covariates: race with $x_1 = 1$ for African American (AA) patients and $x_1 = 0$ otherwise, and Gender with $x_2 = 0$ for male and 1 for female.

We use independent mean 0 and variance 1 Gaussian priors for the regression parameters β_k and γ_k for $k = 1, 2$ to accommodate effectively non-informative prior opinions about the effects of race and gender, ordered uniform priors for vector η in JLCM, and exponential prior for the variance of the Gamma frailty of the shared-frailty model. To summarize the Bayesian analysis under two competing models, Table 5 presents the posterior means as Bayesian estimates (BE), posterior standard deviation (SD) and 95% credible interval (CI) as Bayesian interval estimates of the relevant parameters of interest.

For Bayesian analysis under JLCM, the interval estimates of K, $\boldsymbol{\pi}$ and η in Table 5 show a strong data evidence that this study has three latent classes with no class G_0 ($K = 3$ and $\boldsymbol{\pi}_0$ being very close to 0). This means that this patient population has no latent class for which the number of past NFTR events has no effect on the risk of GvHD event of the patient. To understand and assess the future risk of GvHD for every patient, the effect of his/her past history of NFTR events has to be considered. The Bayesian point estimates of class effects are $\hat{\eta}_1 = 0.504$, $\hat{\eta}_2 = 1.054$, and $\hat{\eta}_3 = 1.661$. Results show strong evidence of increased risk and rate of NFTR recurrence for any AA patient (compared to non-AA patient) with no termination at time t because the CI of $\exp(\beta_1)$ is (1.03, 2.28). However, there is no strong evidence of direct race-effect on the risk of termination because the CI of γ_1 is (−0.967, 0.861), containing 0. Also, the evidence of gender-effects on both recurrence and termination are weak because the CIs of both β_2 and γ_2 contain 0. These suggest that in spite of the strong data evidence of higher risk and higher rate of NFTR recurrences for an AA patient at any time t, there is no good data evidence of the AA patient being at higher risk of death from fatal GvHD after adjustment of the effects of of latent class and number of past recurrences. As a consequence of JLCM's property in (5) and results of our Bayesian analysis imply an increased population lifetime rate $r^*(t \mid x) = \sum_{k=0}^{K} \pi_k r_k^*(t \mid x)$ and an population lifetime mean NFTR recurrence $\mu^*(t \mid x) = \sum_{k=0}^{K} \pi_k \mu_k^*(t \mid x)$ for an AA patient compared to another non-AA patient at time t because when $\gamma = 0$ (as our Bayesian analysis results suggest for this study) we have $r^*(t \mid x) = \exp(\beta x_i)\lambda_0(t)\sum_{k=0}^{K} \pi_k S_k(t)\exp\{\eta_k \Lambda_0(t)\}$ and similar expression for $\mu^*(t|x)$.

The advantages of our JLCM based analysis is that we can compare the expected event profiles of two patients, say, an African American (AA) patient ($x_1 = 1$) versus a non-AA patient ($x_1 = 0$) of same gender within same latent class. The ratio e^{β_1} of their NFTR recurrence rates before termination at any time t has posterior mean 1.53 and CI (1.030,2.288) if they are from the same latent class. The ratio of risks of first recurrence (dynamic comparison given past history of recurrences at time t) between these two patients is also same as the rate-ratio e^{β_1}. However, this ratio of risks of recurrence is $\left(\eta_j + e^{\beta_1}\right)$ if, say, an AA male patient is at risk for the second recurrence and the non-AA male patient is still at risk of first NFTR at that time-point. The Bayesian point estimates for this risk-ratio are 2.038, 2.588 and 3.195 when they from classes 1, 2 or 3 respectively. Because our JLCM based analysis produces moderate number of latent classes, it is possible to compare the dynamic event profiles and mean/rate of events among patients from two different latent classes and even among patients with latent classes unknown. For example, the interval estimate of ratio of mean number of recurrences is (1.5, 4.3) when the latent class is unknown and the model even incorporates covariate effects on termination. Unfortunately, for the sake of brevity, we omit detailed comparisons of future event trajectories of different patients.

JLCM based Bayesian analysis also allows the estimation of probability π_k of any patient being in a latent class G_k and also facilitates the updating the estimates given the past events history of any subject. For example, Bayesian point estimate of π_3 is 0.7 for an AA Male patient with recurrence history as a patient $i = 6$ and without termination compared to this Bayesian estimate being less than 0.2 for a future patient with events history similar to the patient $i = 1$.

In Table 5, the posterior means and CIs of β_1 and β_2 under shared-frailty model are close to the corresponding estimates from JLCM. Overall, analysis from both models have agreement about the evidences of dynamic effects of race and gender on NFTR recurrence and termination conditional on history. However, the shared-frailty model cannot effectively interpret the ratio of rates of NFTR recurrence and ratio of termination risk of two patients with different covariate values. So, the JLCM based analysis is preferable because it allows comparisons of event profiles of two future patients and accommodates a comprehensive interpretation of covariate effects on all relevant functions.

6. Conclusion and Discussion

Our novel JLCM achieves five major practical/clinical goals: (1) explaining the effect of covariates on the future event profiles within each patient; (2) evaluating the risk of events in $[t, t + dt)$ given the history $\mathscr{H}(t-)$; (3) assessing the risk of termination given $\mathscr{H}(t-)$; (4) explaining the heterogeneity among patients via latent class parameters η; (5) providing predictions of future events. Unlike JLCM, existing methods often focus on single main response of interest (say, recurrence) and the corresponding regression function of interest (say, mean number of recurrence), and regression parameters of mean recurrence, in general, do not have any physical interpretation for another regression function, say, for hazard function for termination (Miloslavsky et al., 2004).

We can accommodate right-predictable time-varying covariate $x_i(t)$ within the joint latent class model of (2–3) via re-expressing them as $\lambda_j\left(t \mid \mathscr{H}_i(t-); \eta_j\right) = \lambda_0(t)\left[\eta_j k + \exp(\beta' x_i(t))\right]$ and $h_j\left(t \mid \mathscr{H}_i(t-); \eta_j\right) = h_0(t)e^{\gamma x_i(t) + \alpha\eta_j}$, where the event history $\mathscr{H}_i(t-) = \{N_i(u), D_i(u), A_i(u), x_i(u): u < t\}$ now also contains the information about the sample-path $\mathscr{X}_i(t) = \{x_i(u): u \leqslant t\}$ of the predictable process $\{x_i(\cdot)\}$ up to time t. Our full Bayesian method for studies with time-varying covariates is similar to what is presented in Section 3 as long as the entire sample path of time-varying $x_i(t)$ have been available in the interval when $A_i(t) = 1$. To facilitate the partial likelihoods (10) and (11) for our Bayesian method based on partial likelihoods will only require this time-varying $x_i(t)$ to be measured/known for all subjects at risk/observation at each event time (Li et al., 2016). Instead of (2), $d\mu_j(t \mid \mathscr{X}_i(t)) = E\left[dN_i(t) \mid x_i(t); \eta_j\right] = d\Lambda_0(t)\left[\eta_j\mu_j(t \mid x_i(t)) + e^{\beta x_i(t)}\right.$ is the new differential equation of the mean function (expected number) $\mu_j(t \mid X_i) = E\left[N_i(t) \mid A_i(t) = 1; \mathscr{X}_i(t), \eta_j\right]$ of recurrences given the patient in class G_j, with class-effect η_j, is under observation at time t. For ease of presentation, we consider the special case of piecewise constant $x_i(\cdot)$ with $X_i(t) = x_{ik}$ and for all $t \in I_k = (a_{k-1}, a_k]$ with the grid $0 = a_0 < a_1 < \cdots < a_{K-1} < a_K = \infty$. The solution of this differential equation in this case is the recursive formula

$$\mu_j(t \mid \mathscr{X}_i(t)) = \mu_j(a_{k-1} \mid \mathscr{X}_i(a_{k-1}))e^{\eta_j\Lambda_0(a_{k-1}, t)} + \frac{\theta_{ik}}{\eta_j}\left[e^{\eta_j\Lambda_0(a_{k-1}, t)} - 1\right] \text{ for } t \in (a_{k-1}, a_k], \tag{13}$$

where $\theta_{ik} = \exp(\beta x_{ik})$ and $\Lambda_0(a, b) = \int_a^b \lambda_0(t)dt$ for $0 \leqslant a < b$. Unlike (4), the class-specific rate function $d\mu_j(t \mid \mathscr{X}_i(t))/dt = \left\{\theta_{ik} + \eta\mu_j(a_{k-1} \mid \mathscr{X}_i(a_{k-1}))\right\}\lambda_0(t)\exp\left\{\eta_j\Lambda_0(a_{k-1}, t)\right\}$ as well as the population rate function $d\mu(t \mid \mathscr{X}_i(t))/dt$ given $\{T_i > t\}$ corresponding to (13) can not be expressed as a product of $\exp(\beta' x_i(t))$ and a baseline function free of $\mathscr{X}_i(t)$). However, the expression in (13) shows that similar to the fixed covariate case, the effect of the sample-path $\mathscr{X}_i(t)$ of time-varying $x_{(}\cdot)$ on mean function has two parts. The multiplicative effect of the current covariate value $x_i(t)$ is accommodated in the second-term of right-hand-side of equation (13), and the first part accommodates the effects of past sample path $x_i(u)$ for $u < t$. Obviously for this case, past sample-path $x_i(u)$ for $u < t$ may be different from the current value $x_i(t)$ of the covariate. Using arguments similar to what were used for deriving (5), we obtain the mean $\mu_j^*(t \mid \mathscr{X}_i(t)) = E\left[N_i^*(t) \mid \mathscr{X}_i(t); \eta_j\right] = \int_0^t S_j(u \mid \mathscr{X}_i(u))d\mu_j(u \mid \mathscr{X}_i(u))$ and the corresponding rate function

$$r_j^*(t \mid \mathscr{X}_i(t)) = \left\{\theta_{ik} + \mu_j(a_{k-1} \mid \mathscr{X}_i(a_{k-1}))\right\}S_j(t \mid \mathscr{X}_i(t))\lambda_0(t)\exp\left\{\eta_j\Lambda_0(a_{k-1}, t)\right\} \tag{14}$$

of $N_i^*(t) = N_i(Min\{T_i, t\})$ for $t \in (a_{k-1}, a_k]$, where $S_j(t \mid \mathscr{X}_i(t)) = \exp\left[-\int_0^t h_0(u)\left\{\exp(\gamma x_i(u)) + \alpha\eta_j\right\}du\right]$. In (14), the first term of $r_j^*(t \mid \mathscr{X}_i(t))$ representing the effect of the current value of covariate $x_i(t)$ is proportional to $\theta_{ik} = \exp(\beta' x_i(t))$. Computing the posterior estimates of $\mu_j^*(t \mid \mathscr{X}(t))$ and $r_j^*(t \mid \mathscr{X}(t))$ of any future patient are straightforward within Bayesian analysis as long as we use full Bayesian analysis (instead of partial likelihood based Bayesian analysis) that presents a Bayesian estimate of $\Lambda_0(t)$.

We present an innovative MCMC based tool that is scalable via popular Bayesian software such as JAGS (used in this paper) and WinBUGS because our computational method does not need Reversible Jump MCMC. This code is made available in Appendix II. We note that this JAGS code is not computationally feasible for massive datasets, and in this setting, we suggest optimizing the code using other software besides the standard JAGS option. Our simulation results show that JLCM produces good regression estimates even when the true model is not JLCM. Even though, we only consider non-negative η_j (appropriate for our transplant study), one can, in principle, consider even negative η_j as long as $exp(\beta x) + \eta_j N_i(t-) > 0$ for all observed values of $N_i(t-)$. Irrespective of the true model, JLCM based analysis is preferable because it allows comparisons of event profiles of two future subjects (via estimating class effects) and accommodates a comprehensive interpretation of covariate effects on all relevant functions.

Appendix I:: Partial Likelihood Based Posterior As The Marginal Posterior

We are going to show the partial likelihood based posterior in (12) is an approximation of the marginal posterior after integrating out the cumulative baseline function $\Lambda_0(t)$ and $H_0(t)$ from the joint posterior of (8). For $GP(\Lambda^*, b_\lambda)$ prior on $\Lambda_0(t)$, each increment Λ_{0q} of the $\Lambda_0(t)$ in in interval $I_q = (t_{q-1}, t_q]$, has a Gamma prior $Ga(a_\lambda w_{0q} b_\lambda, b_\lambda)$, where $w_q = (t_q - t_{q-1})$ and $t_1 < \cdots < t_Q$ are the ordered distinct event times. Then we integrate out the increments $d\Lambda_0(t)$ from the (6) as follows,

$$
\begin{aligned}
&PL_R(\beta, \eta, M \mid a_\lambda, b_\lambda; \mathbf{Y}_0)\\
&= \int L_R(\beta, \eta, \Lambda_0, M \mid \mathbf{Y}_0) \times p_1(\Lambda_0 \mid a_\lambda, b_\lambda) d\Lambda_0\\
&= \prod_{q=1}^{Q}\left\{\int \prod_{i=1}^{n}\left\{\Lambda_{0q}\left(N_{iq}W_i^* + \theta_i\right)\right\}^{n_{iq}} \times e^{-\Lambda_{0q}\sum_{i=1}^{n} A_{iq}\left(N_{iq}W_i^* + \theta_i\right)} \times e^{-b_\lambda \Lambda_{0q}}\left(\Lambda_{0q}\right)^{a_\lambda w_q b_\lambda - 1} d\Lambda_{0q}\right\}\\
&= \prod_{q=1}^{Q}\left\{\prod_{i=1}^{n}\left\{\left(N_{iq}W_i^* + \theta_i\right)\right\}^{n_{iq}} \times \int e^{-\Lambda_{0q}\left[\sum_{i=1}^{n} A_{iq}\left(N_{iq}W_i^* + \theta_i\right) + b_\lambda\right]}\right.\\
&\left.\times \left(\Lambda_{0q}\right)^{\sum_{i=1}^{n} n_{iq} + a_\lambda w_q b_\lambda - 1} d\Lambda_{0q}\right\}\\
&= \prod_{q=1}^{Q}\left\{\prod_{i=1}^{n}\left(N_{iq}W_i^* + \theta_i\right)\right\}^{n_{iq}} \frac{\Gamma\left(\sum_{i=1}^{n} n_{iq} + a_\lambda w_q b_\lambda\right)}{\left[b_\lambda + \sum_{i=1}^{n} A_{iq}\left(N_{iq}W_i^* + \theta_i\right)\right]^{\sum_{i=1}^{n} n_{iq} + a_\lambda w_q b_\lambda}}\\
&\propto \prod_{q=1}^{Q}\left\{\prod_{i=1}^{n}\left(N_{iq}W_i^* + \theta_i\right)\right\}^{n_{iq}}\left[b_\lambda + \sum_{i=1}^{n} A_{iq}\left(N_{iq}W_i^* + \theta_i\right)\right]^{-\sum_{i=1}^{n} n_{iq} - a_\lambda w_q b_\lambda}.
\end{aligned}
$$

When we choose a very diffuse Gamma processes with b_λ and $a_\lambda \to 0$, then the above marginal likelihood $PL_R(\beta, \eta, M)$ from recurrent events is approximately (in the limit) $\prod_{q=1}^{Q}\left\{\prod_{i=1}^{n}\left(N_{iq}W_i^* + \theta_i\right)\right\}^{n_{iq}}\left[\sum_{i=1}^{n} A_{iq}\left(N_{iq}W_i^* + \theta_i\right)\right]^{-\sum_{i=1}^{n} n_{iq}}$, same as the partial likelihood of (10) from recurrent events. Using similar steps as above, we can show that the marginal likelihood (after integrating $H_0(\cdot)$) from (y_i, δ_i)

$$
\begin{aligned}
&PL_S(\gamma, \eta, M \mid \mathbf{Y}_0) \int L_S(\gamma, \eta, H_0, M \mid \mathbf{Y}_0) \times p_2(H_0 \mid a_h, b_h) dH_0\\
&\propto \prod_{i=1}^{n} e^{\left(\gamma x_i + \alpha W_i^*\right)\delta_i}\left[b_h + \sum_{i=1}^{n} A_j(y_i)\left(\gamma x_i + \alpha W_i^*\right)\right]^{-\delta_i - a_h w_q b_h} \to PL_S(\gamma, \eta, M \mid \mathbf{Y}_0),
\end{aligned}
$$

as b_h and $a_h \to 0$ (11).

Appendix II:: Model Code in JAGS

```
# input data:
# x1, x2, x3: covariates
# YN[i, j]: number of events happened before time t[j] for subject i.
# Y[i, j]: indicator to show patients i is at risk or not at time t[j]
# t[j]: time point when j-th event happen among all subjects
# T: number of total different event time for all subjects
# N: total subjects number
# final[i]: location of censored time for subject i in variable t.
#Start model
model{
   # compute the log-likelihood by using the zero-trick in Poisson
distribution
   for(i in 1:N) { #Begin loop over subjects
      zeros[i]~ dpois(zeros.mean[i])
      M[i]~ dcat(pi[]) # the group for i-th subject
      for(j in 1:T) {#Begin loop over distinct recurrent event times
         Log.S1[i, j]=-dL0[j]*(K[i, j]*eta[M[i]]
+exp(x1[i]*beta[1]+x2[i]*beta[2]+x3[i]*beta[3]))*Y[i, j]
         Log.Lambda1[i, j]=(log(dL0[j])-log(t[j+1]-t[j])+log(K[i,
j]*eta[M[i]]+exp(x1[i]*beta[1]+x2[i]*beta[2]+x3[i]*beta[3])))*YN[i, j]
         dH[i,
j]=dH0[j]*exp(x1[i]*gamma[1]+x2[i]*gamma[2]+x3[i]*gamma[3]+alpha*eta[M[i]])*
Y[i, j]
      }
      L1[i]=sum(Log.Lambda1[i, 1:T])+sum(Log.S1[i, 1:T])
      log.H1[i]=-sum(dH[i, 1:T])
      log.H2[i]=(log(dH0[final[i]-1])-log(t[final[i]]-t[final[i]-1])
+x1[i]*gamma[1]+x2[i]*gamma[2]+x3[i]*gamma[3]+alpha*eta[M[i]])*fail[i]
      L2[i]=log.H1[i]+log.H2[i]
      LL[i]=L1[i]+L2[i]
      zeros.mean[i]=-LL[i]+C
   }
   # prior settings
   for (j in 1:T) {#Gamma process prior
      dL0[j]~dgamma((t[j+1]-t[j]), 0.001)
      dH0[j]~dgamma((t[j+1]-t[j]), 0.001)
   }
   #prior for regression parameters
   for(i in 1:3){
      beta[i]~dnorm(0, 0.16)
```

```
        gamma[i]~dnorm(0, 0.16)
    }
    alpha~dgamma(1, 1)
    # ordered prior for eta W[1]=0
    for(m in 2:num_class){
        W[m]~dunif(0, 3)
    }
    eta=sort(W)
    #establish a Dirichlet prior
    for(m in 1:num_class){
        a[m]~dgamma(1, 1)
        p[m]=ifelse(m<=KM, 1, 0)
        pi[m]<-a[m]*p[m]
    }
    #number of groups
    KM1~dpois(num_class -1)T(0, num_class -1) # number of groups exclude
group 0.
    KM=KM1+1
}
```

References

Barrett J, Diggle P, Henderson R, and Taylor-Robinson D (2015). Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77, 131–148. [PubMed: 25866468]

Bender R, Augustin T, and Blettner M (2005). Generating survival times to simulate cox proportional hazards models. Statistics in Medicine 24, 1713–1723. [PubMed: 15724232]

Cook RJ and Lawless J (2007). The statistical analysis of recurrent events. Springer Science & Business Media.

Cook RJ, Lawless JF, and Lee K-A (2010). A copula-based mixed poisson model for bivariate recurrent events under event-dependent censoring. Statistics in Medicine 29, 694–707. [PubMed: 20146240]

Cox D (1972). Regression analysis and life table. Journal of the RoyalStatistical Society: Series B (Statistical Methodology) 34, 187–222.

Gelman A et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). Bayesian Analysis 1, 515–534.

Gelman A, Jakulin A, Pittau MG, and Su Y-S (2008). A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics pages 1360–1383.

Han J, Slate EH, and Peña EA (2007). Parametric latent class joint model for a longitudinal biomarker and recurrent events. Statistics in Medicine 26, 5285–5302. [PubMed: 17542002]

Heitjan DF and Rubin DB (1991). Ignorability and coarse data. The Annals of Statistics pages 2244–2253.

Higgins R and Fishman J (2006). Disparities in solid organ transplantation for ethnic minorities: Facts and solutions. American Journal of Transplantation 6, 2556–2562. [PubMed: 16952299]

Hougaard P (2000). Analysis of multivariate survival data. Springer Science & Business Media.

Huang C-Y, Qin J, and Wang M-C (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. Biometrics 66, 39–49. [PubMed: 19459839]

Huang C-Y and Wang M-C (2004). Joint modeling and estimation for recurrent event processes and failure time data. Journal of the American Statistical Association 99, 1153–1165. [PubMed: 24068850]

Ibrahim JG, Chen M-H, and Sinha D (2005). Bayesian survival analysis. Wiley Online Library.

Kalbfleisch JD (1978). Non-parametric bayesian analysis of survival time data. Journal of the Royal Statistical Society. Series B (Methodological) pages 214–221.

Kalbfleisch JD and Prentice RL (2002). Relative risk (cox) regression models. The Statistical Analysis of Failure Time Data, Second Edition pages 95–147.

Kalbfleisch JD, Schaubel DE, Ye Y, and Gong Q (2013). An estimating function approach to the analysis of recurrent and terminal events. Biometrics 69, 366–374. [PubMed: 23651362]

Lancaster T and Intrator O (1998). Panel data with survival: hospitalization of hiv-positive patients. Journal of the American Statistical Association 93, 46–53.

Lawless J (1995). The analysis of recurrent events for multiple subjects. Applied Statistics pages 487–498.

Lee J and Cook RJ (2019). Dependence modeling for multi-type recurrent events via copulas. Statistics in Medicine 38, 4066–4082. [PubMed: 31236985]

Li S, Sun Y, Huang CY, Follmann DA, and Krause R (2016). Recurrent event data analysis with intermittently observed time-varying covariates. Statistics in medicine 35, 30493065.

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, Wolfe RA, and Huang X (2004). Shared frailty models for recurrent events and a terminal event. Biometrics 60, 747–756. [PubMed: 15339298]

Miller JW and Harrison MT (2016). Mixture models with a prior on the number of components. Journal of the American Statistical Association.

Miloslavsky M, Kele s, S., Laan MJ, and Butler S (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66, 239–257.

Neal RM (2000). Markov chain sampling methods for dirichlet process mixture models. Journal of computational and graphical statistics 9, 249–265.

Oakes D (1992). Frailty models for multiple event times. In Survival Analysis: State of the Art, pages 371–379. Springer.

Ouyang B, Sinha D, Slate EH, and Van Bakel AB (2013). Bayesian analysis of recurrent event with dependent termination: an application to a heart transplant study. Statistics in Medicine 32, 2629–2642. [PubMed: 23280968]

Paulon G, Iorio M, Guglielmi A, and Ieva F (2018). Joint modeling of recurrent events and survival: a bayesian non-parametric approach. Biostatistics.

Proust-Lima C, Séne M, Taylor JM, and Jacqmin-Gadda H (2014). Joint latent class models for longitudinal and time-to-event data: A review. Statistical Methods in Medical Research 23, 74–90. [PubMed: 22517270]

Proust-Lima C and Taylor JM (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. Biostatistics 10, 535–549. [PubMed: 19369642]

Qu L, Sun L, and Liu L (2017). Joint modeling of recurrent and terminal events using additive models. Statistics and Its Interface 10, 699–710.

Shih JH and Louis TA (1995). Inferences on the association parameter in copula models for bivariate survival data. Biometrics pages 1384–1399. [PubMed: 8589230]

Sinha D, Maiti T, Ibrahim JG, and Ouyang B (2008). Current methods for recurrent events data with dependent termination: a bayesian perspective. Journal of the American Statistical Association 103, 866–878. [PubMed: 19169433]

Xu G, Chiou SH, Huang C-Y, Wang M-C, and Yan J (2017). Joint scale-change models for recurrent events and failure time. Journal of the American Statistical Association 112, 794–805. [PubMed: 28943684]

Ye Y, Kalbfleisch JD, and Schaubel DE (2007). Semiparametric analysis of correlated recurrent and terminal events. Biometrics 63, 78–87. [PubMed: 17447932]

Zeng D and Lin D (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. Biometrics 65, 746–752. [PubMed: 18945267]

Table 1

Comparison of Bayesian estimates using data simulated from a JLCM with a same covariate effects on recurrence and termination risks: RB is the relative bias, SD is the average posterior Standard-Deviation, and RMSE is the square-root of mean square error based on 500 replicates.

		n=100			n=400		
Methods	**Paramter**	**RB**	**SD**	***RMSE***	**RB**	**SD**	***RMSE***
JLCM with uniform prior for η	α	−0.026	0.421	0.165	−0.004	0.331	0.160
	β	−0.008	0.198	0.193	−0.005	0.139	0.137
	γ	−0.094	0.220	0.214	−0.057	0.154	0.177
	η_1	0.022	0.262	0.097	0.006	0.221	0.083
	η_2	0.078	0.433	0.145	0.004	0.381	0.124
	K	0.159	0.877	0.480	0.150	0.801	0.451
JLCM with Cauchy prior for η	α	0.043	0.497	0.211	0.008	0.361	0.185
	β	−0.081	0.187	0.203	0.005	0.122	0.111
	γ	−0.126	0.218	0.250	−0.074	0.146	0.205
	η_1	−0.942	0.028	0.377	−0.918	0.014	0.367
	η_2	−0.909	0.083	0.727	−0.948	0.016	0.758
	K	0.852	0.475	1.727	0.862	0.542	1.752
Shared-frailty Model	β	−0.163	0.207	0.263	−0.147	0.152	0.212
	γ	−0.127	0.245	0.276	−0.084	0.217	0.178

Table 2

Summary of performances of estimates from different methods when data is simulated from a JLCM with opposite covariate effects on recurrence and termination risks: RB is the relative bias, SD is the average posterior Standard-Deviation, and RMSE is the square-root of mean square error based on 500 replicates.

		n=100			n=400		
Methods	**Parameter**	**RB**	**SD**	***RMSE***	**RB**	**SD**	***RMSE***
JLCM with uniform prior for η	α	−0.034	0.423	0.017	−0.009	0.311	0.011
	β	0.004	0.174	0.004	0.004	0.151	0.003
	γ	0.066	0.284	0.066	0.047	0.236	0.044
	η_1	0.035	0.230	0.014	0.008	0.222	0.009
	η_2	0.033	0.373	0.026	0.005	0.364	0.015
	K	0.150	0.813	0.300	0.114	0.747	0.280
JLCM with Cauchy prior for η	α	−0.076	0.411	0.038	−0.026	0.330	0.033
	β	−0.050	0.166	0.050	−0.016	0.136	0.034
	γ	0.068	0.283	0.068	0.051	0.256	0.062
	η_1	−0.938	0.031	0.375	−0.653	0.013	0.328
	η_2	−0.895	0.096	0.716	−0.613	0.018	0.686
	K	0.923	0.333	1.846	0.935	0.313	1.548
Shared-frailty Model	β	−0.049	0.218	0.222	−0.024	0.217	0.197
	γ	0.186	0.327	0.375	0.173	0.259	0.376

Table 3

Summary statistics for estimates using data simulated from model introduced in Section 4.4 (i.e., a shared-frailty model). RB is the average relative bias, SD is the average posterior Standard-Deviation, and RMSE is the approximate square-root of mean square error.

		n=100			n=400		
Methods	**Parameter**	**RB**	**SD**	***RMSE***	**RB**	**SD**	***RMSE***
JLCM with uniform prior for η	β	−0.031	0.200	0.202	−0.012	0.116	0.121
	γ	−0.157	0.207	0.260	−0.108	0.149	0.159
JLCM with Cauchy prior for η	β	−0.050	0.198	0.207	−0.003	0.153	0.152
	γ	−0.152	0.207	0.257	−0.105	0.151	0.163
Shared-frailty Model	β	−0.036	0.208	0.211	−0.006	0.144	0.143
	γ	−0.122	0.235	0.265	−0.087	0.146	0.151

Table 4

Summary statistics for estimates using data simulated from simulation study 4 to 6 that introduced in Section 4.4. RB is the average relative bias, SD is the average posterior Standard-Deviation, and RMSE is the approximate square-root of mean square error.

		JLCM			Shared-frailty Model		
Simulation Model	**Parameter**	**RB**	**SD**	***RMSE***	**RB**	**SD**	***RMSE***
Simulation Study 4	α	−0.054	0.415	0.167	-	-	-
	β_1	0.020	0.218	0.234	−0.153	0.236	0.281
	β_2	0.021	0.160	0.165	−0.068	0.173	0.182
	β_3	−0.015	0.211	0.216	−0.094	0.238	0.246
	γ_1	−0.194	0.245	0.240	−0.327	0.293	0.283
	γ_2	−0.132	0.170	0.170	−0.080	0.206	0.184
	γ_3	−0.010	0.243	0.235	−0.034	0.292	0.269
	η_1	0.036	0.262	0.091	-	-	-
	η_2	0.085	0.428	0.135	-	-	-
	K	0.107	0.866	0.464	-	-	-
Simulation Study 5	α	−0.061	0.408	0.152	-	-	-
	β_1	−0.002	0.216	0.215	−0.168	0.235	0.470
	β_2	0.027	0.150	0.151	0.026	0.163	0.166
	β_3	0.004	0.183	0.189	0.041	0.220	0.218
	γ_1	0.228	0.259	0.243	0.485	0.304	0.445
	γ_2	0.024	0.177	0.167	0.161	0.213	0.200
	γ_3	0.108	0.276	0.282	0.184	0.324	0.317
	η_1	0.042	0.235	0.090	-	-	-
	η_2	0.040	0.377	0.124	-	-	-
	K	0.107	0.820	0.453	-	-	-
Simulation Study 6	β_1	−0.016	0.157	0.150	−0.088	0.224	0.191
	β_2	−0.051	0.155	0.147	−0.169	0.223	0.181
	β_3	−0.011	0.157	0.154	−0.088	0.223	0.187
	γ_1	−0.240	0.221	0.215	−0.319	0.271	0.261
	γ_2	−0.152	0.221	0.208	−0.205	0.270	0.261
	γ_3	−0.262	0.221	0.209	−0.289	0.270	0.252

Table 5

Results of heart transplant data based on the partial likelihood with the non-informative prior. BE is the posterior mean (Bayesian point estimate), SD is the posterior Standard-Deviation and 95% CI is the 95% credible interval of the parameter.

	JLCM			Frailty Model		
Parameter	**BE**	**SD**	**95% CI**	**BE**	**SD**	**95% CI**
β_1	0.428	0.205	(0.030,0.828)	0.429	0.159	(0.109,0.732)
β_2	0.261	0.212	(−0.153,0.651)	0.177	0.172	(−0.185,0.524)
γ_1	0.063	0.460	(−0.967,0.861)	−0.055	0.977	(−1.974,1.835)
γ_2	−0.076	0.435	(−0.973,0.807)	0.031	1.012	(−1.929,1.933)
α	0.152	0.126	(0.006,0.477)	-	-	-
η_1	0.504	0.399	(0.018,1.455)	-	-	-
η_2	1.054	0.511	(0.218,2.166)	-	-	-
η_3	1.661	0.566	(0.598,2.730)	-	-	-
K	3.212	0.755	(2.000,4.000)	-	-	-