Precision Medicine Using a Joint Longitudinal-Survival Model

Yuxi Zhao* Elizabeth H. Slate*

Abstract

Biomedical studies often monitor subjects using a longitudinal marker that may be informative about a time-to-event outcome of interest. Examples are periodic monitoring of prostate specific antigen (PSA) as the longitudinal marker and time to onset of prostate cancer, and CD4 cell count as marker together with time to death from AIDS. Models that handle the two outcomes jointly and take advantage of their dependence have potential to improve inference for each. We develop a fully Bayesian joint longitudinal-survival model that uses a latent class structure to facilitate discovery of subgroups exhibiting distinct behavior. Subgroups may vary according to covariate effects, for example, time trends or the degree of response to intervention in the context of a clinical trial. Our formulation incorporates estimation of the number of subgroups and offers enhanced flexibility with a subgroup-specific piecewise linear log baseline hazard. We derive the correlation between the longitudinal and survival outcomes induced in our formulation and graphically display this dependence. We further derive prediction of the survival end point conditional on the observed longitudinal marker. Using simulation, we demonstrate the ability of our joint model to recover the true number of subgroups in the population and evaluate prediction of survival. Analysis of data from an AIDS clinical trial illustrates the model and suggests greater precision than prior analyses in the literature.

Key Words: Precision medicine, joint modeling, Bayesian, parametric, latent class

1. Introduction

It is common for biomedical studies to monitor subjects using a longitudinal marker that may be informative about a time-to-event outcome of interest. Once the event occurs or the subject is censored, the longitudinal process is no longer recorded. For example, in the clinical trials of AIDS patients (Abrams et al. 1994), CD4 cells per cubic millimeter of blood was measured at baseline, Month 2, Month 6, Month 12 and Month 18, and death was closely monitored for each participant. Jointly modeling the longitudinal biomarker process and event process offers a way to characterize their dependence and improve inferential efficiency compared to modelling each process separately. Often clinicians and researchers are primarily interested in the survival time, e.g., time to death, remission or relapse. The additional information from the marker process monitored for each individual facilitates precision medicine by enabling dynamic, individualized prediction of the survival event, which can be used to tailor screening schedules or therapy. Therefore, one of our objectives is to derive individualized prediction of the survival outcome taking account of the individual marker process.

There are several approaches available in the literature for characterizing the dependence between the two processes while targeting prediction of the event outcome. A first approach directly models the event process using imputed marker values as a time-dependent covariate (Andersen & Liestøl 2003). However, additional assumptions are needed for the imputation due to nonsynchronous observations of the marker values among the participants, which could introduce bias, especially for any ad hoc imputation without careful theoretical justification. A second approach models the two processes independently conditional on latent variables. Popular methods are the shared random effects

*Department of Statistics, Florida State University, 117 N. Woodward Ave., Tallahassee, FL 32306

model (including early works from Schluchter (1992) and De Gruttola & Tu (1994)), for which the latent variables are continuous, and joint modeling though latent class (Lin et al. 2002), through which the latent variable is discrete. A drawback of the shared random effects model is the need to specify the functional form of the latent variable through which individualized information extracted from the longitudinal process affects the survival outcome. Misspecification of this functional form can lead to loss in estimation efficiency (Huang et al. 2009). In contrast, the latent class joint model postulates conditional independence of the two outcomes given latent class and requires no specification of a functional form for the variable representing the latent class. We adopted this idea and propose a joint model with realization of the dependence between the two processes through latent classes. A mixed effects model is used for the longitudinal marker, and flexibility for the survival outcome is enhanced by using a log-piecewise linear baseline hazard. Parameterization for the longitudinal and survival submodels is designed to balance enhancing flexibility and avoiding nonidentifiability.

A major challenge when using a latent class model is selection of the number of latent classes. Non-Bayesian approaches typically refit the model for a range of numbers of latent classes and rely on post-model fitting assessment via information criteria. Within the Bayesian framework, the Dirichlet process mixture (DPM) induces latent classes, but is known to produce numerous small classes and so overestimate the number of clusters (Miller 2014). Another approach is the mixture of finite mixtures model (Miller & Harrison 2018), which requires a prior distribution for the number of latent classes and, although similar to the DPM, provides unbiased estimation of the number of classes. Estimation in the context of our joint model can be challenging, however, requiring complex approximation (Neal 2000) due to a lack of conjugacy, which has potential to extend computation time considerably. In this paper, a method to automate selection of latent classes is proposed using a spike-and-slab type prior to "activate" or "inactivate" the class in fitting. We use a similar approach to achieve automatic selection of knots in the log-piecewise formulation of the baseline hazard function.

The structure of this paper is as follows: In Section 2 we describe the details of proposed joint latent class model with ability to automatically select the number of latent classes and knots utilizing spike-and-slab priors. We also derive the correlation induced between the two processes and detail posterior prediction of the survival outcome for a new individual conditional on his/her available longitudinal marker readings. In Section 3, a simulation using data with three latent classes is presented. In Section 4, the analysis results of the AIDS data are shown. A discussion is included in Section 5.

2. Methodology

Denote the data as $\{(\mathbf{Y}_i, T_i), i = 1, 2, \ldots, n\}$, where $\mathbf{Y}_i = \{Y_i(t)\}$ is the vector of observed longitudinal measurements to date and T_i is the follow-up time for the i-th individual. Let x_i be a covariate such as treatment group assignment. Let C be a maximum number of latent classes. Also let $z_i = \{z_{i1}, z_{i2}, \cdots, z_{iC}\}$ be the vector of latent class membership, where $z_{ic} = 1$ indicates that individual i belongs to latent class c; define $z_i \sim \text{Categorical}(\pi_1, \cdots, \pi_C)$, where π_c is the membership probability of latent class c. The longitudinal and the event processes are modeled as independent conditional on the latent class. The longitudinal submodel is defined as

$$y_i(t) \mid (z_{ic} = 1) = a_{0c} + a_{1c}t + a_{2c}t^2 + \beta_1 t \cdot x_i + \beta_2 t^2 \cdot x_i + b_i + \epsilon_i(t) \qquad (1)$$

where a_{0c}, a_{1c}, and a_{2c} are the latent class-specific intercept, linear, and quadratic coefficients, respectively; β_1 and β_2 are the effects of the covariate on the slope and quadratic co-

efficients; $b_i \sim N(0, \sigma_b^2)$ denotes the subject-specific random effects and $\epsilon_i(t) \sim N(0, \sigma^2)$ is the error term. This formulation permits a latent class-specific trajectory adjusted for covariate effect and accounts for serial correlation of measurements within subject. For the survival submodel, we define the latent class-specific hazard function for the i-th individual as

$$h_i(t) \mid (z_{ic} = 1) = h_{0c}(t) \exp\{x_i \omega\} \tag{2}$$

where $\log\{h_{0c}\}$ is piecewise linear and ω is the coefficient for covariate x_i. This submodel provides additional flexibility over the formulation with common baseline hazard and class-specific coefficient. Furthermore, the log-piecewise linear hazard function has a closed form solution for the survival probability, which makes it possible to implement using standard software. The log-piecewise linear hazard function is defined as

$$\log\{h_{0c}(t)\} = \alpha_{0c} + \phi_{1c}\min(t, \xi_1) + \phi_{2c}\{\min(t, \xi_2) - \xi_1\} + \cdots + \phi_{K+1,c}\{\min(t, \infty) - \xi_K\} \tag{3}$$

where $\{\xi_1, \xi_2, \cdots, \xi_K\}$ is the vector of prespecified knots and K is the total number of knots. Figure 1 shows an example with three internal knots.

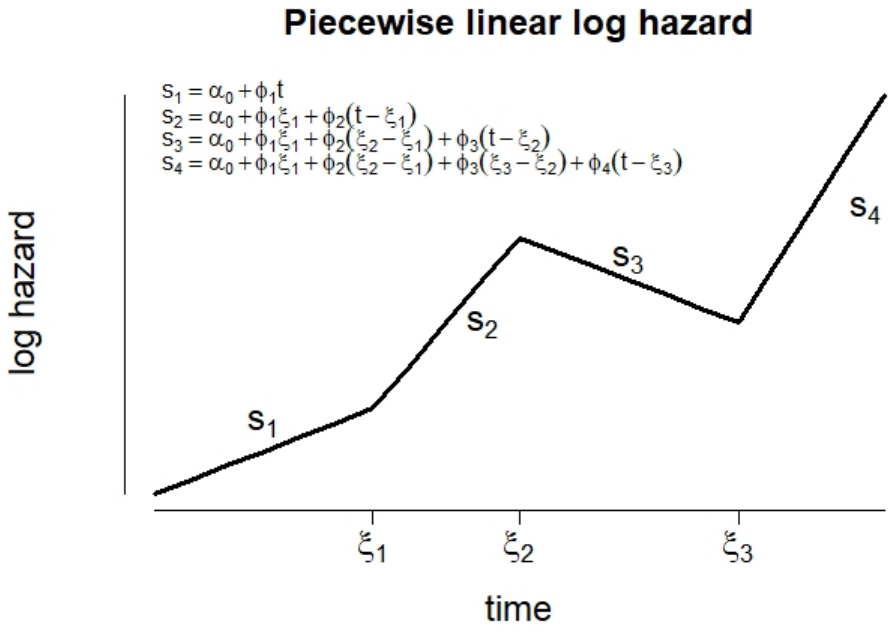

Figure 1: Example of a piecewise linear log hazard with three internal knots.

2.1 Latent Class and Baseline Hazard Knot Selection

2.1.1 Latent Class Section

To enable automatic selection of latent classes, a prior distribution is specified on the class membership probabilities $\{\pi_c, c = 1, 2, \ldots, C\}$ and spike-and-slab ideas are adapted to "activate" or "inactivate" each latent class. Denote the indicator for whether class c is active as Δ_c^0, where $\Delta_c^0 \mid \pi_c \sim$ Bernoulli(π_c) with π_c the membership probability for class c. Details of the latent class selection are specified in Algorithm 1. Linking the indicator to the corresponding membership probability enables the automatic selection: if

the membership probability of the latent class is close to zero, it is likely the latent class will be deactivated. One may use a symmetric Dirichlet distribution for the class membership probabilities, but we use the first latent class as a reference when parametrizing the class-specific baseline hazard functions. To encourage the first latent class to be the major class, we devise an ordered prior distribution for the membership probabilities as follows:

$$\begin{aligned}
\pi_1 &= V_1 \sim \text{Beta}(C, 1) \\
V_c &\sim \text{Beta}(C - c + 1, 1) \text{ for } c = 2, 3, \cdots, C - 1 \\
\pi_c &= V_c \prod_{j<c} (1 - V_j) \text{ for } c = 2, 3, \cdots, C - 1 \\
\pi_C &= 1 - \sum_{c=1}^{C-1} \pi_c
\end{aligned} \tag{4}$$

Algorithm 1 Latent Class Selection

class membership probabilities $\pi_1, \pi_2, \cdots, \pi_C \sim$ Dirichlet
for each baseline hazard knot in first class $k = 1, 2, \cdots, K + 1$ **do**
  generate slope $\phi_{k1} \sim N(0, \tau_\phi)$
end for
for each class $c = 2, 3, \cdots, C$ **do**
  $\Delta_c^0 \sim$ Bernoulli(π_c)
  $\gamma_{kc} \sim N(0, \tau_\phi)$ for $k = 1, 2, \cdots, K + 1$
  $\phi_{kc} = \phi_{1c} + \gamma_{kc}\Delta_c^0$ for $k = 1, 2, \cdots, K + 1$
end for

For the remaining parameters specific to class c in the longitudinal submodel, the same idea is applied to implement the latent class selection by $a_{0c} \cdot \Delta_c^0$, $a_{1c} \cdot \Delta_c^0$ and $a_{2c} \cdot \Delta_c^0$. The priors of γ, β, a_{0c}, a_{1c}, a_{2c}, and ω are specified as conditionally Gaussion with variances that arise from diffuse inverse gamma distributions.

2.1.2 Baseline Hazard Knot Selection

Baseline hazard knot selection is implemented similarly to latent class selection. The details are shown in Aglorithm 2. The first slope within each class is used as an anchor to define the increment for subsequent knots and an indicator Δ_{kc}^1 is used to indicate whether the knot k is selected for latent class c, with $\Delta_{kc}^1 \mid \pi^1 \sim$ Bernoulli(π^1) and $\pi^1 \sim$ Beta or π^1 prespecified (i.e., 0.05).

In combination with the prior distribution that orders the membership probabilities (4), it is useful to "anchor" the initial slope ϕ_{1c} in class $c = 2, 3, \cdots, C$ around the first slope ϕ_{11} in class 1, and then to parameterize using the successive slope increments across the knots. The extended algorithm is described in Algorithm 3.

2.2 Model Characteristics

2.2.1 Correlation

The correlation of the two processes at time t is defined to be Corr$(Y_i(t), I(T_i \leq t))$. It can be visualized on a defined time grid $t \in \mathcal{T}$. To derive this correlation, note that the covariance is

$$\text{Cov}(Y_i(t), I(T_i \leq t)) = \text{E}[\text{E}(Y_i(t), I(T_i \leq t) \mid z_i)] = \text{E}[\text{E}(Y_i(t) \mid z_i)\text{E}(I(T_i \leq t) \mid z_i)].$$

Algorithm 2 Latent Class Selection and Baseline Knot Selection

$\pi_1, \pi_2, \cdots, \pi_C \sim$ Dirichlet
for each class $c = 1, \cdots, C$ **do**
  $\phi_{1c} = \gamma_{1c} \sim N(0, \tau_\phi)$
  for each knot $k = 2, 3, \cdots, K+1$ **do**
    $\Delta_c^0 \sim$ Ber(π_c)
    $\Delta_{kc}^1 \sim$ Ber(π^1)
    $\gamma_{kc} \sim N(0, \tau_\phi)$
    $\phi_{kc} = \phi_{k-1,c} + \gamma_{kc}\Delta_{kc}^1\Delta_c^0$
  end for
end for

Algorithm 3 Latent Class Selection and Baseline Knot Selection

$\pi_1, \pi_2, \cdots, \pi_C \sim$ Dirichlet
for first class $c = 1$ **do**
  $\phi_{11} \sim N(0, \tau_\phi)$
  for $k = 2, 3, \cdots, K+1$ **do**
    $\Delta_{k1}^1 \sim$ Ber(π^1)
    $\phi_{k1} = \phi_{k-1,1} + \gamma_{k1}\Delta_{k1}^1$
  end for
end for
for rest classes $c = 2, 3, \cdots, C$ **do**
  $\Delta_c^0 \sim$ Ber(π_c)
  $\Delta_{1c} \sim$ Ber(π^1)
  $\gamma_{1c} \sim N(0, \tau_\phi)$
  $\phi_{1c} = \phi_{11} + \gamma_{1c}\Delta_{1c}^1\Delta_c^0$
  for $k = 2, 3, \cdots, K+1$ **do**
    $\Delta_{kc}^1 \sim$ Ber(π^1)
    $\gamma_{kc} \sim N(0, \tau_\phi)$
    $\phi_{kc} = \phi_{k-1,c} + \gamma_{kc}\Delta_{kc}^1\Delta_c^0$
  end for
end for

Writing the longitudinal marker submodel (1) more compactly as $Y_i(t) \mid (z_{ic} = 1) = \mu_{ic}(t) + b_i + \epsilon_i(t) = \mu_i(t)^T z_i + b_i + \epsilon_i(t)$, where $\mu_i(t) = (\mu_{i1}(t), \mu_{i2}(t), \cdots, \mu_{iC}(t))^T$, we have $\mathrm{E}(Y_i(t) \mid z_i) = \mu_i(t)^T z_i$. Similarly write $E(I(T_i \leq t) \mid z_i) = (1 - S_i(t))^T z_i$, where $S_i(t) = (S_{i1}(t), S_{i2}(t), \cdots, S_{iC}(t))^T$ and $S_{ic}(t) = \exp\{-\int_0^t h_{ic}(t)dt\}$. Let $Q = \mathrm{Var}(z_i)$ with diagonal elements equal to $\pi_c(1 - \pi_c)$ and off diagonal elements equal to $-\pi_c \pi_{c'}$. Then

$$\mathrm{Cov}(Y_i(t), I(T_i \leq t)) = \mu_i(t)^T Q(1 - S_i(t)),$$

and similarly

$$\begin{aligned} \mathrm{Var}(Y_i(t)) &= \mu_i(t)^T Q \mu_i(t) \\ \mathrm{Var}(I(T_i \leq t)) &= (1 - S_i(t))^T Q (1 - S_i(t)) \end{aligned}$$

Because in the longitudinal process there are additional variabilities caused by the random effects b_i and error term $\epsilon_i(t)$, we define the adjusted correlation as

$$\mathrm{Corr}(Y_i(t), I(T_i \leq t)) = \frac{\frac{\mu_i(t)^T Q(1-S_i(t))}{\sqrt{\mu_i(t)^T Q \mu_i(t)}\sqrt{(1-S_i(t))^T Q(1-S_i(t))}}}{\sqrt{1 + \frac{\sigma_b^2 + \sigma^2}{\mu_i(t)^T Q \mu_i(t)}}}. \tag{5}$$

Given realizations from the posterior sampler $j \in \{1, 2, \cdots, \texttt{n.iter}\}$ (post convergence), the correlation $\mathrm{Corr}(Y_i(t), I(T_i \leq t))$ is computable for each $t \in \mathcal{T}$, providing the means to determine pointwise credible bands. One characteristic of this correlation: it can be shown that the parameters in the longitudinal submodel common across classes, i.e., β_1 and β_2, do not affect the value of the correlation. Note that because the determinant of Q is always zero by definition, to ensure $\sqrt{(1 - S_i(t))^T Q (1 - S_i(t))}$ is evaluable when $S_i(t)$ close to zero, a small number (1E-16) is added to the diagonal elements of Q in calculation.

2.2.2 Prediction of Survival Outcome Conditional on Longitudinal Observations

Consider now that our model has been fit to training data $\mathcal{D} = \{(\mathbf{Y_i}, T_i), i = 1, \cdots, n\}$, and we seek prediction for a new subject $n + 1$. More specifically, at time s, we have observed $\mathbf{Y_{n+1}}$, the longitudinal marker for subject $n + 1$ at discrete visit times prior to s, and our goal is to predict the survival outcome T_{n+1} given that $T_{n+1} > s$. We derive $\Pr(T_{n+1} \geq u \mid \mathcal{D}, \mathbf{Y_{n+1}}, T_{n+1} \geq s)$ as follows (where $u > s$ throughout):

$$\Pr(T_{n+1} \geq u \mid \mathcal{D}, \mathbf{Y_{n+1}}, T_{n+1} \geq s) = \frac{\Pr(T_{n+1} \geq u, \mathbf{Y_{n+1}} \mid \mathcal{D})}{\Pr(T_{n+1} \geq s, \mathbf{Y_{n+1}} \mid \mathcal{D})}.$$

The numerator is

$$\begin{aligned} &\Pr(T_{n+1} \geq u, \mathbf{Y_{n+1}} \mid \mathcal{D}) \\ &\quad = \int \Pr(T_{n+1} \geq u, \mathbf{Y_{n+1}} \mid \theta) p(\theta \mid \mathcal{D}) d\theta \\ &\quad = \int \sum_c \Pr(T_{n+1} \geq u \mid \theta, z_{n+1,c} = 1) f(\mathbf{Y_{n+1}} \mid \theta, z_{n+1,c} = 1) \Pr(z_{n+1,c} = 1 \mid \theta) p(\theta \mid \mathcal{D}) d\theta \\ &\quad = \frac{1}{\texttt{n.iter}} \sum_j \sum_c \Pr(T_{n+1} \geq u \mid \theta_c^{(j)}) f(\mathbf{Y_{n+1}} \mid \theta_c^{(j)}) \pi_c^{(j)}, \end{aligned}$$

where `n.iter` is the number of iterations of the selected posterior sampler, θ denotes all the parameters and $\theta_c^{(j)}$ denotes the j-th posterior sample specific to latent class c, and

Table 1: Simulation setting.

		Class-specific			
Submodel	Parameter	Class 1	Class 2	Class 3	Common
Longitudinal	a_{0c}	15	2	8	-
	a_{1c}	-2	3	1	-
	a_{2c}	0.5	-0.5	0.5	-
	β_1	-	-	-	2
	β_2	-	-	-	0.5
	σ_b^2	-	-	-	1
	σ^2	-	-	-	1
Survival	α_0	-8	-6	-6	-
	Knots	(2,3,4)	(2,4,6)	(2,3,5)	-
	ϕ	(1,0,2,1)	(2,1,-0.5,-2)	(2,-1,0.5,3.5)	-
	ω	-	-	-	2

$\pi_c^{(j)}$ denotes the j-th posterior membership probability of latent class c. Similarly, the denominator is

$$\begin{aligned}&\Pr(T_{n+1} \geq s, \mathbf{Y_{n+1}} \mid \mathcal{D}) \\ &\quad = \frac{1}{\texttt{n.iter}} \sum_j \sum_c p(T_{n+1} \geq s \mid \theta_c^{(j)}) f(\mathbf{Y_{n+1}} \mid \theta_c^{(j)}) \pi_c^{(j)}.\end{aligned}$$

3. Simulation

This section describes a simulation for 300 individuals in each of three classes (total $n = 900$). The parameter values for the longitudinal and survival submodels are shown in Table 1. The only covariate included was treatment group assignment simulated as Bernoulli(0.5). The longitudinal measurements were obtained for each integer time $t_i = \{0, 1, \cdots\}$ until the end of follow up for each subject $i = 1, 2, \cdots, n$. The censoring for the survival submodel was assumed to be noninformative. The true baseline hazard functions, density estimates of observed follow-up times, and marker trajectories are shown in Figure 2.

The maximum number of classes for posterior fitting was set to be $C = 5$ and the knots in the baseline hazard were pre-specified at $(2, 3, 4, 5, 6)$. The Markov chain Monte Carlo (MCMC) of Algorithm 3 was implemented in R using the `runjags` package (Denwood 2016) and JAGS (Plummer 2003) with multiple chains; convergence was achieved around $25,000$ iterations (Figure 3). As shown in Figure 3, three classes with non-zero posterior membership probabilities were selected. Posterior inferences were based on the last $1,000$ iterations from the sampler. The posterior baseline hazard functions for the three nonzero classes are in accordance with the true hazard functions except for the regions with no or low event occurrences (Figure 4). Figure 4 shows the observed marker trajectories and follow-up times color coded by class assignment as determined using a best posterior realization in a least squares sense (Dahl 2006).

A test dataset of size of 100 subjects for each latent class was generated to evaluate the performance of the fitted model. Conditional on the observed longitudinal marker process, the area under the receiver-operator characteristic curve (AUC) based on the prediction of survival at the true times was 0.83.

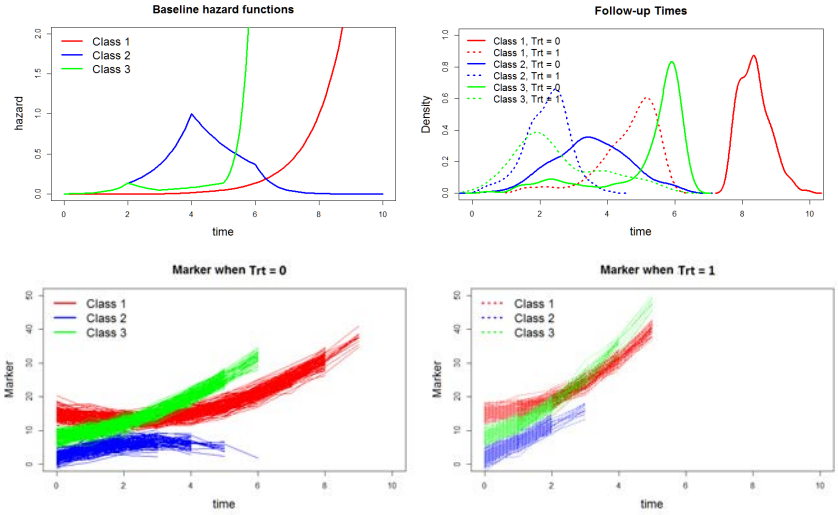

Figure 2: Simulation for three latent classes: Underlying baseline hazard functions, follow-up times density estimates, and observed marker trajectories. There are 300 subjects in each of the three classes.

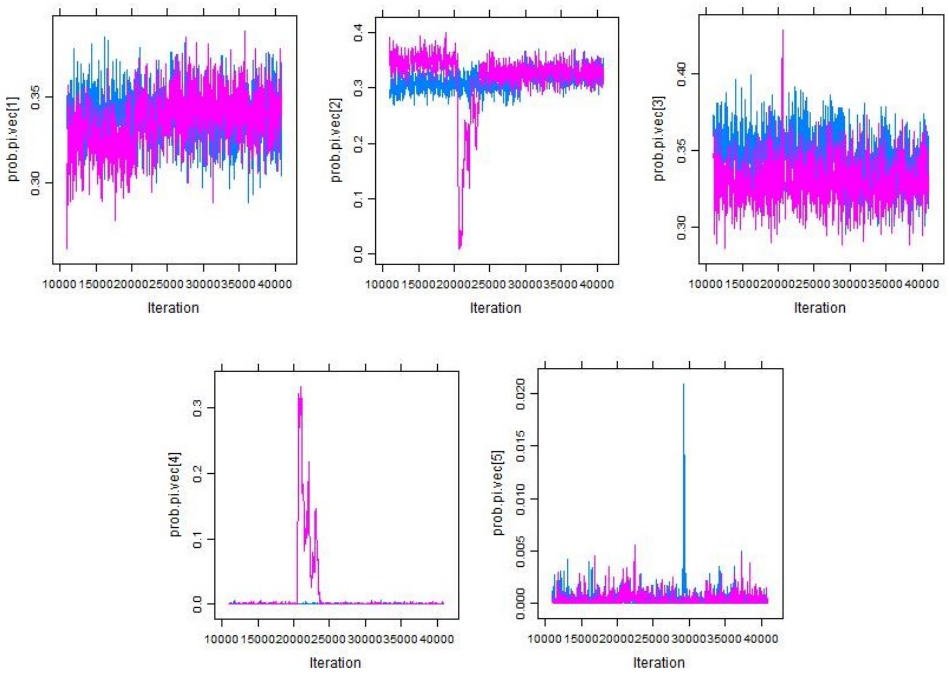

Figure 3: Simulation: Trace plots for the class membership probabilities for two MCMC chains. Convergence is evident at about 25,000 iterations.

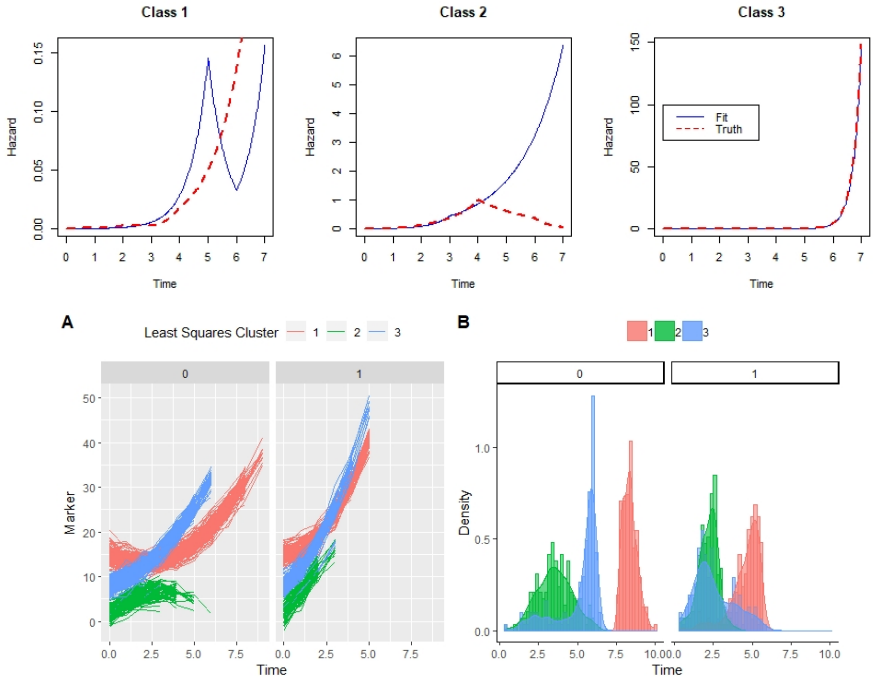

Figure 4: Simulation: Posterior inference. The top panels show the posterior estimates of the baseline hazards together with the true values. The bottom panels show the observed trajectories and follow-up times labeled according to class assignment as determined using Dahl's method (Dahl 2006).

4. AIDS Data Analysis

In the clinical trial (Abrams et al. 1994) assessing the efficacy and safety of two drugs in patients who had failed or exhibited intolerance to zidovudine therapy(AZT), 467 HIV-infected participants were recruited and randomly assigned to receive didanosine (ddI) or zalcitabine (ddC); 51% of the participants were enrolled in ddC. Patients were followed until death (40%) or censoring (60%). The associated biomarker was CD4 cells per cubic millimeter of blood with measurements taken at baseline, Month 2, Month 6, Month 12 and Month 18 (Figure 5). Additional covariates are gender (Female: 10%, Male: 90%), previous infection status with $0 =$ no previous infection (34%) and $1 =$ previous infection or AIDS diagnosis (66%), and a variable indicating either intolerance of AZT (63%) or failure of AZT therapy (37%).

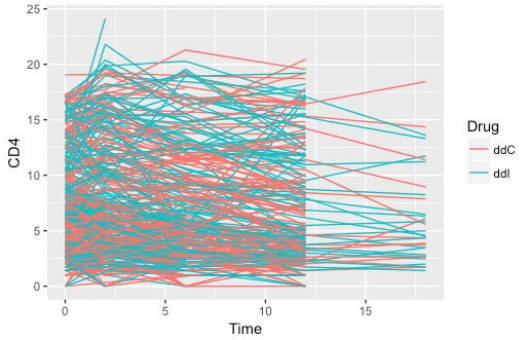

Figure 5: Trajectories of CD4 cell counts for 467 HIV-positive patients enrolled in a clinical trial (Abrams et al. 1994).

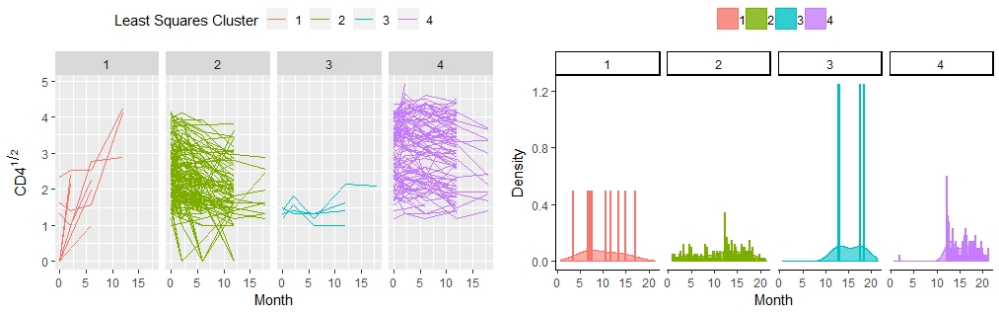

Figure 6: Analysis of the AIDS trial data: Classification of the 467 HIV-positive subjects according to the best (Dahl, 2006) inferential iteration.

Table 2: Analysis of the AIDS trial data: Covariate effects with 95% credible intervals.

Effect	Longitudinal $CD4^{1/2}$		Survival	
ddC on int	0.033	(-0.11, 0.16)	-0.49	(-0.94, -0.07)
ddC on t	-0.032	(-0.06, -0.01)		
ddC on t^2	0.002	(-0.00, 0.00)		
AIDS dx	-0.67	(-0.94, -0.42)	1.97	(1.25, 2.67)
AZT intol.	0.15	(-0.04, 0.32)	-0.08	(-0.57, 0.38)
Male	0.37	(0.18, 0.58)	-1.54	(-2.25, -0.65)

The data was split into training (70%) and test (30%). The biomarker CD4 was transformed by $\sqrt{\text{CD4}}$. The longitudinal and survival submodels used in this analysis are as follows:

$$\begin{aligned} \text{CD}_{ic}^{1/2}(t) &= a_{0c} + a_{1c}t + a_{2c}t^2 + \beta_1 t \cdot \text{ddC}_i + \beta_2 t^2 \cdot \text{ddC}_i \\ &\quad + \beta_{\text{ddC}} \cdot \text{ddC}_i + \beta_{\text{AIDS}} \cdot \text{AIDS}_i + \beta_{\text{intolerance}} \cdot \text{AZTintol}_i \\ &\quad + \beta_{\text{male}} \cdot \text{male}_i + b_i + e_i(t) \\ h_{ic}(t) &= h_{0c}(t) \exp[\omega_{\text{ddC}} \cdot \text{ddC}_i + \omega_{\text{AIDS}} \cdot \text{AIDS}_i + \omega_{\text{intolerance}} \cdot \text{AZTintol}_i \\ &\quad + \omega_{\text{male}} \cdot \text{male}_i] \end{aligned}$$

The number of classes for posterior fitting was set to be $C = 5$ and the knots were pre-specified at $(2, 4, \cdots, 18)$. MCMC was run for $30,000$ iterations with multiple chains and convergence was achieved around $27,000$ iterations. Posterior inference was based on the last $1,000$ iterations. From the realization with the best cluster configuration according to least squares (Dahl 2006), four classes were selected with $10, 205, 4$, and 108 subjects each (Figure 6).

From Table 2, there is evidence for significant effects of ddC, previous AIDS infection or diagnosis (AIDS dx), and sex (see highlighted). This analysis reveals more covariate effects than Guo & Carlin (2004), which found only prior AIDS dx statistically significant, suggesting greater precision. The correlation plot for subjects enrolled in the ddC treatment group and with AIDS diagnosis are shown in Figure 7, together with the ddI and no prior diagnosis situation as reference. As can be seen, the marginal correlation for subjects with AIDS diagnosis is of slightly greater magnitude and better precision than the correlation for the other two groups shown. The calculated AUC is 0.68 for the prediction of death in the test data (30% holdout) conditional on observed CD4 readings. Figure 8 shows the

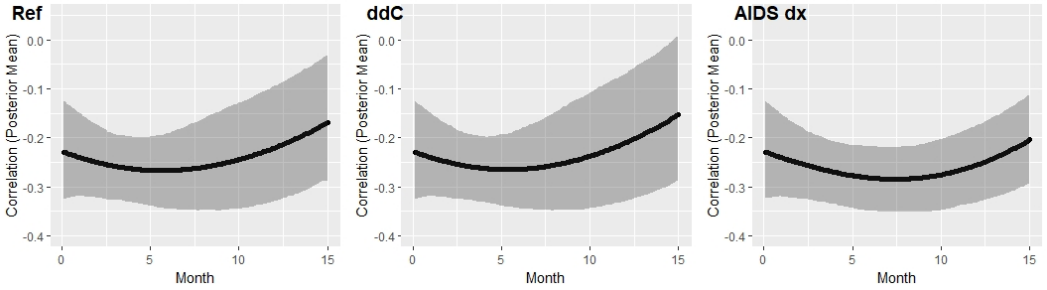

Figure 7: Analysis of the AIDS trial data: Correlation between the CD4 and time-to-death processes with 90% credible bands.

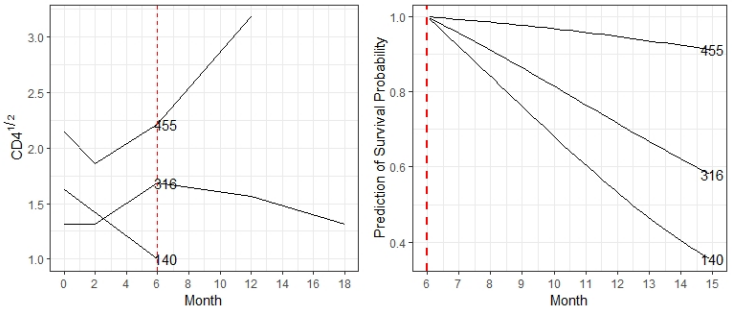

Figure 8: Analysis of the AIDS trial data: Prediction of survival conditional on observed CD4 cell counts for selected subjects.

predicted survival probability for three selected test subjects given observed CD4 readings to Month 6. As can be seen, with higher level of CD4 at baseline or increasing trend, subjects 456 and 316 have higher predicted survival probability than subject 140.

5. Discussion

The proposed model with automatic selection of number of latent classes shows the ability to recover the true number of latent classes effectively by specifying a relatively large maximum number of classes for fitting as shown in Section 3. The derived correlation and plot provides a way to visualize the relationship between the longitudinal marker and the event process; it enriches our understanding of the strength of the dependence and aids interpretation. In AIDS data analysis, negative correlations are observed for each time $t \in \mathcal{T}$ in general, which means higher CD4 cell counts are associated with higher probability of survival; moreover, the lack of a prior diagnosis of AIDS further increases the association with higher probability of survival. As compared to the findings from Guo & Carlin (2004), our analysis reveals better precision for covariate effects. The prediction of survival conditional on the observed longitudinal marker makes it possible for individualized prediction, which has great utility for helping practitioners tailor screening and therapy schedules.

There are a number of ways to further explore or extend this model. First, since this model uses a spike-and-slab method for selection of latent classes and knots, variable selection using spike-and-slab could be also easily incorporated. Second, there is a need for diagnostics to verify the underlying model assumption of conditional independence to ensure the validity of the analysis. Third, the goodness-of-fit needs to be examined in order to support our conclusions for the AIDS clinical trial data analysis and the comparisons to the work of Guo & Carlin (2004). Last, it is necessary to directly compare the performance

of our method to other joint model formulations such as the shared random effects model to further assess the strengths.

References

Abrams, D. I., Goldman, A. I., Launer, C., Korvick, J. A., Neaton, J. D., Crane, L. R., Grodesky, M., Wakefield, S., Muth, K., Kornegay, S. et al. (1994), 'A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection', *New England Journal of Medicine* **330**(10), 657–662.

Andersen, P. K. & Liestøl, K. (2003), 'Attenuation caused by infrequently updated covariates in survival analysis', *Biostatistics* **4**(4), 633–649.

Dahl, D. B. (2006), 'Model-based clustering for expression data via a dirichlet process mixture model', *Bayesian inference for gene expression and proteomics* **4**, 201–218.

De Gruttola, V. & Tu, X. M. (1994), 'Modelling progression of cd4-lymphocyte count and its relationship to survival time', *Biometrics* pp. 1003–1014.

Denwood, M. J. (2016), 'runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS', *Journal of Statistical Software* **71**(9), 1–25.

Guo, X. & Carlin, B. P. (2004), 'Separate and joint modeling of longitudinal and event time data using standard computer packages', *The American Statistician* **58**(1), 16–24.

Huang, X., Stefanski, L. A. & Davidian, M. (2009), 'Latent-model robustness in joint models for a primary endpoint and a longitudinal process', *Biometrics* **65**(3), 719–727.

Lin, H., Turnbull, B. W., McCulloch, C. E. & Slate, E. H. (2002), 'Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer', *Journal of the American Statistical Association* **97**(457), 53–65.

Miller, J. W. (2014), Nonparametric and variable-dimension Bayesian mixture models: Analysis, comparison, and new methods, PhD thesis, Brown University.

Miller, J. W. & Harrison, M. T. (2018), 'Mixture models with a prior on the number of components', *Journal of the American Statistical Association* **113**(521), 340–356.

Neal, R. M. (2000), 'Markov chain sampling methods for dirichlet process mixture models', *Journal of computational and graphical statistics* **9**(2), 249–265.

Plummer, M. (2003), JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, *in* 'Proceedings of the 3rd international workshop on distributed statistical computing', Vol. 124, Vienna, Austria, pp. 1–10.

Schluchter, M. D. (1992), 'Methods for the analysis of informatively censored longitudinal data', *Statistics in medicine* **11**(14-15), 1861–1870.