Received: 30 May 2022 | Revised: 3 August 2023 | Accepted: 23 August 2023

DOI: 10.1002/sim.9896

RESEARCH ARTICLE

Statistics in Medicine WILEY

Semiparametric multivariate joint model for skewed-longitudinal and survival data: A Bayesian approach

Jiaqing Chen[1] | **Yangxin Huang**[2] | **Qing Wang**[3]

[1]Department of Statistics, College of Science, Wuhan University of Technology, Wuhan, China

[2]College of Public Health, University of South Florida, Tampa, Florida, USA

[3]Yunnan Key Laboratory of Statistics Modeling and Data Analysis, Yunnan University, Kunming, China

Correspondence
Yangxin Huang, College of Public Health, University of South Florida, Tampa, FL 33612, USA.
Email: yhuang@usf.edu

Joint models and statistical inference for longitudinal and survival data have been an active area of statistical research and have mostly coupled a longitudinal biomarker-based mixed-effects model with normal distribution and an event time-based survival model. In practice, however, the following issues may stand-out: (i) Normality of model error in longitudinal models is a routine assumption, but it may be unrealistically violating data features of subject variations. (ii) Data collected are often featured by the mixed types of multiple longitudinal outcomes which are significantly correlated, ignoring their correlation may lead to biased estimation. Additionally, a parametric model specification may be inflexible to capture the complicated patterns of longitudinal data. (iii) Missing observations in the longitudinal data are often encountered; the missing measures are likely to be informative (nonignorable) and ignoring this phenomenon may result in inaccurate inference. Multilevel item response theory (MLIRT) models have been increasingly used to analyze the multiple longitudinal data of mixed types (ie, continuous and categorical) in clinical studies. In this article, we develop an MLIRT-based semiparametric joint model with skew-t distribution that consists of an extended MLIRT model for the mixed types of multiple longitudinal data and a Cox proportional hazards model, linked through random-effects. A Bayesian approach is employed for joint modeling. Simulation studies are conducted to assess performance of the proposed models and method. A real example from primary biliary cirrhosis clinical study is analyzed to estimate parameters in the joint model and also evaluate sensitivity of parameter estimates for various plausible nonignorable missing data mechanisms.

KEYWORDS

Bayesian analysis, longitudinal-survival data, Markov chain Monte Carlo, multilevel item response theory, multivariate joint model, skew-t distribution

Jiaqing Chen, Yangxin Huang, and Qing Wang share first authorship on the manuscript.

1 INTRODUCTION

In many clinical studies, multiple longitudinal biomarkers are measured with mixed types (eg, continuous and categorical), and time-to-event outcomes may be also recorded. The interrelationship between a time-to-event outcome and longitudinal biomarkers is often of interest. For example, in primary biliary cirrhosis (PBC) clinical study,[1] a number of publications[2-4] have investigated the association of longitudinal (repeated) measurements of time-dependent biomarkers such as serum bilirubin (SB), serum albumin (SA), hepatomegaly (HM), and histologic stage (HS) data with the risk of death times. In addition, information on some covariates may affect the processes of both longitudinal biomarkers and time-to-event outcome. For such studies, interest often lies in how a disease biomarker progresses with time and how it is influenced by covariates. However, there have been limited studies on simultaneously considering for longitudinal data with inherent features—skewness, missingness, correlation and mixed types along with linkage in specifying a time-to-event outcome together for joint modeling and inference.

Joint modeling analysis of longitudinal biomarkers and survival outcome is an active area of statistical research. One of the main reasons for the increasing interest is that joint modeling can be applied in various scientific problems with various data features and joint modeling methods may yield more precise inference.[5-10] As a result, many statistical modeling and analysis methods have been investigated for longitudinal-survival data analysis and inference. However, (i) the commonly assumed distribution for model error is normal due to mathematical tractability and computational convenience, but the continuously longitudinal outcomes may suffer from a serious departure of normality in which normality assumption may cause lack of robustness and subsequently lead to invalid inference, as exemplified in (Figure 1 [top panel]). Although one often handles skewed longitudinal data using transformation (eg, log-transformation),[2,3,11] the results may be difficult to interpret in their own or interpreted with different clinical meaning based on log-transformed longitudinal data. Thus, it is of practical interest to investigate models with skewed distributions including the skew-t (ST) distribution[12-15] for nonnormal longitudinal data. In the mean time, the (continuous) longitudinal measurements-based trajectories often display irregular nonlinear trends over time (Figure 1 [bottom panel]). Thus a parametric linear mixed-effects model is not appropriated to fit such data and a partially linear (semiparametric) mixed-effects model, which inherits the advantages from both parametric and nonparametric models, may be applied. (ii) In practice, many studies are designed to collect data with multiple longitudinal variables of mixed types (eg, continuous and categorical) which may be significantly correlated or associated such as SB and SA (continuous), HM (binary) and HS (ordinal) to be adopted in our real data application, and ignoring their interrelationships may result in biased results in estimation. (iii) Although most studies are designed to collect longitudinal data from every subject in the study at each visit time, missing values in the longitudinal data are often encountered because of subject dropout or discontinuation from studies. The missing data may be informative (nonignorable) in the sense that missingness may be related to the missing values; ignoring this phenomenon may result in inaccurate statistical inference. Additionally, longitudinal studies are commonplace in clinical and medical research, where repeated measurements, possibly censored time-to-event observations and other covariates are collected for each subject. One often focuses on interrelationships among these variables.[5,8-10,16] A joint model that links the hazard to these longitudinal measurements and covariates with incorporating some typical data features of interest, is becoming increasingly powerful in many fields.

To analyze mixed types (eg, continuous and categorical) of the multiple longitudinal data, the multilevel item response theory (MLIRT) models have been often applied.[17-20] According to MLIRT, the observed multivariate outcomes are viewed as imperfect clinical manifestations of a univariate subject-specific latent variable measuring disease severity. The MLIRT model normally consists of two levels. The first level measurement model quantifies the association between a subject-specific latent disease severity and the response to the multivariate outcomes. In the second level structural model, the latent disease severity is regressed on predictors (eg, treatment, disease duration and time) and subject-specific random-effects (describing between-subject variations) to study the overall treatment effects.[21-23] Advantages of the MLIRT models include better reflection of multilevel data structure, simultaneous estimation of measurement-specific parameters and covariate effects, and accurate inference about high-level measures.[24-26] To gain valid inference from the MLIRT models, marginal maximum likelihood methods,[23] and Bayesian methods[19,24,27-32] have been widely adopted. To the best of our knowledge, no studies have done to explore semiparametric multivariate joint models (SMJM) with the ST distribution, coupled with an MLIRT submodel for multivariate longitudinal variables of mixed types with various data features and a Cox proportional hazards submodel for survival endpoint, linked through random-effects. Inferential procedures become dramatically complicated when multivariate longitudinal data with nonnormality, correlation and missingness features involved and their interrelationships are conducted in conjunction with time-to-event into the SMJM.

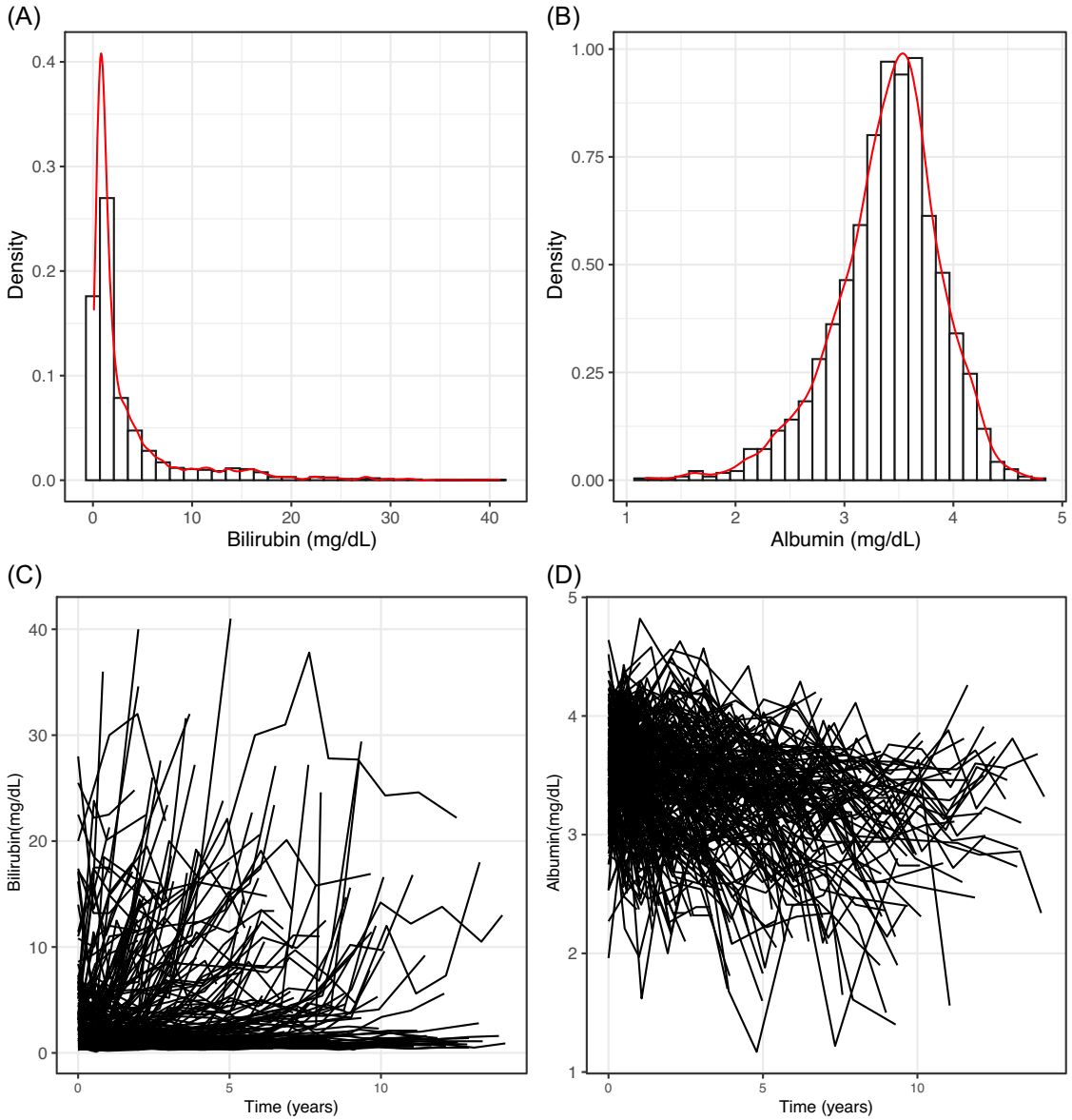

FIGURE 1 Histograms of serum bilirubin (SB) and serum albumin (SA) measurements for all patients from this dataset (top panel) and trajectory profiles of observed SB and SA values (bottom panel).

The remainder of this article evolves as follows. In Section 2, we formulate SMJM in which an extended MLIRT model is introduced for the multiple longitudinal variables of mixed types, where the continuous-based longitudinal submodel follows the ST distribution, and a Cox proportional hazards model for survival endpoint. Section 3 is to investigate associated Bayesian inferential approach in general form so that it can be applicable to various scientific fields. The proposed Bayesian method via Markov chain Monte Carlo (MCMC) algorithm is offered for inference. Section 4 is devoted to simulation studies for evaluating performance of the proposed joint models and method. In Section 5, we describe the PBC clinical study and data structure that motivated this research, present the specific SMJM formulation and analyze the PBC dataset to demonstrate the methodology. A concluding discussion is summarized in Section 6.

2 | SETUP OF SMJMS

In this section, we present the SMJM in full generality for the multiple longitudinal data with non-normality, correlation and missing values, and time-to-event outcome with censoring, illustrating that our model may be applicable in different

scientific fields. Let y_{ijk} be a measure of the kth longitudinal variable ($k = 1, 2, \dots, K$) from the ith subject ($i = 1, 2, \dots, N$) at time t_{ij} of visit j ($j = 1, 2, \dots, J_i$). Let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^T, \dots, \boldsymbol{y}_{iK}^T)^T$ be the K-variate vector of longitudinal responses, where $\boldsymbol{y}_{ik} = (y_{i1k}, \dots, y_{iJ_ik})^T$. Similarly, we can define the other related vectors. Let $\boldsymbol{r}_{ik} = (r_{i1k}, \dots, r_{iJ_ik})^T$ be an indicator vector of the kth longitudinal variable so that $r_{ijk} = 1$ if y_{ijk} is missing and 0 otherwise. Thus, we denote $\boldsymbol{y}_{\text{mis},ik}$ and $\boldsymbol{y}_{\text{obs},ik}$ are the collections of the missing and observed components of $\boldsymbol{y}_{ik}$, respectively. Let $\boldsymbol{\theta}_{ij}$ be a univariate latent variable measuring disease severity of subject i at visit j, with a higher value denoting more severe status. Let T_i be the observed event time for subject i, and $\boldsymbol{\varsigma}_i = (\varsigma_{i1}, \dots, \varsigma_{iJ_i})^T$ be a vector of event indicator for individual i, where $\varsigma_{ij} = 1$ or 0 if the event is observed or not by time t_{ij}. We introduce SMJM which consists of an extended MLIRT submodel for the multiple longitudinal variables of mixed types and a Cox proportional hazards submodel for the time-to-event outcome as follows.

2.1 | Extended MLIRT submodel for longitudinal data with missing values

In traditional MLIRT model, the longitudinal continuous response is specified by the usual linear model with the normal distribution,[19,24,27,28,31] where in the first-level measurement model, one modeled the continuous variables, the binary variables and the cumulative probabilities of ordinal variables by a two-parameter model,[27] graded response model,[27] and common factor model,[33] respectively. However, the trajectories of longitudinal continuous variables often show complex nonlinear patterns and such longitudinal data may exhibit nonnormal feature. Thus, we consider an extended version of MLIRT model which is specified by the semiparametric (partially linear) mixed-effects model-based MLIRT framework to take into account the complex nonlinear effect of time on the longitudinal continuous variables, where the semiparametric model is specifically composed of a linear component and a nonparametric part.[34,35] For the extended MLIRT model, in the first-level measurement model, the continuous variables, the binary variables and the cumulative probabilities of ordinal variables are modeled by the following specifications, respectively.

$$\begin{aligned} &y_{ijk} = a_k + b_k\theta_{ij} + w_k(t_{ij}) + h_{ik}(t_{ij}) + \epsilon_{ijk}, \\ &\text{logit}\{p(y_{ijk} = 1|\theta_{ij})\} = a_k + b_k\theta_{ij} + w_k(t_{ij}) + h_{ik}(t_{ij}), \\ &\text{logit}\{p(y_{ijk} \le l|\theta_{ij})\} = a_{kl} - b_k\theta_{ij} + w_k(t_{ij}) + h_{ik}(t_{ij}), \text{with } l = 1, 2, \dots, n_k - 1, \end{aligned} \tag{1}$$

where the random error for continuous variables $\boldsymbol{\epsilon}_i = (\boldsymbol{\epsilon}_{i1}^T, \dots, \boldsymbol{\epsilon}_{ik}^T, \dots, \boldsymbol{\epsilon}_{iK_1}^T)^T$ follows a multivariate ST distribution $ST_{K_1J_i,\nu}\left(-J(\nu)[\boldsymbol{\Delta}_{K_1} \otimes \mathbf{1}_{J_i}], \boldsymbol{\Sigma}_{K_1} \otimes \boldsymbol{I}_{J_i}, \boldsymbol{\Delta}_{K_1} \otimes \boldsymbol{I}_{J_i}\right)$[14,15] with unknown variance-covariance matrix $\boldsymbol{\Sigma}_{K_1} = (\sigma^2_{kk'})_{K_1\times K_1}$ ($k, k' = 1, 2, \dots, K_1$), unknown skewness parameter matrix $\boldsymbol{\Delta}_{K_1} = \text{diag}(\delta_1, \dots, \delta_{K_1})$, the vector of skewness parameters $\boldsymbol{\delta}_{K_1} = (\delta_1, \dots, \delta_{K_1})^T$, the degrees of freedom $\nu, J(\nu) = (\nu/\pi)^{1/2}[\Gamma((\nu-1)/2)/\Gamma(\nu/2)]$ and $\mathbf{1}_{J_i} = (1, \dots, 1)^T$. Note that $-J(\nu)[\boldsymbol{\Delta}_{K_1} \otimes \mathbf{1}_{J_i}]$ is specified here to make the mean zero of ST distribution and K_1 is the number of continuous variables. a_k is the outcome-specific difficulty parameter and b_k (always positive) is the outcome-specific discriminating parameter representing the discrimination of outcome k, that is, the degree to which outcome k discriminates between individuals with different disease severity θ_{ij} for patient i at visit time j, with higher value denoting more severe status, where θ_{ij} is a continuous latent variable, where t_{ij} is time for subject i at visit j. For the kth ordinal outcome with n_k categories, the order constraint $a_{k1} < \cdots < a_{kl} < \cdots < a_{kn_k-1}$ must be satisfied and the probability that patient i is in category l on outcome k at visit j is $p(y_{ijk} = l|\theta_{ij}) = p(y_{ijk} \le l|\theta_{ij}) - p(y_{ijk} \le l-1|\theta_{ij})$. $w_k(t)$ and $h_{ik}(t)$ are unknown nonparametric smooth fixed-effects and random-effects functions, respectively, $h_{ik}(t)$ are *iid* realizations of a zero-mean stochastic process.

In the second-level structural multilevel model, the latent disease severity θ_{ij} is regressed on covariates of interest, visit time and random-effects.

$$\theta_{ij} = (\boldsymbol{X}_{i0}^T\boldsymbol{\beta}_0 + u_{i0}) + (\boldsymbol{X}_{i1}^T\boldsymbol{\beta}_1 + u_{i1})t_{ij} + (\boldsymbol{X}_{i2}^T\boldsymbol{\beta}_2)t_{ij}^2, \tag{2}$$

where t_{ij} is time since treatment initiation for subject i at visit j ($t_{i0} = 0$ for baseline); $\boldsymbol{X}_{i0}, \boldsymbol{X}_{i1}$, and $\boldsymbol{X}_{i2}$ are vectors of covariates associated with baseline disease severity, progressive disease rate, and concavity of underlying individual trajectories of θ_{ij}, respectively; $\boldsymbol{X}_{i0}, \boldsymbol{X}_{i1}$, and $\boldsymbol{X}_{i2}$ may or may not be the same each other; the random intercept u_{i0} and random slope u_{i1} determine the subject-specific baseline disease severity and progressive disease rate, respectively. A limitation of this model is that it is not possible to add additional random-effects in the latent variable model of θ_{ij} above. The random-effects vector $\boldsymbol{u}_i = (u_{i0}, u_{i1})^T$ follows $N_2(0, \boldsymbol{\Sigma}_u)$, with covariance matrix $\boldsymbol{\Sigma}_u$ being denoted by $((1, \rho\sigma_u), (\rho\sigma_u, \sigma_u^2))$, where the variance of u_{i0} is set to 1 for identifiability, σ_u^2 is the variance of u_{i1} and ρ is the correlation coefficient. The random-effects vector

$\boldsymbol{u}_i$ accounts for three sources of correlations within the same subject: (i) inter-source (different outcomes at the same visit time), (ii) longitudinal (same outcome at different visit times) and (iii) cross-correlation (different outcomes at different visit times).[36] The vectors of regression parameters $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$ represent the covariate effects on the baseline disease severity, progressive disease rate and concavity of underlying individual trajectories of θ_{ij}, respectively. For example, if $\theta_{ij} = (\beta_{01}x_i + u_{i0}) + (\beta_{10} + \beta_{11}x_i + u_{i1})t_{ij} + (\beta_{20} + \beta_{21}x_i)t_{ij}^2$, where x_i is an indicator variable of treatment (1 if treatment, 0 otherwise), then β_{01} is the group difference of baseline disease severity, β_{10} and $\beta_{10} + \beta_{11}$ are the progressive disease rates for the placebo and treatment patients, respectively, and β_{20} and $\beta_{20} + \beta_{21}$ are the concavities of disease severity for the placebo and treatment patients, respectively. Note that no fixed intercept β_{00} included in the equation of θ_{ij} is an identifiability constraint induced by the model due to the latent process. The significantly negative β_{11} value indicates that the treatment is efficacious in slowing down the disease progression.

As mentioned in Section 1, a common problem of longitudinal studies is that some subjects may dropout of the study or miss scheduled visits due to various reasons. Missingness is likely to be informative or non-ignorable in the sense that the probability of dropouts or missing data may depend on the missing values. It is well known that, when dropouts or missing data are informative or nonignorable, inference-based statistical analysis of ignoring the missing data mechanism may lead to biased parameter estimates,[37] where a missing data mechanism is usually a model relating the probability of missing data to the missing or observed values. Here we consider reliable estimates of parameters based on a reasonable missing data mechanism. Although the assumed missing data models cannot be verified based on the observed data, subject-area knowledge and sensitivity analyses based on various plausible models may still lead to reasonable models. We first focus on the following simple nonignorable missing (NIM) data model (to avoid too many nuisance parameters) which may be reasonable in practice.[38,39]

$$\prod_{i=1}^{N}\prod_{k=1}^{K} f(\boldsymbol{r}_{ik}|\boldsymbol{y}_{ik}, \boldsymbol{\varphi}_k) = \prod_{i=1}^{N}\prod_{k=1}^{K}\prod_{j=1}^{J_i}[P(r_{ijk} = 1|\boldsymbol{\varphi}_k)]^{r_{ijk}}[1 - P(r_{ijk} = 1|\boldsymbol{\varphi}_k)]^{1-r_{ijk}}, \tag{3}$$

where $\text{logit}[P(r_{ijk} = 1|\boldsymbol{\varphi}_k)] = \varphi_{0k} + \varphi_{1k}y_{ijk}$, $\boldsymbol{\varphi}_k = (\varphi_{0k}, \varphi_{1k})^T$, and the missingness of the longitudinal response r_{ijk} depends on current response value, which may be missing, so the missing mechanism is nonignorable.[38,39] Note that the several alternative missing data models are considered for sensitivity analysis in Section 5.3.2.

The combined formulations (1)–(3) form the extended MLIRT submodel with subject-specific covariance (referred to as subject-specific MLIRT model).[24,27,28,40,41]

To fit model (1), we apply a regression spline method to the nonparametric smooth components $w_k(t)$ and $h_{ik}(t)$. The working principle is briefly described as follows and more details can be found in the literature.[16,35,42] The main idea of regression spline is to approximate $w_k(t)$ and $h_{ik}(t)$ by using a linear combination of spline basis functions. For instance, $w_k(t)$ and $h_{ik}(t)$ can be approximated by a linear combination of basis functions $\boldsymbol{\Psi}_p(t) = \{\psi_0(t), \psi_1(t), \ldots, \psi_{p-1}(t)\}^T$ and $\boldsymbol{\Phi}_q(t) = \{\phi_0(t), \phi_1(t), \ldots, \phi_{q-1}(t)\}^T$, respectively. That is,

$$w_k(t) \approx w_{kp}(t) = \sum_{l=0}^{p-1}\gamma_{kl}\psi_l(t) = \boldsymbol{\Psi}_p(t)^T\boldsymbol{\gamma}_{kp}, \quad h_{ik}(t) \approx h_{ikq}(t) = \sum_{l=0}^{q-1}\xi_{ikl}\phi_l(t) = \boldsymbol{\Phi}_q(t)^T\boldsymbol{\xi}_{ikq},$$

where $\boldsymbol{\gamma}_{kp} = (\gamma_{k0}, \ldots, \gamma_{k,p-1})^T$ is a $p \times 1$ vector of fixed-effects, and $\boldsymbol{\xi}_{ikq} = (\xi_{ik0}, \ldots, \xi_{ik,q-1})^T$ $(q \le p)$ is a $q \times 1$ vector of random-effects. We assume that $\boldsymbol{\xi}_{ikq} \overset{\text{iid}}{\sim} N_q(\boldsymbol{0}, \boldsymbol{\Sigma}_\xi)$ with $\boldsymbol{\Sigma}_\xi$ being unrestricted covariance matrix. Based on the assumption of $h_{ik}(t)$, we can regard $\boldsymbol{\xi}_{ikq}$ as *iid* realizations of a zero-mean random vector. We consider natural cubic spline bases with the percentile-based knots in the nonparametric part. The optimal degree of regression spline p and numbers of knots q can be also determined according to the Akaike information criterion (AIC) or the Bayesian information criterion (BIC).[42]

2.2 | Survival submodel for time-to-event

The event time T_i is likely related to the longitudinal process. This association is of much interest in many practical situations. We assume that the distribution of T_i, the event time for the ith subject, depends on the multivariate longitudinal processes through the random-effects $\boldsymbol{u}_i$, representing individual-specific longitudinal process and covariates and/or risk factors $\boldsymbol{X}_i$, respectively. We therefore consider a frailty model for T_i which is linked to the MLIRT submodel (1)–(3) through the random-effects $\boldsymbol{u}_i$. Additional covariates $\boldsymbol{X}_i$ are assumed to be associated with the time-to-event. Thus, we specify the

association by assuming that, conditional on the random-effects $\boldsymbol{u}_i$ in model (2), the event time $T_i \sim f(\cdot|\boldsymbol{u}_i)$, where $f(\cdot|\boldsymbol{u}_i)$ is a generic notation denoting a density function of the event time variable in the argument. That is, we assume that the event time is related to the multivariate longitudinal process through the subject-specific random-effects $\boldsymbol{u}_i$. In particular, letting t_i denote the time-to-event for the ith subject, we consider the following Cox proportional hazards model.

$$\lambda(t_i) = \lambda_0(t_i)\exp(\boldsymbol{X}_i^T\boldsymbol{\alpha} + \boldsymbol{u}_i^T\boldsymbol{\tau}), \tag{4}$$

where the function $\lambda(t)$ is the instantaneous hazard rate at time t, $\lambda_0(t)$ is an unspecified baseline hazard function. The association parameter vector $\boldsymbol{\tau} = (\tau_0, \tau_1)^T$ linking the random-effects $\boldsymbol{u}_i$ measuring the association between the two submodels. The SMJM is defined by the longitudinal MLIRT submodel (1)–(3) and the survival submodel (4), where two submodels are linked together through the shared random-effects $\boldsymbol{u}_i$, which is a popular approach in joint modeling.[5,6,8,35,43,44] The covariate vector $\boldsymbol{X}_i$ can be the same as or different from $\boldsymbol{X}_{i0}, \boldsymbol{X}_{i1}$, and $\boldsymbol{X}_{i2}$. An alternative method can be adopted to approximate the Cox proportional hazards model (4) through counting process[45] which is used for our joint modeling and can obviously reduce computational burdens; the detailed discussion for the alternative method can be found in References 34,35, and 46 and also provided in Section 1 of Appendix S1.

3 | SIMULTANEOUS BAYESIAN INFERENTIAL APPROACH

In general, estimation of a joint model for longitudinal and time-to-event data could be conducted in two ways. The first approach is based on the likelihood inferential methods, such as Expectation-Maximization (EM) algorithm and Monte Carlo Expectation-Maximization (MCEM) algorithm.[47] A simultaneous inference method based on a joint likelihood may be favorable, but the computational burden for proposed SMJM with the ST distribution can be extremely intensive with problems of algorithm convergence, and even sometimes infeasible.[10,48] The second approach, Bayesian inferential method shows the advantage. Thus, we simultaneously estimate parameters for the SMJM under a fully Bayesian framework via Markov chain Monte Carlo (MCMC) procedure based on the joint likelihood of the longitudinal-survival data and specified prior distributions. Thus, Bayesian joint modeling approach may pave a way to alleviate the computational burdens and to overcome convergence problems for such complex model setting. Bayesian analysis rests upon computing the conditional posterior distributions of the unknown parameters for inference, given the observed data and weighted by the prior information.

We assume that $\boldsymbol{\epsilon}_i$, $\boldsymbol{u}_i$, and $\boldsymbol{\xi}_{ikq}$ $(k = 1, 2)$ are mutually independent of each other. According to the property and stochastic representation of ST distribution described in the publications,[14,15] it can be shown that, by introducing the random vector $\boldsymbol{w}_i$ and random variable η_i, the SMJM specified by (1)–(4) can be hierarchically formulated as follows.

$$\begin{aligned}
&\boldsymbol{y}_{iK_1} \sim N_{K_1J_i}\Big(\boldsymbol{g}_{iK_1}(\boldsymbol{t}_i) + \boldsymbol{\Delta}_{K_1} \otimes [\boldsymbol{w}_i - J(\nu)\mathbf{1}_{J_i}], \eta_i^{-1}\boldsymbol{\Sigma}_{K_1} \otimes \boldsymbol{I}_{J_i}\Big),\\
&\boldsymbol{w}_i \sim N_{J_i}(\mathbf{0}, \eta_i^{-1}\boldsymbol{I}_{J_i})\boldsymbol{I}(\boldsymbol{w}_i > \mathbf{0}), \eta_i \sim \Gamma(\nu/2, \nu/2),\\
&r_{ijk}|\boldsymbol{\varphi}_k \sim \text{Bern}(p_{ijk}), \boldsymbol{\xi}_{ikq} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_\xi), \boldsymbol{u}_i \sim N_2(\mathbf{0}, \boldsymbol{\Sigma}_u),\\
&\zeta_{ijk} \sim \text{Bern}(p_{ijk}), \kappa_{ijk} \sim \text{Cat}(p_{ijkl}), T_i \sim F(t_i) = \int f(\boldsymbol{\varsigma}_i|\boldsymbol{u}_i),
\end{aligned} \tag{5}$$

where $\boldsymbol{g}_{iK_1}(\cdot) = (\boldsymbol{g}_{i1}(\cdot)^T, \ldots, g_{iK_1}(\cdot)^T)^T$ and $\boldsymbol{g}_{ik}(\cdot) = (g_{i1k}(t_{ij}), \ldots, g_{iJ_ik}(t_{ij}))^T$ with $g_{ijk}(t_{ij}) = a_k + b_k\theta_{ij} + \boldsymbol{\Psi}_p(t_{ij})^T\boldsymbol{\gamma}_{kp} + \boldsymbol{\Phi}_q(t_{ij})^T\boldsymbol{\xi}_{ikq}$, K_1 is the number of continuous variables. $p_{ijk} = p(\zeta_{ijk} = 1)$ for binary variable which follows Bernoulli distribution denoted by $\text{Bern}(p_{ijk})$ and $p_{ijkl} = p(\kappa_{ijk} = l)$ $(l = 1, 2, \ldots, n_k - 1)$ for ordinal variable which follows Categorical distribution denoted by $\text{Cat}(p_{ijkl})$. The contribution to the likelihood from the time-to-event model for the ith subject is denoted by $f(\boldsymbol{\varsigma}_i|\boldsymbol{u}_i)$. Note that the ST distribution here can be reduced to the following three special cases: (i) as $\nu \to \infty$ and $\eta_i \to 1$ with probability 1, then the ST distribution becomes an skewed normal (SN) distribution; (ii) as $\boldsymbol{\Delta}_{K_1} = \mathbf{0}$, then it reduces to a standard multivariate t-distribution; (iii) as $\nu \to \infty$, $\eta_i \to 1$ with probability 1 and $\boldsymbol{\Delta}_{K_1} = \mathbf{0}$, then it reverts to a standard multivariate normal distribution.

Let $\boldsymbol{\Theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}_{kp}, \boldsymbol{\delta}_{K_1}, \boldsymbol{\tau}, \rho, \sigma_u^2, \nu, \boldsymbol{\Sigma}_{K_1}, \boldsymbol{\Sigma}_\xi, a_k, b_k, a_{kl}(k = 1, \ldots, K; l = 2, \ldots, n_k - 1)\}$ be the collection of unknown population parameters in the joint model. The prior distribution represents the information about the uncertain parameter vector $\boldsymbol{\Theta}$ that is combined with the probability distribution of data to yield the posterior distribution for inference involving $\boldsymbol{\Theta}$. Under the Bayesian framework, we specify prior distributions for all unknown population parameters in our

joint model as follows.

$$\begin{aligned}
&\boldsymbol{\alpha} \sim N_3(\mathbf{0}, \boldsymbol{\Lambda}_1), \boldsymbol{\beta} \sim N_3(\mathbf{0}, \boldsymbol{\Lambda}_2), \boldsymbol{\gamma}_{kp} \sim N_p(\mathbf{0}, \boldsymbol{\Lambda}_3), \boldsymbol{\delta}_{K_1} \sim N_{K_1}(\mathbf{0}, \boldsymbol{\Lambda}_4), \boldsymbol{\tau} \sim N_2(\mathbf{0}, \boldsymbol{\Lambda}_5),\\
&\rho \sim \text{Unif}(-\vartheta, \vartheta), \sigma_u^2 \sim \text{IG}(\vartheta_1, \vartheta_2), \nu \sim Exp(\nu_0)I(\nu > 2), \boldsymbol{\Sigma}_{K_1} \sim \text{IW}(\boldsymbol{\Lambda}_6, \boldsymbol{\varepsilon}_1),\\
&\boldsymbol{\Sigma}_\xi \sim \text{IW}(\boldsymbol{\Lambda}_7, \boldsymbol{\varepsilon}_2), a_k \sim N(0, \sigma_k^2), b_k \sim \Gamma(\vartheta_5, \vartheta_6), a_{kl} \sim N(0, \sigma_{kl}^2), \eta_l \sim \Gamma(\vartheta_{7l}, \vartheta_{8l}),
\end{aligned} \tag{6}$$

where $a_{kl} = a_{k,l-1} + \eta_l, (k = 1, \ldots, K; l = 2, \ldots, n_k - 1)$; we assume that the Normal (N), Uniform (Unif), Inverse Gamma (IG), exponential (Exp), Inverse Wishart (IW), and Gamma (Γ) prior distributions are mutually independent.

Let $f(\cdot), f(\cdot|\cdot), F(\cdot|\cdot)$ and $\pi(\cdot)$ denote a density function, a conditional density function, a cumulative distribution function and a prior density function, respectively. Since the elements of parameter vector $\boldsymbol{\Theta}$ are assumed to be independent of each other, $\pi(\boldsymbol{\Theta})$ is the product of all prior distributions for each element of $\boldsymbol{\Theta}$. Subsequently, after specifying the joint model for the observed data and the prior distributions for the unknown parameters, we can make Bayesian inference for the parameters based on their posterior distributions. Thus, the joint posterior density of $\boldsymbol{\Theta}$ based on the observed data $\mathcal{D}$ can be given by

$$\begin{aligned}
f(\boldsymbol{\Theta}|\mathcal{D}) \propto \Bigg\{ \prod_{i=1}^{N} \prod_{k=1}^{K} \int \int & f(\boldsymbol{y}_{iK_1}) f(\boldsymbol{w}_i|\boldsymbol{w}_i > 0) f(\eta_i) f(\boldsymbol{r}_{ik}) \\
& \times f(\boldsymbol{\xi}_{ikq}) f(\boldsymbol{u}_i) f(\boldsymbol{\zeta}_{ik}) f(\boldsymbol{\kappa}_{ik}) f(\boldsymbol{\varsigma}_i|\boldsymbol{u}_i) d\boldsymbol{u}_i d\boldsymbol{\xi}_{ikq} \Bigg\} \pi(\boldsymbol{\Theta}).
\end{aligned} \tag{7}$$

Generally, the integrals in (7) are of high dimension and do not have a closed form. Analytic approximations to the integrals may not be sufficiently accurate. Therefore, it is prohibitive to directly calculate the posterior distribution of $\boldsymbol{\Theta}$ based on the observed data. As an alternative, the MCMC procedure can be used to sample population parameters $\boldsymbol{\Theta}$ and random-effects $\boldsymbol{u}_i$ and $\boldsymbol{\xi}_{ikq}$ $(i = 1, \ldots, N; k = 1, \ldots, K)$ from posterior distributions based on (7). We adopt a hybrid MCMC algorithm to speed up calculation. This process is repeated in iterations of MCMC procedure until convergence is reached.

4 | SIMULATION STUDIES

To evaluate performance of the proposed SMJMs with NIM data mechanism (Model JM-NIM) and no missing data mechanism (Model JM-NM) in longitudinal biomarkers and associated Bayesian modeling approach, we conduct the following simulation studies. The design of the simulated data is similar to the real data used in Section 5. We generate the sample size $N = 300$ (150 in both treatment and placebo groups), and assume that each subject has 15 scheduled longitudinal measurements. In the simulation, the NIM data model (3) is adopted and the measurement time points used similar to those in the real data and the true parameter values mimic those obtained in the real data analysis so that all simulated data make biological sense. Specifically, we simulate two continuous (y_{ij1} and y_{ij2}), one binary (y_{ij3}) and one ordinal (y_{ij4} with three categories) longitudinal data. The covariates considered which mimic real data include age with mean 50 years of old generated from the Uniform distribution ranged between 40 and 60, sex (male as reference) generated from Bernoulli(0.8), and treatment variable (1 if treatment and 0 if placebo) generated from Bernoulli(0.5). The true parameter values are selected as follows based on the simple case of SMJM (8)–(10) specified in Section 5 in order to limit intensively computational burden as suggested by the associate editor, where the binary and ordinal outcomes do not consider nonparametric parts in (8) and the disease severity θ_{ij} is specified by the following function $\theta_{ij} = (\beta_{01}x_i + u_{i0}) + (\beta_{10} + \beta_{11}x_i + u_{i1})t_{ij}$. $\boldsymbol{\beta} = (\beta_{01}, \beta_{10}, \beta_{11})^T = (-0.10, 1.35, -0.17)^T$, $\rho = 0.05$, $\sigma_u^2 = 1.25$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T = (-0.10, 0.10, -0.63)^T$, $\boldsymbol{\tau} = (\tau_0, \tau_1)^T = (1.70, 0.80)^T$; $(a_1, b_1)^T = (-0.85, 0.90)^T$, $(a_2, b_2)^T = (1.52, 0.50)^T$, $(a_3, b_3)^T = (-0.39, 2.85)^T$, $(a_{41}, a_{42}, b_4)^T = (-3.78, 1.78, 3.50)^T$; $\boldsymbol{\Sigma}_2 = (\sigma_{kk'}^2)_{2\times 2}$ $(k, k' = 1, 2)$, where $(\sigma_{11}^2, \sigma_{12}^2, \sigma_{22}^2)^T = (0.07, 0.03, 0.1)^T$; $\boldsymbol{\delta} = (\delta_1, \delta_2)^T = (2.10, -0.55)^T$; $\boldsymbol{\gamma}_1 = (\gamma_{11}, \gamma_{21}, \gamma_{31})^T = (1.6, 8.7, 1.8)^T$, $\boldsymbol{\gamma}_2 = (\gamma_{12}, \gamma_{22}, \gamma_{32})^T = (0.2, -0.6, -2.3)^T$; $\boldsymbol{\Sigma}_\xi$ values are mimic to those estimated in the real data. $\boldsymbol{u}_i$ follows $N_2(\mathbf{0}, \boldsymbol{\Sigma}_u)$ and $\boldsymbol{\epsilon}_i$ follows two-dimensional ST distribution with variance–covariance matrix $\boldsymbol{\Sigma}_2$, skewness parameter matrix $\boldsymbol{\Delta}_2 = \text{diag}(\delta_1, \delta_2)$, and the degrees of freedom $\nu = 5.0$. We set $\boldsymbol{\varphi}_1 = (-5.6, 0.04)^T$ and $\boldsymbol{\varphi}_2 = (-3.7, 0.1)^T$ in the NIM data model (3), that is, $P(r_{ijk} = 1|\boldsymbol{\varphi}_k) = \exp(\varphi_{0k} + \varphi_{1k}y_{ijk})/[1 + \exp(\varphi_{0k} + \varphi_{1k}y_{ijk})]$ so that there are approximately 15% missing data rates. To generate the time-to-event data, we use an exponential distribution with

TABLE 1 Simulation results of the true parameter (TP) values, average estimates (EST), bias, mean squared error (MSE)m and coverage probability (CP) based on the 300 sample sizes and 500 simulated datasets for the most interesting parameters.

	JM-NM				JM-NIM			
TP	**EST**	**Bias**	**MSE**	**CP**	**EST**	**Bias**	**MSE**	**CP**
Parameter estimates in longitudinal submodel								
$\beta_{01} = -0.10$	−0.12	−20.0	23.8	0.942	−0.11	−10.0	14.7	0.948
$\beta_{10} = 1.35$	1.30	−3.70	5.78	0.937	1.33	−1.48	2.03	0.944
$\beta_{11} = -0.17$	−0.15	11.8	14.5	0.927	−0.16	5.88	7.83	0.936
$\rho = 0.05$	0.06	20.0	23.1	0.963	0.06	20.0	22.9	0.958
$\sigma_u^2 = 1.25$	1.31	4.80	5.69	0.938	1.27	1.60	2.27	0.940
$\delta_1 = 2.10$	2.31	10.0	12.7	0.944	2.17	3.33	5.56	0.952
$\delta_2 = -0.55$	−0.70	−27.3	30.4	0.921	−0.59	−7.27	9.90	0.938
$a_1 = -0.85$	−0.80	5.88	9.65	0.939	−0.84	1.18	4.94	0.947
$b_1 = 0.90$	0.94	4.44	8.54	0.929	0.92	2.22	6.31	0.938
$a_2 = 1.52$	1.47	−3.29	7.78	0.937	1.49	−1.97	6.45	0.940
$b_2 = 0.50$	0.43	−14.0	17.1	0.931	0.47	−6.01	9.05	0.945
$a_3 = -0.39$	−0.44	−12.8	16.0	0.931	−0.41	−5.13	8.36	0.951
$b_3 = 2.85$	2.72	−4.56	8.66	0.925	2.80	−1.75	5.86	0.939
$a_{41} = -3.78$	−3.66	3.17	6.18	0.940	−3.75	0.79	3.80	0.949
$a_{42} = 1.78$	1.69	−5.06	8.64	0.942	1.75	−1.68	5.27	0.953
$b_4 = 3.50$	3.38	−3.43	7.12	0.933	3.40	−2.86	6.55	0.944
Parameter estimates in survival submodel								
$\alpha_1 = -0.10$	−0.13	−30.0	37.5	0.938	−0.11	−10.0	15.9	0.942
$\alpha_2 = 0.10$	0.12	20.0	24.4	0.942	0.11	10.0	14.2	0.951
$\alpha_3 = -0.63$	−0.60	4.76	6.88	0.922	−0.62	1.59	2.97	0.940
$\tau_0 = 1.70$	1.73	1.76	2.98	0.919	1.71	0.59	1.25	0.938
$\tau_1 = 0.80$	0.83	3.75	5.09	0.937	0.81	1.25	1.68	0.948

mean equal to 0.1 to generate censoring time. The simulation is repeated 500 times (ie, 500 simulated datasets), and the resulting estimates are averaged.

The SMJM modeling is implemented using the same MCMC strategies as those in real application below and the prior distributions considered are all close to noninformative; that is, prior distributions with large variances are considered. Thus, we expect the results to be somewhat robust with respect to prior distributions. In the simulation studies, we compare Model JM-NIM with Model JM-NM to assess how the missing data influences modeling results.

Table 1 summarizes simulation results that include the true parameter (TP) values, average estimates (EST) of some important fixed-effects, associated the bias (quantified by relative percent bias $= 100 \times (\text{EST} - \text{TP})/|\text{TP}|$), mean squared error (MSE)(quantified by relative percent square root of MSE $= 100 \times \sqrt{\text{MSE}}/|\text{TP}|$) and coverage probabilities (CP) of 95% equal-tail credible intervals.

To compare the numerical results of Models JM-NIM and JM-NM considered in the simulation studies, it is of interest to see that the estimated biases for β_{01} and β_{10} are negative, indicating that these parameters (the estimated baseline group difference and disease progression rate for placebo group) are underestimated, while estimated bias for β_{11} is positive, suggesting that the disease progression rate in treatment group is overestimated. The average estimates of skewness parameters δ_1 for SB and δ_2 for SA in both Models indicate a departure from (symmetric) normal distribution for both SB and SA longitudinal measurements. Moreover, both models provide reasonable estimates to the difficulty and discriminating parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. For the parameters in the survival submodel, all parameters in Model JM-NIM are consistently closer the true parameter values than the corresponding counterparts in Model JM-NM. Because the

parameters τ_0 and τ_1 are set to be positive, the patients with worse baseline disease severity (larger u_{i0}) and faster disease progression rate (larger u_{i1}) tend to have a terminal event earlier. These results suggest that Model JM-NIM can recover the true values better than Model JM-NM in the presence of dependent event.

In summary, it is seen that, when missing data are not at random, Model JM-NIM performs well in comparison with Model JM-NM in terms of estimate, bias, MSE and CP, suggesting that it is important to take NIM data into account in our SMJM in order to achieve less biased and more accurate estimates as well as closer CPs to nominal level in the presence of missing data in longitudinal measurements. Note that both Models JM-NIM and JM-NM provide reasonable estimates to the difficulty and discriminating parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, but Model JM-NIM generally provides more accurate estimates than Model JM-NM. We also expand simulation studies using larger sample sizes ($N = 500$) and larger number of replications (ie, 800 simulated datasets) for further simulation studies, and the results (please refer to Table S1 in Section 2 of Appendix S1 in detail) indicate that the simulation results are quite similar to those reported in Table 1.

5 | APPLICATION

5.1 | PBC clinical study and data

We examine the effect of multiple longitudinal biomarkers on the prognosis for patients with PBC using data collected by the Mayo Clinic.[1,49] The PBC is a chronic, fatal, but rare liver disease characterized by inflammatory destruction of the small bile ducts within the liver, which eventually leads to cirrhosis of the liver, followed by death. If the PBC is not treated or reaches an advanced stage, it can lead to several major complications, including mortality. The PBC is a chronic liver disease with no effective treatment other than liver transplantation.[50]

Various longitudinal biomarkers such as SB, SA, HM, and HS data were collected, and interest is on examining whether these longitudinal biomarkers relate to the natural history of disease. This dataset has been widely analyzed using various statistical modeling methods including joint modelling approach.[2,51,52] The data were collected to examine the progress of PBC patients. A total of 424 PBC patients met the eligibility criteria for the randomized placebo controlled trail of the drug D-penicillamine, referred to Mayo Clinic during the ten-year interval, but 112 patients did not participate in the clinical trail. Thus, 312 patients, who had a baseline measurement and were followed longitudinally at 6 months and at yearly intervals thereafter, participated in the randomized trail and were randomized to receive D-penicillamine ($n = 158$) or placebo ($n = 154$). The PBC data are publicly accessible at website: https://stat.ethz.ch/R-manual/R-devel/library/survival/html/pbc.html.

Patients with PBC typically have abnormalities in several blood tests; hence, during follow-up several biomarkers associated with liver function were serially recorded for these patients. The original clinical protocol for these patients specified visits at 6 months, 1 year, and annually thereafter. This is an ideal dataset to illustrate the various features of the MLIRT-based SMJM. The data set includes clinical, biochemical, and demographic risk factors for each patient. Particularly, in our application, we consider the following data variables. (i) Demographic factors (covariates) are age and sex of patients; biochemical factor is drug (D-penicillin and placebo group). (ii) Longitudinal data of mixed types: HM (liver growth status, presence of hepatomogaly 0 = No, 1 = Yes) which is binary outcome, HS (1≤ stage 2, 2 = stage 3 and 3 = stage 4) which is ordinal outcome, and the continuous outcomes SB (mg/dL) and SA (mg/dL) values were taken as biochemical properties. As illustrated in Section 1, the repeated SB (mg/dL) and SA (mg/dL) measurements appeared departures from normality (ie, skewed distribution) and the trajectory profiles for SB and SA longitudinal biomarkers depicted irregular nonlinear time trends (see Figure 1). (iii) Survival data: since the patients were started to be followed, the survival time (year) were taken as the period until death. Here the failure is considered as death and other patients are expressed as censored. At the end of study, 140 of the 312 patients (44.9%) died and 172 (55.1%) was censored.

The demographic and clinical characteristics of patients of the PBC dataset by treatment groups (placebo vs D-penicillin) are summarized in Table 2. Of the 312 patients, the mean ages at baseline for D-penicillin and placebo groups are 51.42 ± 11.01 years and 48.58 ± 9.96 years, respectively. The baselines values of SB for D-penicillin and placebo groups are 0.53 ± 0.95 g/dL and 0.61 ± 1.10 mg/dL, respectively. The baselines values of SA for D-penicillin and placebo groups are 3.52 ± 0.44 mg/dL and 3.52 ± 0.40 mg/dL, respectively. A total of 1945 longitudinal measurements are collected from 312 subjects. The numbers of repeated measurement per patient are ranged between 2 and 15 with the mean number of follow-up being 8.

TABLE 2 Descriptive statistics of the demographic and clinical characteristics.

Variables	D-penicillin	Placebo	Total
Number of patients	158	154	312
Gender			
Female	137 (86.71%)	139 (90.26%)	276 (88.46%)
Male	21 (13.29%)	15 (9.74%)	36 (11.54%)
Survival status			
Death	71 (44.94%)	69 (44.81%)	140 (44.87%)
Alive	87 (55.06%)	85 (55.19%)	172 (55.13%)
HM			
Yes	73 (46.20%)	87 (56.49%)	160 (51.28%)
No	85 (53.80%)	67 (43.51%)	152 (48.72%)
HS			
1	47 (29.75%)	36 (23.38%)	83 (26.60%)
2	53 (33.54%)	49 (31.82%)	102 (32.69%)
3	58 (36.71%)	69 (44.80%)	127 (40.71%)
Baseline age			
Mean (SD)	51.42 (11.01)	48.58 (9.96)	50.02 (10.58)
Max, Min	78.44, 26.28	74.53, 30.57	78.44, 26.28
Baseline SB			
Mean (SD)	0.53 (0.95)	0.61 (1.10)	0.57 (1.03)
Max, Min	2.99, −1.20	3.33, −1.20	3.33, −1.20
Baseline SA			
Mean (SD)	3.52 (0.44)	3.52 (0.40)	3.52 (0.42)
Max, Min	4.64, 2.10	4.38, 1.96	4.64, 1.96

Note: A total of 1945 longitudinal measurements are collected from 312 subjects. The numbers of repeated measurements per patient are ranged between 1 and 15 with the mean number of follow-up being 8. The mean and standard deviation(SD) for continuous variables and proportion for categorical variables.
Abbreviations: HM, hepatomegaly; HS, histologic stage; SA, serum albumin; SB, serum bilirubin.

When patient's repeated measurements are not observed during a study period described above, the values are considered as missing and their missing rates were 11% (HM), 16% (HS), 14% (SB), and 14% (SA), respectively. In the PBC clinical study, some patients may drop out early due to death and other reasons. In practice, patient's dropouts may be related the value being missing. Because the follow-up visits are pre-designed, there could be some reasons to drop out or miss scheduled visits that we just do not know. So missing data for patient's longitudinal measurements may be nonignorable, and a possible missing data model should be incorporated in the analysis to avoid biased results.

5.2 | Specific SMJM and implementation

To apply the extended MLIRT submodel for analyzing the PBC dataset, we focus on a specific MLIRT-based SMJM which considers continuous variables SB (y_{ij1}) and SA (y_{ij2}), binary HM variable (y_{ij3}) and ordinal HS variable (y_{ij4}) ($i = 1, \ldots, N; j = 1, \ldots, J_i; K = 1, \ldots, 4$) as multivariate longitudinal biomarkers, and time to death as the event endpoint. Based on the general SMJM discussed in Section 2, for the nonparametric part of the longitudinal submodel in (1), we consider natural cubic spline bases with the percentile-based knots in the nonparametric part. The optimal degree of regression spline p and numbers of knots q can be also determined according to AIC/BIC values.[42] Here, the AIC/BIC values were evaluated for various models with $(p, q) = \{(1,1), (2,1), (2,2), (3,1), (3,2), (3,3)\}$. Among them, we

found that the model with $(p, q) = (3, 3)$ has the smallest AIC/BIC values. Thus, the specific SMJM can be constructed as follows.

$$\begin{aligned}
&y_{ijk} = a_k + b_k\theta_{ij} + \boldsymbol{\Psi}_3(t_{ij})^T\boldsymbol{\gamma}_{k3} + \boldsymbol{\Phi}_3(t_{ij})^T\boldsymbol{\xi}_{ik3} + \epsilon_{ijk}(k = 1, 2),\\
&\text{logit}\{p(y_{ij3} = 1|\theta_{ij})\} = a_3 + b_3\theta_{ij} + \boldsymbol{\Psi}_3(t_{ij})^T\boldsymbol{\gamma}_{k3} + \boldsymbol{\Phi}_3(t_{ij})^T\boldsymbol{\xi}_{ik3},\\
&\text{logit}\{p(y_{ij4} \le l|\theta_{ij})\} = a_{4l} - b_4\theta_{ij} + \boldsymbol{\Psi}_3(t_{ij})^T\boldsymbol{\gamma}_{k3} + \boldsymbol{\Phi}_3(t_{ij})^T\boldsymbol{\xi}_{ik3}, \text{with}(l = 1, 2),\\
&f(r_{ijk}|y_{ijk}) = [p(r_{ijk} = 1|\boldsymbol{\varphi}_k)]^{r_{ijk}}[1 - p(r_{ijk} = 1|\boldsymbol{\varphi}_k)]^{1-r_{ijk}},
\end{aligned} \tag{8}$$

where $\text{logit}[p(r_{ijk} = 1|\boldsymbol{\varphi}_k)] = \varphi_{0k} + \varphi_{1k}y_{ijk}$, $\boldsymbol{\varphi}_k = (\varphi_{0k}, \varphi_{1k})^T$, the random error for two continuous variables $\boldsymbol{\epsilon}_i = (\boldsymbol{\epsilon}_{i1}^T, \boldsymbol{\epsilon}_{i2}^T)^T$ follows a multivariate ST distribution $ST_{2J_i,\nu}\big(-J(\nu)[\boldsymbol{\Delta}_2 \otimes \mathbf{1}_{J_i}], \boldsymbol{\Sigma}_2 \otimes \boldsymbol{I}_{J_i}, \boldsymbol{\Delta}_2 \otimes \boldsymbol{I}_{J_i}\big)$[14,15] with unknown variance–covariance matrix $\boldsymbol{\Sigma}_2 = (\sigma^2_{kk'})_{2\times2}$, degrees of freedom ν and skewness parameter matrix $\boldsymbol{\Delta}_2 = \text{diag}(\delta_1, \delta_2)$, where the vector of skewness parameters $\boldsymbol{\delta}_2 = (\delta_1, \delta_2)^T$. $\boldsymbol{\gamma}_{k3} = (\gamma_{1k}, \gamma_{2k}, \gamma_{3k})^T$ $(k = 1, 2)$ and $\boldsymbol{\xi}_{ik3} \overset{\text{iid}}{\sim} N_3(\mathbf{0}, \boldsymbol{\Sigma}_\xi)$ with $\boldsymbol{\Sigma}_\xi$ being unrestricted covariance matrix. The basis functions $\boldsymbol{\Psi}_3(t) = \{\psi_0(t), \psi_1(t), \psi_2(t)\}^T$ and $\boldsymbol{\Phi}_3(t) = \{\phi_0(t), \phi_1(t), \phi_2(t)\}^T$, respectively.

In the second-level structural multilevel model for the latent disease severity θ_{ij}, we use the treatment indicator variable x_i as only covariate which was adopted in most of the literature.[19,24,28,31] Thus, θ_{ij} is regressed on biochemical factor x_i which is an indicator variable of treatment (1 if treatment, and 0 otherwise) and random-effects $\boldsymbol{u}_i = (u_{i0}, u_{i1})^T$.

$$\theta_{ij} = (\beta_{01}x_i + u_{i0}) + (\beta_{10} + \beta_{11}x_i + u_{i1})t_{ij} + (\beta_{20} + \beta_{21}x_i)t_{ij}^2, \tag{9}$$

where β_{01} is the group difference of baseline disease severity, β_{10} and $\beta_{10} + \beta_{11}$ are the progressive disease rates for the placebo and treatment patients, respectively and β_{20} and $\beta_{20} + \beta_{21}$ are the concavities of disease severity for the placebo and treatment patients, respectively. Note that no fixed intercept β_{00} in the equation of θ_{ij} (9) is an identifiability constraint induced by the creation of the latent variable. The significantly negative β_{11} value indicates that the treatment is efficacious in slowing down the disease progression.

Under Cox proportional hazards model, the hazard of having an event at time t_i is

$$\lambda(t_i|\boldsymbol{X}_i, \boldsymbol{u}_i) = \lambda_0(t_i)\exp(\boldsymbol{X}_i^T\boldsymbol{\alpha} + \boldsymbol{u}_i^T\boldsymbol{\tau}), \tag{10}$$

where $\boldsymbol{\tau} = (\tau_0, \tau_1)^T$ with τ_0 and τ_1 measuring the association between the two submodels, the vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T$ is the coefficient parameters corresponding to the risk factor vector $\boldsymbol{X}_i$ which includes treatment indicator variable, age at baseline, gender (male as reference group).

To carry out a Bayesian inference, we need to specify the values of the hyper-parameters in the prior distributions. We assume weakly informative prior distributions for all the parameters. In particular, (i) the fixed-effects are taken to be independent normal distribution $N(0, 100)$ for each element of the population parameter vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_{k3}(k = 1, 2)$, $\boldsymbol{\delta}_2$ and $\boldsymbol{\tau}$, respectively. (ii) For the variance-covariance matrices $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_\xi$, we assume the inverse Wishart distributions $IW(\boldsymbol{\Lambda}_6, 3)$ and $IW(\boldsymbol{\Lambda}_7, 4)$, respectively, where $\boldsymbol{\Lambda}_6$ and $\boldsymbol{\Lambda}_7$ are the matrices with diagonal elements being 0.01 and off-diagonal elements being 0. (iii) We use the prior distribution $\rho \sim Unif(-1, 1)$ and a noninformative inverse Gamma prior distribution $IG(0.01, 0.01)$, which has mean 1 and variance 100, for variance parameter σ_u^2. (iv) The degrees of freedom parameter $\nu \sim Exp(0.1)I(\nu > 2)$, a truncated exponential distribution. (v) a_k (k = 1,2,3) and a_{41} follow normal distribution $N(0, 100)$, respectively; b_k $(k = 1, \ldots, 4)$ and η_2 follow $\Gamma(0.01, 0.01)$.

We now conduct the following scenarios to analyze the PBC dataset to illustrate our models and method. We firstly estimate the model parameters using the joint modeling (JM) approach, proposed in Section 3 based on the SMJM (8)–(10), with considering NIM data model (3) (denoted by Model JM-NIM), to compare with those in which no missing data mechanism in longitudinal biomarkers is considered (denoted by Model JM-NM). That is, the Model JM-NM uses only the observed longitudinal values y_{ijk} in the SMJM (8)–(10) ignoring the missing data for modeling. We note that the NIM data model (3) is not testable based on the observed data, so we will consider alternative missing data models to perform sensitivity analysis (Section 5.3.2 in detail below).

The SMJM modeling is implemented using MCMC procedure and the program codes have been posted in Section 4 of Appendix S1. When the MCMC procedure is applied to the actual data, the convergence of the generated samples is assessed using standard tools such as trace plots and Gelman-Rubin (GR) diagnostics[53] as depicted in Figure S1 in Section 3 of Appendix S1.

5.3 | Modeling results

5.3.1 | Comparison of modeling results.

The Bayesian joint modeling approach is used to fit PBC longitudinal measurements of mixed types data and time to death outcome. Table 3 presents the posterior mean (PM), the corresponding standard deviation (SD) and 95% credible interval (CI) for fixed-effects (population) parameters based on the two specific SMJMs (ie, JM-NIM and JM-NM), respectively, but some estimates of the less important or nuisance parameters such as $\boldsymbol{\gamma}_{k3}$, $\boldsymbol{\psi}_k$, and $\boldsymbol{\Sigma}_\xi$ are not reported to save space here. The following findings are obtained for estimated results of population parameters.

It can be seen from Table 3 that, for the longitudinal MLIRT submodel, Model JM-NIM and Model JM-NM give different parameter estimates, although the same set of parameters are identified for significance by both Models. For instance, Model JM-NIM suggests that D-penicillin is associated with 0.132 unit decrease [β_{01}, 95% CI: (−0.540, 0.322)] in disease severity, comparing with the placebo, while it is associated with 0.137 unit decrease [β_{01}, 95% CI: (−0.591, 0.244)] in disease severity in Model JM-NM. The results from Model JM-NIM indicate that the placebo patients show significant disease progression cross time at the rate of 5.278 units per year [β_{10}, 95% CI: (3.979, 6.520)], whereas its counterpart is 5.981 [95% CI: (4.582, 7.492)] in Model JM-NM. In comparison, Model JM-NIM suggests that the D-penicillin patients have disease progression rate of 4.614 units per year [$\beta_{10} + \beta_{11}$, 95% CI: (2.004, 7.364)] with insignificant D-penicillin treatment effect of slowing down the disease progression rate by −0.664 per year [β_{11}, 95% CI: (−1.975, 0.844)]; the similar conclusions can be found in Model JM-NM.

For the survival submodel, it can be seen from Table 3 that, in comparison with placebo, D-penicillin decreases the hazard of death by 10% [$1 - \exp(\alpha_1) = 0.10$, where $\alpha_1 = -0.103$ with 95% CI being (−0.577, 0.437)] for Model JM-NIM and by 6% [$1 - \exp(\alpha_1) = 0.06$, where $\alpha_1 = -0.065$ with 95% CI being (−0.513, 0.375)] for Model JM-NM. The insignificant treatment effect is consistent with those reported in the literature.[2,11] There is a 2.43% increase (HR = 1.0243, 95% CI:(1.004,1.046)) in hazard for death relative to a 1-year older in age for Model JM-NIM which is consistent with the finding by Dil and Karasoy;[2] the comparable result can be found from Model JM-NM. Female patients decrease the hazard of death by 46.58% [$1 - \exp(\alpha_3) = 0.4658$, where $\alpha_3 = -0.627$ with 95% CI being (−1.147, −0.060)] in comparison with male patients in Model JM-NIM, while female patients decrease the hazard of death by 47.74% [$1 - \exp(\alpha_3) = 0.4774$, where $\alpha_3 = -0.649$ with 95% CI being (−1.165, −0.114)] in comparison with male patients in Model JM-NM. This finding is consistent with that in publications.[2-4]

We observe that τ_0 and τ_1 are positive and significantly different from zero, [$\tau_0 = 0.490$ with 95% CI being (0.309,0.677), and $\tau_1 = 0.216$ with 95% CI being (0.129,0.344)] in Model JM-NIM, suggesting that the patients with worse baseline disease severity (larger u_{i0}) and faster disease progression rate (larger u_{i1}) tend to have higher hazard of death and vice versa; the similar conclusions can be made from Model JM-NM. Model JM-NIM and Model JM-NM give quite different estimates for the majority of outcome-specific parameters for the continuous SB and SA biomarkers as well as the binary HM and ordinal HS biomarkers (**a** and **b**; see Table 3 for details).

We also find from Table 3 that the estimated skewness parameters of the continuous SB outcome (δ_1) and the continuous SA outcome (δ_2) from Model JM-NIM are $\delta_1 = 1.429$ with 95% CI (1.303, 1.556) and $\delta_2 = -0.055$ with 95% CI (−0.077, −0.034), respectively; the corresponding results from Model JM-NM are $\delta_1 = 1.850$ with 95% CI (0.897, 2.199) and $\delta_2 = -0.090$ with 95% CI (−0.119, −0.061). These findings suggest that there is a significantly positive skewness in SB outcome and negative skewness in SA outcome, confirming the fact that the distribution of the repeated SB and SA measurements is skewed which are consistent with those displayed graphically in Figure 1 (top panel). Thus, it is recommended to consider the model error with skewed distribution in the modeling of both SB and SA data.

It is seen from Table 3 that the both DIC = 9721.9 and the posterior mean of the deviance ($\overline{D} = 7319.4$) in Model JM-NIM is smaller than its counterparts (DIC = 10 479.7 and $\overline{D} = 7922.0$) in Model JM-NM. Therefore, Model JM-NIM performs better than Model JM-NM, suggesting that it is important to take NIM data mechanism in the longitudinal data into account when collected data are "incompletely" measured in order to achieve more reliable results, particularly when the missing observations of longitudinal data are informative (nonignorable). With these findings, we conduct sensitivity analysis in conjunction with different NIM data models (Models I–IV) specified in Section 5.3.2 below. Further, we report some of interesting results in Section 5.3.3 for the better Model JM-NIM.

TABLE 3 Summary of estimated posterior mean (PM), standard deviation (SD) of population (fixed-effects) parameters and the corresponding 95% equal-tail credible interval (CI) as well as DIC value and the posterior mean of the deviance $\overline{D}$.

	JM-NM			JM-NIM		
Parameter	**PM**	**SD**	**95% CI**	**PM**	**SD**	**95% CI**
	Parameter estimates in longitudinal submodel					
β_{01}	−0.137	0.212	(−0.591, 0.244)	−0.132	0.210	(−0.540, 0.322)
β_{10}	5.981	0.790	(4.582,7.492)	5.278	0.696	(3.979,6.520)
β_{11}	−0.892	0.805	(−2.489, 0.677)	−0.664	0.703	(−1.975, 0.844)
β_{20}	−1.378	0.778	(−2.790, 0.464)	−0.767	0.665	(−1.913, 0.770)
β_{21}	0.381	1.023	(−1.792, 2.367)	0.297	0.895	(−1.701, 1.520)
ρ	0.223	0.150	(0.078, 0.511)	0.212	0.185	(0.142, 0.499)
σ_u^2	4.160	1.743	(2.441, 5.500)	3.820	1.122	(1.892, 4.180)
ν	4.461	0.078	(3.375, 4.773)	3.021	0.022	(3.001, 3.082)
δ_1	1.850	0.077	(0.897, 2.199)	1.429	0.064	(1.303, 1.556)
δ_2	−0.090	0.015	(−0.119, −0.061)	−0.055	0.011	(−0.077, −0.034)
For SB and SA						
a_1	−1.150	0.097	(−1.339, −0.959)	−1.385	0.260	(−1.819, −0.948)
b_1	0.684	0.078	(0.545, 0.831)	0.726	0.079	(0.592, 0.895)
a_2	1.042	1.481	(−1.919, 2.801)	0.431	0.461	(−0.671, 0.953)
b_2	0.330	0.099	(0.166, 0.567)	0.174	0.055	(0.076, 0.292)
σ_{11}^2	0.078	0.020	(0.044, 0.125)	0.030	0.005	(0.021, 0.040)
σ_{12}^2	0.025	0.014	(−0.002, 0.052)	0.006	0.005	(−0.004, 0.015)
σ_{22}^2	0.101	0.004	(0.094, 0.109)	0.068	0.004	(0.059, 0.076)
For HM						
a_3	3.242	1.727	(0.291, 6.079)	1.379	1.935	(0.579, 4.276)
b_3	0.126	0.016	(0.096, 0.157)	0.120	0.016	(0.092, 0.152)
For HS						
a_{41}	0.854	0.327	(0.243, 1.539)	2.666	1.468	(0.412, 4.814)
a_{42}	6.394	0.501	(5.402, 7.371)	8.410	1.542	(5.911, 9.990)
b_4	4.116	0.515	(3.291, 5.194)	4.714	0.470	(3.878, 5.762)
	Parameter estimates in survival submodel					
α_1	−0.065	0.234	(−0.513, 0.375)	−0.103	0.255	(−0.577, 0.437)
α_2	0.023	0.011	(0.003, 0.043)	0.024	0.010	(0.006, 0.044)
α_3	−0.649	0.266	(−1.165, −0.114)	−0.627	0.278	(−1.147, −0.060)
τ_0	0.463	0.094	(0.294, 0.665)	0.490	0.094	(0.309, 0.677)
τ_1	0.227	0.050	(0.140, 0.329)	0.216	0.054	(0.129, 0.344)
DIC		10 479.7			9721.9	
$\overline{D}$		7922.0			7319.4	

5.3.2 | Sensitivity for NIM data models

It is important to evaluate the sensitivity of parameter estimates to various plausible NIM data models. Subject-area knowledge may help us to determine alternative missing data models. It is conceivable that missing values may be related to current and previous measurements and/or missingness may link to the random-effects that characterize the true response processes in practice. Note that we should avoid building too complicated missing data model since the parameters may become nonidentifiable. Thus, to conduct sensitivity analysis we consider SMJM (8) to (10) for the longitudinal responses y_{ijk}, $(k = 1, 2, 3, 4)$ in conjunction with the several alternative missing data models as follows.

- Model I: SMJM with $\text{logit}[p(r_{ijk} = 1|\boldsymbol{\varphi}_k)] = \varphi_{0k} + \varphi_{1k}y_{ijk}$;
- Model II: SMJM with $\text{logit}[p(r_{ijk} = 1|\boldsymbol{\varphi}_k)] = \varphi_{0k} + \varphi_{1k}r_{i,j-1,k} + \varphi_{2k}y_{ijk}$;
- Model III:SMJM with $\text{logit}[p(r_{ijk} = 1|\boldsymbol{\varphi}_k)] = \varphi_{0k} + \varphi_{1k}y_{i,j-1,k} + \varphi_{2k}y_{ijk}$.
- Model IV: SMJM with $\text{logit}[p(r_{ijk} = 1|\boldsymbol{\varphi}_k)] = \varphi_{0k} + \varphi_{1k}y_{ijk} + \varphi_{2k}u_{i0} + \varphi_{3k}u_{i1}$.

TABLE 4 Model comparison: deviance information criterion (DIC), and estimated population posterior means of parameters selected.

Model	DIC	β_{01}	β_{10}	β_{11}	β_{20}	β_{21}	τ_0	τ_1	ρ	σ_u^2	δ_1	δ_2
Model I	9721.9	−0.132	5.278	−0.664	−0.767	−0.078	0.490	0.216	0.212	3.820	1.429	−0.055
Model II	9758.4	−0.140	5.301	−0.648	−0.756	−0.080	0.491	0.218	0.206	3.793	1.427	−0.053
Model III	9821.3	−0.137	5.373	−0.673	−0.749	−0.069	0.487	0.220	0.217	3.851	1.431	−0.051
Model IV	9860.7	−0.143	5.343	−0.658	−0.770	−0.073	0.489	0.221	0.219	3.795	1.426	−0.049

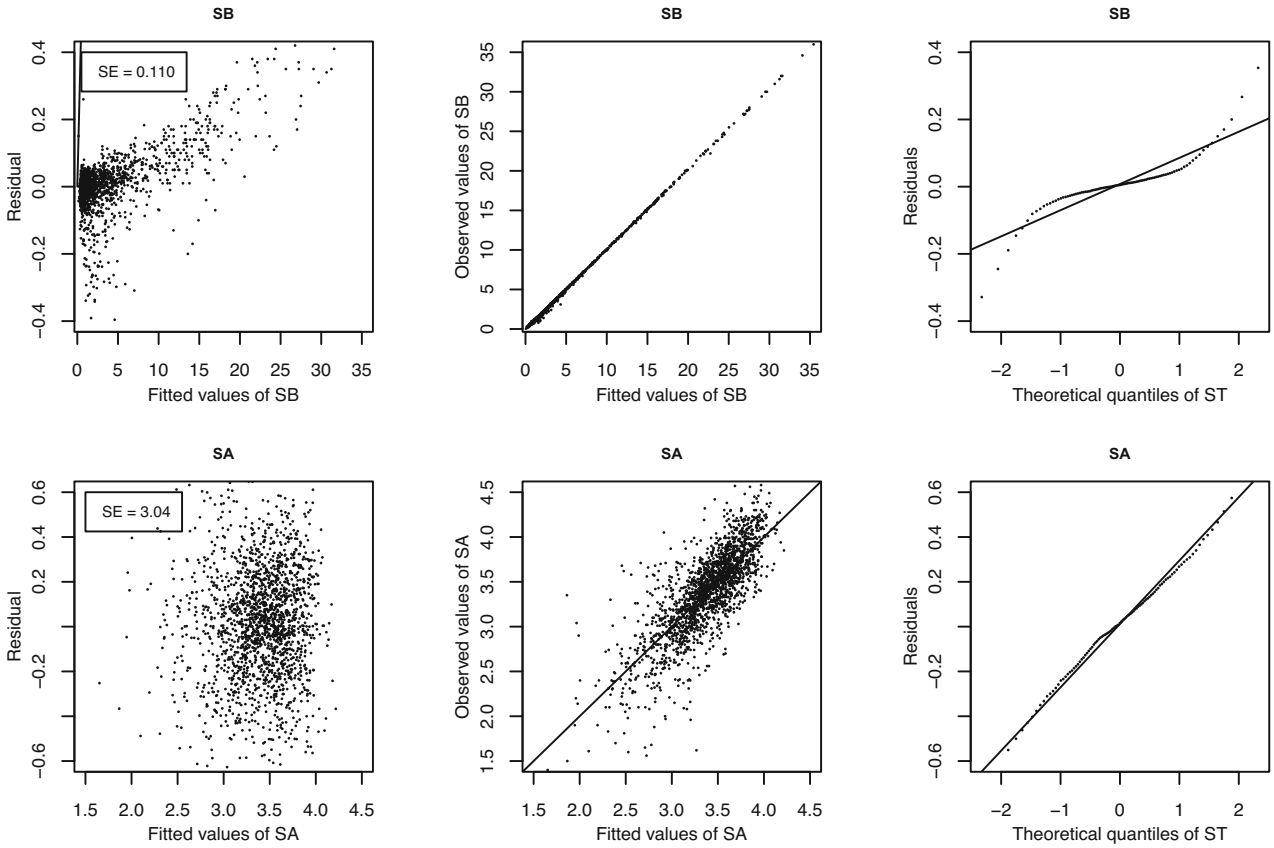

FIGURE 2 Goodness of fit: (A) Scatter plots of residuals versus fitted values of serum bilirubin (SB) and serum albumin (SA) (left panel), where SE denotes standard error of the residual. (B) Observed values vs fitted values of SB and SA (middle panel). (C) skew-t Q-Q plots with line (right panel).

Model I depends on the current longitudinal values being missing which is SMJM discussed in Section 5.2; in Model II, the missingness is related to the previous missing status and current longitudinal value being missing; Model III incorporates previous and current longitudinal measurements in the missing data mechanism; in Model IV, the correlation is given by inducing combination of current measurement and random-effects which characterize longitudinal disease severity.

Table 4 provides the results of model comparison including DIC and estimated population posterior means of parameters selected, respectively. We can see that all models have very similar posterior mean estimates of the parameters selected. We also find that the resulting estimates of other parameters in SMJM (8) to (10) with NIM data models are all similar (results not shown). This suggests that the parameter estimation may be robust against the non-ignorable missing data models and, thus, the parameter estimates in Model JM-NIM presented in Table 3 may be reliable.

It is seen that adding previous missing status (Model II) or previous longitudinal measurement (Model III) gives larger the DIC value in comparison with Model I; this may be explained by the fact that the current longitudinal measurement may be most predictive for missing data since longitudinal measurements are relatively sparse in this study and early values may not have much influence on subjects' missing values. Adding random-effects in Model IV (DIC = 9860.7)

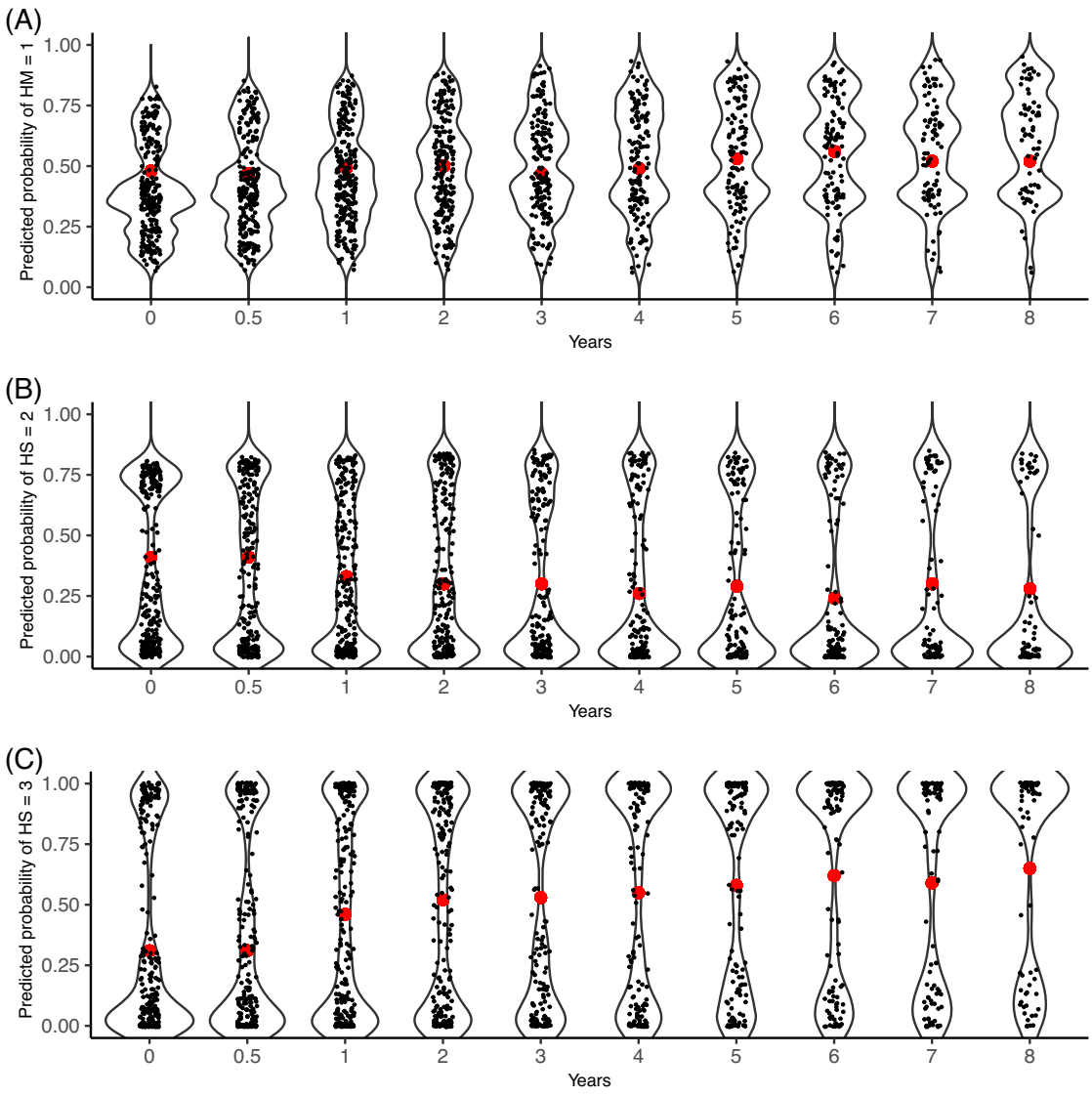

FIGURE 3 Goodness of fit by comparing the observed proportions (red solid circles) of each value at each visiting time with the predicted probabilities: (A) binary hepatomegaly = 1 (top panel); (B) ordinal histologic stage (HS) = 2 (middle panel); (C) ordinal HS = 3 (bottom panel).

provides largest DIC value in comparison with Models I, II, and III (the differences in DIC values are 138.8, 102.3 and 39.4, respectively). Thus, Model I (ie, Model JM-NIM) is the favored model.

5.3.3 | Model fitting results based on Model I.

To assess the goodness-of-fit to the observed SB and SA data for the proposed joint modeling approach, the diagnosis plots including the residuals vs the fitted values (left panel), the observed values versus the fitted values (center panel), and ST Q-Q plots (right panel) are presented in Figure 2. The residual plots (left panel) indicate that no major systematic patterns are found and in general the model fit the observed data well. Further, both the observed values versus the fitted values and ST Q-Q plots confirm that the model fitting performs reasonably well.

The fit of the binary HM (liver growth status: 0 = No, 1 = Yes) and ordinal HS (1≤ stage 2, 2 = stage 3 and 3 = stage 4) markers is presented in Figure 3. Due to the limited space and smaller numbers of patients being followed after year 8, Figure 3 presents the fitting results of the first 8 years of study period. To interpret fitting results, we take panel (a) as an example. At each follow-up time, the predicted probability of HM (1 = Yes) is plotted on the Y-axis with distribution summarized using violin shapes. The red solid circle denotes the observed proportion of HM with liver growth at the corresponding time of visit. The predicted probabilities capture the main pattern in the changes of the observed proportions across times of visit. Similarly, the fitting results of panels (b) and (c) for HS with stage 3 and stage 4 can be interpreted, respectively.

6 | CONCLUDING DISCUSSION

With an increased focus on complex longitudinal studies conducted that take repeated measures over time in an effort to evaluate a patient's health status for some events, an MLIRT-based SMJM modeling approach is powerful tool to fit these complicated longitudinal-survival models with a variety of data features ranging from evaluating nonnormality, missing values, correlated multivariate longitudinal measures and others, along with uncertainty about the distributional assumptions of model errors in practice. In clinical and medical studies, it is quite common to have various longitudinal outcomes with mixed types (continuous, binary, and ordinal) subject to dependent time-to-event. Previous work of joint modeling for this type of data has been predominantly focused on a single type of longitudinal data setting accounting for the dependent censoring. In this article, we have extended the previous work by proposing an SMJM modeling framework that consists of an extended MLIRT longitudinal submodel for the multivariate longitudinal outcomes with mixed types (continuous, binary, and ordinal) while modeling the missing data process and the survival submodel for time-to-event endpoint. Two submodels are linked together via shared random-effects representing the subject-specific baseline disease severity and disease progression rate, respectively. The data analysis using a joint model with random-effects variables that linked the longitudinal submodel and survival submodel together resulted in more accurate estimates that were evidenced in the simulation studies.[5,9] The Bayesian modeling approach adopted in this article provided not only accurate parameter estimates, but also obtained the subject-specific PBC disease severity estimation. A benefit of directly estimating parameters of key interest in our SMJM, including the (latent) disease severity for all individuals at each visit through the baseline disease severity and disease progression rates, is that these subject-specific quantities can be served as surrogate indicators of characteristic of longitudinal data and, in turn, incorporate into the survival submodel as covariates to assess the effects of longitudinal profiles on the risk of time-to-event. This kind of joint modeling approach is important in many statistical application areas, allowing accurate inference of parameters while adjusting for nonnormality, incompleteness and correlation in the multivariate longitudinal data setting. Although this article is motivated by a PBC study, the basic concepts of the developed Bayesian SMJM modeling approach have generally broader applications whenever the two different sources of dependence among longitudinal measures over time and between different variables are presented and the relevant technical specifications are met.

We have found that the treatment D-penicillin is insignificant in slowing the PBC disease progression. Moreover, we have identified a significant positive correlation between the multiple longitudinal outcomes and the time to death, in addition to the positive significant correlation between the baseline disease severity and disease progression rate. We estimated subject-specific disease severities and disease progression rates for all patients at each visit. The proposed joint models have the capability to accommodate hierarchically structured data, to estimate latent disease severity at different levels, and to investigate the relationships between disease severity and predictors at different levels. The proposed

joint models can be easily implemented using the publicly available BUGS language and can be readily accessible to and extended by applied researchers. Further, our results suggested that it is important to incorporate nonignorable missing mechanism and nonnormality into account in the SMJM in order to achieve less biased and more accurate estimates in the presence of missing data in longitudinal observations and nonnormal feature in SB and SA measurements.

The following notes related to this research should be made. (i) Identifiability is an important but difficult problem, when a lot of model parameters must be estimated simultaneously. Fortunately, in this article, the MCMC algorithm converged without any problems and we did not observe potential identifiability issues. (ii) Our modeling framework provides great flexibility for extension. This article only considers a single-type terminal event (death). In the presence of multiple "failure types," for example, death and dropout due to disease-related causes, the proposed SMJM can be extended to accommodate competing risks survival data. These interesting topics are beyond the focus of this article, but are warranted in our research pipeline under investigation.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Editor, an Associate Editor and two anonymous referees for their significant time, insightful comments and constructive suggestions that led to a marked improvement of the article.

DATA AVAILABILITY STATEMENT

The data used in this article are publicly accessible at website: https://stat.ethz.ch/R-manual/R-devel/library/survival/html/pbc.html.

ORCID

Yangxin Huang https://orcid.org/0000-0002-9792-8061

REFERENCES

1. Murtaugh PA, Dickson ER, Van Dam GM, et al. Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology*. 1994;20(1):126-134.
2. Dil E, Karasoy D. Joint modeling of a longitudinal measurement and parametric survival data with application to primary biliary cirrhosis study. *Pakistan J Stat Oper Res*. 2020;16(2):295-304.
3. Ding J, Wang JL. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*. 2008;64(2):546-556.
4. Papageorgiou G, Mauff K, Tomer A, et al. An overview of joint modeling of time-to-event and longitudinal outcomes. *Annu Rev Stat Appl*. 2019;6:223-240.
5. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000;1(4):465-480.
6. Huang Y, Dagne G, Wu L. Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. *Stat Med*. 2011;30(24):2930-2946.
7. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*. 2011;67(3):819-829.
8. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: with Applications in R*. New York: CRC Press; 2012.
9. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin*. 2004;14:809-834.
10. Wu L, Liu W, Hu X. Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. *Biometrics*. 2010;66(2):327-335.
11. Albert PS, Shih JH. An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. *Ann Appl Stat*. 2010;4(3):1517-1532.
12. Arellano-Valle RB, Genton MG. On fundamental skew distributions. *J Multivar Anal*. 2005;96(1):93-116.
13. Azzalini A, Capitanio A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J R Stat Soc Series B Stat Methodol*. 2003;65(2):367-389.
14. Huang Y, Dagne GA, Zhou S, et al. Piecewise mixed-effects models with skew distributions for evaluating viral load changes: a Bayesian approach. *Stat Methods Med Res*. 2015;24(6):730-746.
15. Sahu SK, Dey DK, Branco MD. A new class of multivariate skew distributions with applications to Bayesian regression models. *Can J Stat*. 2003;31(2):129-150.
16. Huang Y. Quantile regression-based Bayesian semiparametric mixed-effects models for longitudinal data with non-normal, missing and mismeasured covariate. *J Stat Comput Simul*. 2016;86(6):1183-1202.
17. Douglas JA. Item response models for longitudinal quality of life data in clinical trials. *Stat Med*. 1999;18(21):2917-2931.
18. Glas CA, Geerlings H, van de Laar MA, et al. Analysis of longitudinal randomized clinical trials using item response models. *Contemp Clin Trials*. 2009;30(2):158-170.
19. He B, Luo S. Joint modeling of multivariate longitudinal measurements and survival data with applications to parkinson's disease. *Stat Methods Med Res*. 2016;25(4):1346-1358.

20. Wang C, Douglas J, Anderson S. Item response models for joint analysis of quality of life and survival. *Stat Med*. 2002;21(1):129-142.
21. Adams RJ, Wilson M, Wu M. Multilevel item response models: an approach to errors in variables regression. *J Educ Behav Stat*. 1997;22(1):47-76.
22. Andersen EB. Latent regression analysis based on the rating scale model. *Psychol Sci*. 2004;46:209-226.
23. Mislevy RJ. Estimation of latent group effects. *J Am Stat Assoc*. 1985;80(392):993-997.
24. Fox JP. Multilevel IRT using dichotomous and polytomous response data. *Br J Math Stat Psychol*. 2005;58(1):145-172.
25. Kamata A. Item analysis by the hierarchical generalized linear model. *J Educ Meas*. 2001;38(1):79-93.
26. Maier KS. A rasch hierarchical measurement model. *J Educ Behav Stat*. 2001;26(3):307-330.
27. Fox J. *Bayesian Item Response Modeling: Theory and Applications*. New York, USA: Springer-Verlag; 2010.
28. Fox JP, Glas CA. Bayesian estimation of a multilevel IRT model using gibbs sampling. *Psychometrika*. 2001;66(2):271-288.
29. Huang Y, Chen J, Xu L, et al. Bayesian MLIRT-based joint models for multivariate longitudinal and survival data with multiple features. *J of Med Stat Inform*. 2021;9(4):1-14.
30. Hung LF, Wang WC. The generalized multilevel facets model for longitudinal data. *J Educ Behav Stat*. 2012;37(2):231-255.
31. Luo S. A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Stat Med*. 2014;33(4):580-594.
32. Natesan P, Limbers C, Varni JW. Bayesian estimation of graded response multilevel models using gibbs sampling: formulation and illustration. *Educ Psychol Meas*. 2010;70(3):420-439.
33. Lord F, Novick M, Birnbaum A. *Statistical Theories of Mental Test Scores*. Boston, MA: Addison-Wesley; 1968.
34. Chen J, Huang Y. A Bayesian mixture of semiparametric mixed-effects joint models for skewed-longitudinal and time-to-event data. *Stat Med*. 2015;34(20):2820-2843.
35. Zhang H, Huang Y. Bayesian joint modeling for partially linear mixed-effects quantile regression of longitudinal and time-to-event data with limit of detection, covariate measurement errors and skewness. *J Biopharm Stat*. 2021;31(3):295-316.
36. O'Brien LM, Fitzmaurice GM. Analysis of longitudinal multiple-source binary data using generalized estimating equations. *J R Stat Soc Ser C Appl Stat*. 2004;53(1):177-193.
37. Little RJ, Rubin DB. *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley & Sons; 2014.
38. Huang Y, Dagne G. Bayesian semiparametric nonlinear mixed-effects joint models for data with skewness, missing responses, and measurement errors in covariates. *Biometrics*. 2012;68(3):943-953.
39. Liu W, Wu L. Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. *Biometrics*. 2007;63(2):342-350.
40. Curtis S. Bugs code for item response theory. *J Stat Softw*. 2010;36:1-34.
41. Tong P, Coombes KR. Integirty: a method to identify genes altered in cancer by accounting for multiple mechanisms of regulation using item response theory. *Bioinformatics*. 2012;28(22):2861-2869.
42. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurement Data*. Vol 62. New York: CRC Press; 1995.
43. Wulfsohn MS, Tsiatis AA A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997;53:330-339.
44. Huang Y, Chen J. Bayesian quantile regression-based nonlinear mixed-effects joint models for time-to-event and longitudinal data with multiple features. *Stat Med*. 2016;35(30):5666-5685.
45. Clayton DG. A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*. 1991;47:467-485.
46. Chen J, Huang Y, Tang NS. Bayesian change-point joint models for multivariate longitudinal and time-to-event data. *Stat Biopharm Res*. 2022;14(2):227-241.
47. Rizopoulos D. JM: an R package for the joint modelling of longitudinal and time-to-event data. *J Stat Softw*. 2010;35(9):1-33.
48. Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*. 2003;59(2):221-228.
49. Fleming T, Harrington D. *Counting Processes and Survival Analysis*. Hoboken, NJ: John Wiley & Sons; 1991.
50. Gordon RD, Iwatsuki S, Tzakis AG, et al. The Denver-Pittsburgh liver transplant series. *Clin Transpl*. 1987; Chapter 7;1:43-49.
51. Andrinopoulou ER, Rizopoulos D. Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Stat Med*. 2016;35(26):4813-4823.
52. Crowther MJ, Abrams KR, Lambert PC. Joint modeling of longitudinal and survival data. *Stata J*. 2013;13(1):165-184.
53. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457-472.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Chen J, Huang Y, Wang Q. Semiparametric multivariate joint model for skewed-longitudinal and survival data: A Bayesian approach. *Statistics in Medicine*. 2023;1-18. doi: 10.1002/sim.9896