BIOSTATISTICS & EPIDEMIOLOGY
https://doi.org/10.1080/24709360.2019.1693198

Taylor & Francis
Taylor & Francis Group

Dynamic prediction using joint models of longitudinal and recurrent event data: a Bayesian perspective

Xuehan Ren[a], Jue Wang[b] and Sheng Luo [c]

[a]Gilead Sciences Inc., Foster City, CA, USA; [b]Genentech Inc., South San Francisco, CA, USA; [c]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA

ABSTRACT

In cardiovascular disease (CVD) studies, the events of interest may be recurrent (multiple occurrences from the same individual). During the study follow-up, longitudinal measurements are often available and these measurements are highly predictive of event recurrences. It is of great clinical interest to make personalized prediction of the next occurrence of recurrent events using the available clinical information, because it enables clinicians to make more informed and personalized decisions and recommendations. To this end, we propose a joint model of longitudinal and recurrent event data. We develop a Bayesian approach for model inference and a dynamic prediction framework for predicting target subjects' future outcome trajectories and risk of next recurrent event, based on their data up to the prediction time point. To improve computation efficiency, embarrassingly parallel MCMC (EP-MCMC) method is utilized. It partitions the data into multiple subsets, runs MCMC sampler on each subset, and applies random partition trees to combine the posterior draws from all subsets. Our method development is motivated by and applied to the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), one of the largest CVD studies to compare the effectiveness of medications to treat hypertension.

ARTICLE HISTORY
Received 18 March 2019
Accepted 30 October 2019

KEYWORDS
ALLHAT study; personalized prediction; cardiovascular disease; parallel EP-MCMC

1. Introduction

As a leading cause of death and disability in both men and women, cardiovascular disease (CVD) causes one in every four deaths in the United States and has became a major public health concern [1]. The estimated direct costs for CVD is 199.2 billion dollars in the United States [2]. It is critically important to identify risk factors of CVD which yields important insight into prediction, prevention, and treatment of disease progression. To this end, National Heart, Lung, and Blood Institute (NHLBI) has funded many clinical studies aiming to identify risk factors for predicting future CVD events, e.g. stroke, reoccurred heart failure, and recurrent heart attack [3]. Anti-hypertensive and Lipid-lowing Treatment to Prevent Heart Attack Trial (ALLHAT) study [4] is one of the largest CVD studies that

CONTACT Sheng Luo sheng.luo@duke.edu Department of Biostatistics and Bioinformatics, Duke University, 2424 Erwin Rd, Durham, NC 27705, USA

was sponsored by NHLBI. Previous literatures have identified multiple important CVD risk factors, e.g. hypertension, high cholesterol level, and diabetes [5,6].

An important feature of cardiovascular disease is that the CVD event occurrences are often recurrent, i.e. subjects may experience multiple CVD events during the follow-up period. There exists intrinsic correlation among these recurrent events within the same subject. Ignoring the correlation leads to biased estimated risk [3]. Hence, the commonly used Cox proportional hazard model does not apply, because it does not account for the correlation. Moreover, the recurrent event process tends to be correlated to some longitudinal measures, e.g. people with high systolic blood pressure are more likely to experience recurrent CVD events. Analyzing these two processes separately may lead to biased estimates. Hence, it is essential to jointly model the longitudinal measures and recurrent events.

Joint models (JM) of longitudinal measurements and time-to-event data (including terminal event, competing risks, and recurrent events) have been widely studied in the past two decades and a increasing number of studies applied various joint models to large clinical studies, e.g. Henderson et al. [7], Han et al. [8], and Crowther et al. [9]. A joint model often involves two submodels, a mixed effects submodel for the longitudinal measurements and a survival submodel for the time-to-event data, linked by shared random effects or latent structures. Many extensions of joint models have been developed, e.g. relaxing the normality assumptions of random effects [10]. Of the existing literature about joint modeling of longitudinal measurements and recurrent events, Liu and Huang [11] considered a more complex setting where a repeated measures process and a recurrent event process were correlated, both subject to a terminal event.

One novel use of JM is to provide dynamic predictions of the risk of target event and the trajectories of longitudinal outcomes [12]. The predictions are dynamic because they can be updated given the additional information of longitudinal trajectories and longer history of survival process. In the literature of dynamic prediction of recurrent event risks, Krol et al. [13] extended the models of Liu and Huang [11] to dynamic prediction. They derived the estimated probability of having a terminal event in a specific time interval, given all previous history and no event occurred before the prediction time. Regarding prediction of recurrent events, Musoro et al. [14] extended dynamic prediction by landmarking to recurrent event data. Using landmark method, they handled the longitudinal data as a time-fixed covariate at different landmark time points. To the best of our knowledge, providing personalized dynamic predictions of future longitudinal trajectories and recurrent event risks using joint model approach remains an open research question. In this article, we develop a Bayesian joint model that incorporates longitudinal clinical information and the recurrent disease history. The proposed model consists of two submodels, the linear mixed effect submodel for the longitudinal trajectory and the recurrent event model with an intensity function of a Poisson process. The two submodels are linked via a shared latent trajectory. Moreover, the proposed joint model provides dynamic predictions of the risk of recurrent event occurrences and the future longitudinal trajectories.

The rest of the article is organized as follows. In Section 2, we introduce a cardiovascular disease study that motivates the methodological development of this article. The joint model framework and dynamic prediction method are illustrated in Section 3. We conduct simulations in Section 4 to assess our proposed model's performance in inference accuracy and prediction. In Section 5, we apply our proposed model to the motivating dataset. In

Section 6, we summarize the findings, and discuss limitation and some future research directions.

2. Motivating dataset

The methodological development is motivated by Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), a randomized, double-blinded, active-controlled clinical trial conducted from February 1994 through March 2002 [4]. The aim of the ALLHAT study is to determine whether newer anti-hypertensive agents, including α-blocker doxazosin, differ from the reference treatment chlorthalidone in coronary heart disease events and other CVD events. The primary outcome was composite fatal coronary heart disease (CHD) or non-fatal myocardial infarction (MI). During the study follow-up period, 15,255 subjects were randomized to chlorthalidone arm and 9,061 subjects were randomized to doxazosin arm, our analysis is based on subjects from these two arms. The maximum follow-up time in the analytic dataset was 5.57 years. Subjects who were alive at the end of the trial, or lost to follow-up during the trial are considered as right censored. We are interested in the composite outcome of three types of recurrent CVD events (CHD, stroke, and heart failure) before death or censoring. To visualize the data structure, Figure 1 displays the time plot of four selected subjects who experienced composite CVD events before death or censoring. For example, subject 1 experienced one CVD event in month

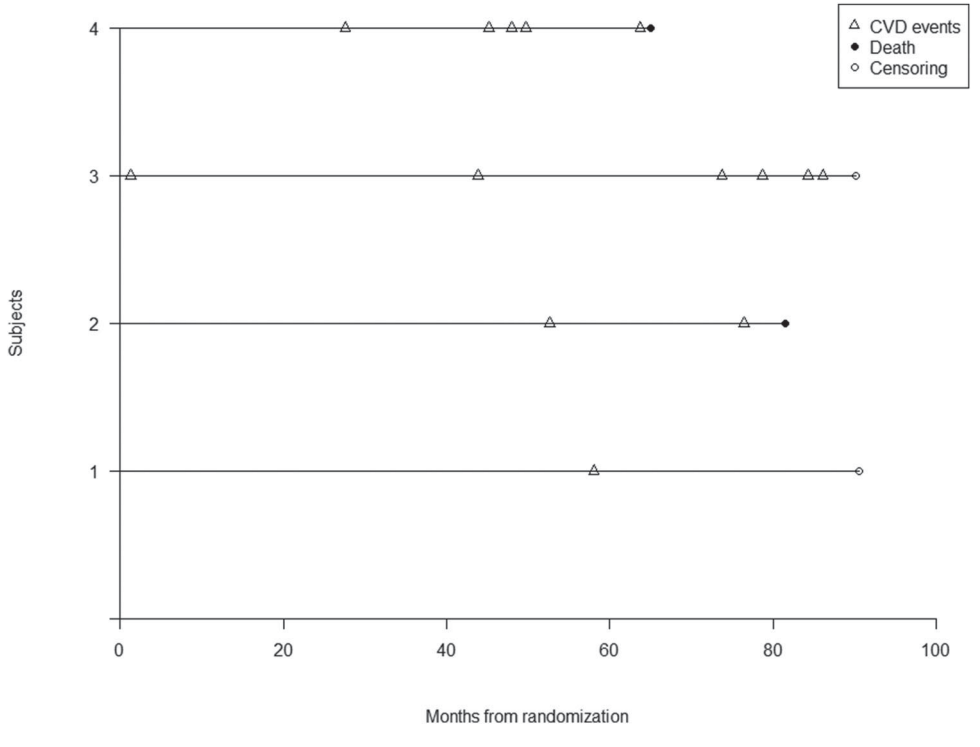

Figure 1. Data display of CVD events in ALLHAT study.

59 and was censored on month 92, while subject 4 experienced 5 recurrent CVD events and died in 67 month.

During the study, participants' clinical information was collected and recorded at each visit. Systolic blood pressure (SBP) is the most frequently measured clinical outcome in ALLHAT study. It is well documented that high blood pressure (hypertension) is one of the main risk factors for CVD events, and previous literature has studied the mechanisms of how hypertension causes CHD, stroke, and heart failure [15]. In the risk prediction of CVD events, systolic blood pressure is usually included in the statistical models. Staessen et al. [16] found that among untreated patients with isolated systolic hypertension, systolic blood pressure predicted cardiovascular risk. More recent publications suggested that there was significant positive relationship between higher SBP and CVD events [17,18]. On the other hand, subjects with previous CVD events have high risk of future CVD recurrences [19]. Thus, incorporating the SBP trajectory and recurrent disease history is essential to personalized prediction of next occurrence of a CVD event.

3. Methods

3.1. Joint model specification

Let $\boldsymbol{y}_i(t) = \{y_i(t_{ij})\}$ be the vector of longitudinal observation for subject i at time t_{ij}, where $i = 1, \ldots, n, j = 1, \ldots, m_i$, and n is the number of subjects and m_i is the number of visits. Let T_{ik} be the recurrent event occurrence times from study onset for subject $i, k = 0, \ldots, n_i$, where n_i denotes the number of recurrent CVD events (including CHD, stroke, and HF). We propose a joint model consisting of a linear mixed effects submodel for the longitudinal health outcome and a poisson intensity submodel for the recurrent CVD events (under the assumption that the number of recurrent events in non-overlapping time intervals is a poisson process).

$$\begin{aligned} y_i(t) &= \boldsymbol{X}_i^Y(t)\boldsymbol{\alpha} + \boldsymbol{Z}_i(t)\boldsymbol{u}_i + e_i(t) = f_i(t) + e_i(t) \\ r_i(t) &= r_0(t)\exp\{\boldsymbol{Z}_i^R\boldsymbol{\beta} + \nu f_i(t) + v_i\}, \end{aligned} \tag{1}$$

where $f_i(t)$ is the unobserved true value of the longitudinal outcome and it connects two submodels, $r_0(t)$ is the baseline hazard function for the recurrent events. The covariate vectors $\boldsymbol{X}_i^Y(t)$ and $\boldsymbol{Z}_i^R$ are covariates related to the longitudinal process and recurrent event process, respectively, and they may be identical or different. Parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the corresponding coefficients. Vector $\boldsymbol{Z}_i(t)$ is a covariate vector corresponding to the random effects vector $\boldsymbol{u}_i$. Random effects vector $\boldsymbol{u}_i$ accounts for the correlation among repeated measurements and it is assumed to be normally distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_u$. Random effects $v_i \sim N(0, \sigma_v^2)$ is independent of $\boldsymbol{u}_i$ and it accounts for the heterogeneity in the rate of recurrent events due to random effects not shared with the repeated measurements. The measurement error $e_i(t) \sim N(0, \sigma_e^2)$ is independent from random effects $\boldsymbol{u}_i$ and v_i. The two submodels are linked via an association parameter ν, which quantifies the strength of correlation between the expected longitudinal outcome and the risk of recurrent CVD events.

One key assumption in the linear mixed effects submodel is that all measurements from each subject are independent conditional on the random effects vector $\boldsymbol{u}_i$. Thus, the

likelihood from the longitudinal process of subject i is:

$$l_i^Y = \prod_{j=1}^{m_i} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{ -\frac{[y_i(t_{ij}) - \boldsymbol{X}_i^Y(t_{ij})\boldsymbol{\alpha} - \boldsymbol{Z}_i(t_{ij})\boldsymbol{u}_i]^2}{2\sigma_e^2} \right\}. \tag{2}$$

We approximate the baseline hazard function $r_0(t)$ using piecewise constant functions [20]. Specifically, we divide the total follow-up time interval $[0, \tau]$ by quantiles of event times $\boldsymbol{\tau}_t = (0, \tau_1, \ldots, \tau_R)'$, and denote the baseline hazard vector as $\boldsymbol{g} = (g_0, g_1, \ldots, g_{R-1})'$. The piecewise constant hazard function is expressed as $h_0(t) = \sum_{r=0}^{R-1} g_r I_r(t)$, where indicator function $I_r(t) = 1$, if $\tau_r \le t < \tau_{r+1}$ and 0 otherwise. Let $S_i(x_i)$ denote the corresponding survival function. The likelihood of the recurrent event process for subject i is:

$$\begin{aligned} l_i^R &= \prod_{k=0}^{n_i} h_i(t_{ik})^{\delta_{ik}} S_i(x_i) \\ &= \prod_{k=0}^{n_i} \left[r_0(t_{ik}) \exp\left\{ \boldsymbol{Z}_i^R \boldsymbol{\beta} + \nu f_i(t_{ik}) + \nu_i \right\} \right]^{\delta_{ik}} \\ &\quad \cdot \exp\left[- \int_0^{\tau_i} r_0(t) \exp\left\{ \boldsymbol{Z}_i^R \boldsymbol{\beta} + \nu f_i(t) + \nu_i \right\} \, \mathrm{d}t \right], \end{aligned}$$

where δ_{ik} is the indicator of a recurrent event at time t_{ik} and τ_i is the observed follow-up time of subject i. The longitudinal outcome $\boldsymbol{y}_i$ is assumed to be independent of recurrent event times, conditioning on the random effects $\boldsymbol{u}_i$ and ν_i. Hence, the full likelihood for subject i is

$$l_i = l_i^Y \cdot l_i^R \cdot f(\boldsymbol{u}_i) \cdot f(\nu_i), \tag{3}$$

where $f(\boldsymbol{u}_i)$ and $f(\nu_i)$ are the density functions of $\boldsymbol{u}_i$ and ν_i, respectively. The parameter vector is denoted by $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \nu, \boldsymbol{\Sigma}_u, \sigma_\nu, \sigma_e\}$.

3.2. Dynamic prediction

Next, we illustrate the dynamic prediction framework based on the proposed joint model. Suppose a new subject i had n_i (e.g. $n_i = 0, 1, 2, \ldots$) recurrent events up to prediction time t, with longitudinal profile $\boldsymbol{y}_i(t) = \{y_i(t_{ij}); 0 \le t_{ij} \le t\}$, we would like to predict his/her probability of having the next recurrent event $(n_i + 1)$ before future time $t' = t + \Delta t$ (e.g. prediction time window Δt is 1 year from time t), denoted by $\pi_i(t'|t) = P(T_{i,n_i+1} \le t' | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the parameter vector.

We derive $\pi_i(t'|t)$ as follows:

$$\begin{aligned} \pi_i(t'|t) &= \int \int P(T_{i,n_i+1} \le t' | T_{i,n_i+1} > t, \boldsymbol{u}_i, \nu_i, \boldsymbol{y}_i(t), \boldsymbol{\theta}) \\ &\quad \cdot P(\boldsymbol{u}_i | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}, \nu_i) d\boldsymbol{u}_i P(\nu_i | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}, \boldsymbol{u}_i) \, \mathrm{d}\nu_i \\ &\approx \frac{1}{M} \sum_{m=1}^{M} 1 - \exp\left[- \int_t^{t'} r_i^{(m)}(s | \boldsymbol{u}_i^{(m)}, \nu_i^{(m)}, \boldsymbol{\theta}^{(m)}) \, \mathrm{d}s \right], \end{aligned} \tag{4}$$

where the integration with respect to $\boldsymbol{u}_i$ and v_i in the first equality is approximated using Monte Carlo method, by plugging in the samples of the parameter vector and random effects $\{\boldsymbol{\theta}^{(m)}, \boldsymbol{u}_i^{(m)}, v_i^{(m)}, m = 1, \ldots, M\}$ into the proposed model, where M is the number of posterior samples. The term $r_i^{(m)}$ denotes the intensity function of Poisson process based on the mth sample. Detailed derivation can be found in Appendix.

The key step to approximate the event probability $\pi_i(t'|t)$ is to obtain samples for random effects $\boldsymbol{u}_i$ and v_i. Specifically, conditional on the mth posterior sample $\boldsymbol{\theta}^{(m)}$ and the $m-1$th posterior sample $v_i^{(m-1)}$, we draw samples of $\boldsymbol{u}_i$ from its posterior distribution.

$$P(\boldsymbol{u}_i|T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}^{(m)}, v_i^{(m-1)}) = \frac{P\left(\boldsymbol{y}_i(t), T_{i,n_i+1} > t, \boldsymbol{u}_i|\boldsymbol{\theta}^{(m)}, v_i^{(m-1)}\right)}{P\left(\boldsymbol{y}_i(t), T_{i,n_i+1} > t|\boldsymbol{\theta}^{(m)}, v_i^{(m-1)}\right)}$$

$$\propto P\left(\boldsymbol{y}_i(t), T_{i,n_i+1} > t, \boldsymbol{u}_i|\boldsymbol{\theta}^{(m)}, v_i^{(m-1)}\right) = P\left(\boldsymbol{y}_i(t)|\boldsymbol{u}_i, \boldsymbol{\theta}^{(m)}\right)$$
$$P\left(T_{i,n_i+1} > t|\boldsymbol{u}_i, \boldsymbol{\theta}^{(m)}, v_i^{(m-1)}\right) P\left(\boldsymbol{u}_i|\boldsymbol{\theta}^{(m)}\right),$$

where

$$P\left(T_{i,n_i+1} > t|\boldsymbol{u}_i, \boldsymbol{\theta}^{(m)}, v_i^{(m-1)}\right)$$
$$= \prod_{k=0}^{n_i}\left[r_0^{(m)}(t_{ik}) \exp\left\{\boldsymbol{Z}_i^R\boldsymbol{\beta}^{(m)} + \nu^{(m)}f_i^{(m)}(t_{ik}) + v_i^{(m-1)}\right\}\right]^{\delta_{ik}}$$
$$\cdot \exp\left[-\int_0^t r_0^{(m)}(s) \exp\left\{\boldsymbol{Z}_i^R\boldsymbol{\beta}^{(m)} + \nu^{(m)}f_i^{(m)}(s) + v_i^{(m-1)}\right\} \mathrm{d}s\right].$$

We then draw samples of v_i from its posterior distribution, conditional on $\boldsymbol{\theta}^{(m)}$ and the mth posterior sample $\boldsymbol{u}_i^{(m)}$.

$$P\left(v_i|T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}^{(m)}, \boldsymbol{u}_i^{(m)}\right) = \frac{P\left(T_{i,n_i+1} > t, v_i|\boldsymbol{y}_i(t), \boldsymbol{\theta}^{(m)}, \boldsymbol{u}_i^{(m)}\right)}{P\left(T_{i,n_i+1} > t|\boldsymbol{y}_i(t), \boldsymbol{\theta}^{(m)}, \boldsymbol{u}_i^{(m)}\right)}$$

$$\propto P\left(T_{i,n_i+1} > t, v_i|\boldsymbol{y}_i(t), \boldsymbol{\theta}^{(m)}, \boldsymbol{u}_i^{(m)}\right) = P\left(T_{i,n_i+1} > t|\boldsymbol{y}_i(t), \boldsymbol{\theta}^{(m)}, \boldsymbol{u}_i^{(m)}, v_i\right) P\left(v_i|\boldsymbol{\theta}^{(m)}\right),$$

where

$$P\left(T_{i,n_i+1} > t|\boldsymbol{y}_i(t), \boldsymbol{\theta}^{(m)}, \boldsymbol{u}_i^{(m)}, v_i\right)$$
$$= \prod_{k=0}^{n_i}\left[r_0^{(m)}(t_{ik}) \exp\left\{\boldsymbol{Z}_i^R\boldsymbol{\beta}^{(m)} + \nu^{(m)}f_i^{(m)}(t_{ik}) + v_i\right\}\right]^{\delta_{ik}}$$
$$\cdot \exp\left[-\int_0^t r_0^{(m)}(s) \exp\left\{\boldsymbol{Z}_i^R\boldsymbol{\beta}^{(m)} + \nu^{(m)}f_i^{(m)}(s) + v_i\right\} \mathrm{d}s\right].$$

For each of $\boldsymbol{\theta}^{(m)}$, $m = 1, \ldots, M$, we adopt adaptive rejection metropolis sampling (ARMS) [21] implemented in `R HI` package to draw 50 samples of $\boldsymbol{u}_i$ and v_i and retain the last sample. This process is repeated for the M saved values of $\boldsymbol{\theta}$ to obtain M samples of $\boldsymbol{u}_i$ and v_i. Based on all the samples, we can also predict the new subject's future outcome trajectory. Specifically, the outcome trajectory at time t' is $y_i(t')|\boldsymbol{u}_i \sim N(\boldsymbol{X}_i^Y(t')\boldsymbol{\alpha}^{(m)} + \boldsymbol{Z}_i(t')\boldsymbol{u}_i^{(m)}, \sigma_e^{2(m)})$. Suppose that subject i does not have a recurrent event between times t and t', then the outcome history is updated to $\boldsymbol{y}_i(t')$. We draw new samples of $\boldsymbol{u}_i$ and v_i from their dynamically updated posterior distributions $P(\boldsymbol{u}_i|T_{i,n_i+1} > t', \boldsymbol{y}_i(t'), \boldsymbol{\theta}, v_i)$ and $P(v_i|T_{i,n_i+1} > t', \boldsymbol{y}_i(t'), \boldsymbol{\theta}, \boldsymbol{u}_i)$ and obtain the updated predictions.

3.3. Bayesian inference

We use Bayesian inference based on Markov chain Monte Carlo (MCMC) posterior simulations to infer unknown parameters. We use non-informative priors on all parameters in vector $\boldsymbol{\theta}$. Specifically, the prior distribution of all elements in the coefficient vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and the association parameter ν is normal distribution with mean 0 and standard deviation being 10. To ensure the positivity of variances σ_v, σ_e and the variance terms in $\boldsymbol{\Sigma}_u$, we use inverse gamma distribution with $\alpha = 0.01, \beta = 0.01$ so that the prior variance is 100. We impose the uniform prior distribution $U(-1, 1)$ for the correlation parameters in $\boldsymbol{\Sigma}_u$.

The model fitting is performed in `Stan` by specifying the full likelihood function and the prior distributions of all unknown parameters. `Stan` adopts a No-U-Turn sampler [22], which offers faster convergence and parameter space exploration compared with other MCMC algorithms such as Gibbs sampling. To diagnose the convergence of posterior samples, we use trace plots and view the absence of apparent trends in the plots as evidence of convergence. Moreover, we ensure Gelman and Rubin [23] potential scale reduction statistic $\hat{R}$ of all parameters are smaller than 1.1, as an additional way of convergence assessment. In simulation, 2,000 after burn-in samples are drawn for each parameter in $\boldsymbol{\theta}$, the convergence of these posterior samples are ensured by meeting aforementioned criteria.

3.4. Parallelizing MCMC with random partition trees

Bayesian inference for large clinical studies can be demanding in computation and memory. Multiple articles attempted to address these issues [24,25], but they either rely on asymptotic normality of posterior distributions, or have the drawback of insufficient resampling. To the end, Wang et al. [26] proposed an embarrassingly parallel MCMC (EP-MCMC) approach to address the computation and memory challenges in datasets with large sample size, without assuming any distribution of posterior samples. The theoretical property of this EP-MCMC approach grants its good performance when being applied to different models. The algorithm consists of two steps. First, it partitions the data into multiple subsets and independently runs MCMC sampler on each of them, then it applies random partition trees to combine the posterior draws from all subsets.

Next, we illustrate how to fit our proposed model using EP-MCMC method. We randomly divide the dataset into 10 subsets (for example), each subset contains sufficient number of subjects and recurrent events. We then conduct Bayesian inference of our

Table 1. Parameter estimation in the simulation study.

	Results using regular MCMC					Results using EP-MCMC				
Parameter	BIAS	SD	SE	CP	RMSE	BIAS	SD	SE	CP	RMSE
$\alpha_0 = -1.000$	0.001	0.169	0.184	0.960	0.169	0.010	0.175	0.184	0.955	0.175
$\alpha_1 = -0.200$	−0.007	0.235	0.244	0.955	0.235	−0.011	0.241	0.244	0.960	0.240
$\alpha_2 = 0.800$	0.000	0.037	0.038	0.955	0.037	0.000	0.037	0.038	0.955	0.037
$\sigma_1 = 1.500$	−0.005	0.123	0.138	0.970	0.123	−0.003	0.123	0.138	0.970	0.123
$\sigma_2 = 0.150$	0.008	0.025	0.030	0.980	0.026	0.007	0.025	0.030	0.985	0.026
$\sigma_v = 0.200$	0.001	0.039	0.041	0.960	0.039	0.001	0.040	0.041	0.955	0.040
$\rho = 0.400$	0.006	0.183	0.244	0.995	0.183	0.008	0.184	0.245	0.995	0.183
$\sigma_e = 5.000$	−0.001	0.068	0.065	0.940	0.067	0.000	0.065	0.065	0.950	0.065
$\beta_1 = -0.12$	0.003	0.025	0.025	0.955	0.025	0.002	0.025	0.025	0.960	0.025
$\nu = 0.750$	−0.001	0.039	0.042	0.980	0.039	−0.001	0.038	0.042	0.980	0.038

proposed model in each subset, resulting into 2000 after burn-in MCMC posterior samples for every parameter of interest. For example, for the 10 unknown parameters listed in Table 1, we have 10 matrices of size 2000 by 10, with one matrix from one subset. We then implement `PART` algorithm [26] to combine the posterior samples from 10 subsets. This algorithm aggregates sub-chain posterior MCMC samples and draws a certain number (e.g. 20,000) of samples for each parameter from the combined posterior with k-dimensional tree partition rules. In simulation, the aggregation step produces a matrix of size $20,000 \times 10$ as the posterior samples of the 10 parameters. The simulation results in Table 1 indicates the validity of the EP-MCMC approach in making Bayesian inference of our proposed model.

Moreover, EP-MCMC method has additional computational efficiency when conducting internal cross-validation in dynamic prediction. When using traditional Bayesian inference, we first randomly split the whole dataset into 10 (for example) subsets of about equal sample size. Then we estimate parameters in the training set with 9 subsets (90% of data) and conduct the dynamic prediction on the validation subset (10% of data). The step is repeated 10 times. This procedure is computationally intensive because the model building step is always based on about 90% of the data and this step has to be repeated 10 times. In contrast, EP-MCMC estimates parameters in each subset in a parallel fashion. Then it performs the aggregation step based on the posterior samples of the unknown parameters from 9 selected subsets and uses the aggregated samples to conduct dynamic prediction on the remaining subset. This process is repeated 10 times without repeating the model building steps 10 times based on 90% of the data. Therefore, EP-MCMC method is much more computationally efficient than the traditional Bayesian inference.

3.5. Assessing predictive performance

We assess the prediction performance of the proposed model in both discrimination (discriminate between subjects who will have the next recurrent event from subjects who will not) and calibration (how well the model predicts the observed data). Specifically, we employ receiver operating characteristic (ROC) curve and the area under the ROC curves (AUC) to assess discrimination and the expected Brier score (BS) for calibration. AUC and BS measure different aspects of the predictive performance. While AUC assesses model's overall discrimination ability, BS measures the bias between the predicted and observed

event risks. Moreover, we adopt the methodology in Blanche et al. [17] to quantify and compare the predictive accuracy of different models in terms of AUC and BS.

For a given threshold $c \in [0,1]$, the time-dependent sensitivity and specificity are defined as $P\{\pi_i(t'|t) > c | D(t,t') = 1, T_i^* > t\}$ and $P\{\pi_i(t'|t) \leq c | D(t,t') = 0, T_i^* > t\}$ respectively, where $D(t,t')$ is an indicator function equals to 1 when a new event happens in time interval $(t,t']$ and 0 otherwise, and T_i^* denotes the true event time for subject i.

Therefore, for probability $p \in [0,1]$, the ROC curve is $\text{ROC}_t^{t'}(p) = TP_t^{t'}[FP_t^{t'}]^{-1}(p)$, where $TP_t^{t'}$ denotes the true positive rate, $FP_t^{t'}$ denotes the false positive rate [27]. We calculate a standard "concordance" summary: the time-dependent Area Under Curve (AUC) [28] $\text{AUC}(t,t') = \int_0^1 \text{ROC}_t^{t'}(p)\,\mathrm{d}p$. In practice, the model with higher AUC has higher discrimination capability.

Using the extension of the Brier score (BS) defined in survival models to joint model framework [29], we define the expected BS for dynamic prediction as $\text{BS}(t,t') = E[(D(t,t') - \pi(t'|t))^2]$, which can be expressed as $\text{BS}(t,t') = E[(E[D(t,t')] - \pi(t'|t))^2] + E[(D(t,t') - E[D(t,t')])^2]$. The first term measures how close the prediction is to the expected event status $E[D(t,t')]$, i.e. how well the model predicts the observed data. The second term is the aggregation of resolution and uncertainty and it does not depend on the distribution of the predictions. In general, BS = 0 indicates perfect prediction and BS = 0.25 means no better than a random guess.

Blanche et al. [17] proposed to compute confidence regions of AUC and BS and to perform significance tests of between-model differences when comparing different models. Let us denote either $\text{AUC}(t,t')$ or $\text{BS}(t,t')$ as $\theta(t,t')$ and let $\Delta\hat{\theta}(t,t')$ denote the estimated difference of AUC or BS between two predictive models. The pointwise confidence interval is $\Delta\hat{\theta}(t,t') \pm z_{1-\alpha/2}(\hat{\sigma}_{\Delta,t,t'}/\sqrt{n})$, where $z_{1-\alpha/2}$ is the critical value of standard normal distribution at the significance level α and $\hat{\sigma}^2_{\Delta,t,t'}$ is the empirical estimate of variance, which can be consistently estimated by the influence function [17].

4. Simulation setting

In this section, we assess the performance of our proposed model in inference and prediction. Consider three binary covariates x_{i1}, x_{i2}, and z_i which equal to 0 or 1 with probability 0.5. We generate the longitudinal measurements and the recurrent events from the models

$$y_i(t_{ij}) = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 t_{ij} + u_{i1} + u_{i2} t_{ij} + e_i(t_{ij}) = f_i(t_{ij}) + e_i(t_{ij})$$
$$r_i(t_{ik}) = r_0 \exp\{\beta_1 z_i + \nu f_i(t_{ij}) + \nu_i\},$$

where $e_i(t_{ij}) \sim N(0, \sigma_e^2)$, $\boldsymbol{u}_i = (u_{i1}, u_{i2})' \sim N(\mathbf{0}, \Sigma_u)$, and Σ_u is the covariance matrix with the variance terms being σ_1^2 and σ_2^2 and the correlation coefficient term being ρ. The censoring time C_i is sampled from the uniform distribution $Uniform(9, 10)$ which generates around 50% censoring.

The inference results are based on 200 replications with the sample size being 800. We randomly set aside 200 subjects as the validation dataset and the inference is based on the remaining 600 subjects. There are on average 2 recurrent events per subject, while around 30% of the simulated subjects have more than 1 event. Table 1 displays the true parameters, bias, standard deviation (SD, the standard deviation of the posterior means), standard error (SE, the square root of the average of the posterior variance), coverage probabilities (CP) of

Table 2. Time-dependent AUCs and BSs for the simulation study.

t	Δt	Bias	AUC	BS
5	1	0.002	0.746	0.014
	2	0.003	0.756	0.016
	3	0.004	0.759	0.017
	4	0.004	0.761	0.020
6	7	0.003	0.788	0.016
	1	0.004	0.788	0.017
	2	0.005	0.789	0.019
7	1	0.003	0.814	0.018
	2	0.004	0.812	0.021

Notes: AUC: area under the ROC curve; BS: Brier score. Notation: t is prediction time; Δt is prediction time window.

95% equal-tail credible intervals (CI), and root mean squared error (RMSE) from results using regular MCMC (left panel) and EP-MCMC (right panel, with 3 subsets). The results suggest that parameter estimates have negligible bias, small RMSE, SE being close to SD, and the coverage probability close to the 95% nominal value, indicating that both regular MCMC and EP-MCMC provide valid Bayesian inference. The computation time is about 12 h when using regular MCMC. It is significantly reduced to 3 h when using EP-MCMC method.

We then apply the dynamic prediction method to the validation dataset. Table 2 displays the bias (average of the difference between the estimated $\pi(t'|t)$ and the true probability), AUC (average of all AUCs calculated within the given time interval from 200 replications), and BS. The bias is all under 0.005 and estimated AUC and BS suggest our model's good performance in discrimination and calibration.

5. ALLHAT study application results

In this section, we apply the proposed joint model to the motivating ALLHAT study. We select the non-fatal cardiovascular disease events (composite CHD, stroke, and heart failure) as the recurrent event outcome and SBP as the longitudinal variable which has significant association with time to CVD events. We consider the following risk factors: treatment (1 if doxazosin and 0 if chlorthalidone), age (in years), gender (1 for male and 0 for female), race (1 for black and 0 for white and others), diabetes (yes/no), history of MI or Stroke at baseline (yes/no), history of coronary revascularization (yes/no), antihypertensive treatment before trial (yes/no). These covariates are significant either in the linear mixed effect model or in the recurrent event model. After excluding subjects with pending death confirmation and those with missing data in the aforementioned covariates, we obtain the analysis dataset with 19,804 subjects. Among them, 16,811 subjects did not have CVD events until censoring, 2179 subjects experienced 1 event, and 814 subjects experienced more than 1 events. Around 15% of ALLHAT subjects experienced recurrent CVD events during follow-up period.

Due to the large sample size of the ALLHAT study, we adopt the EP-MCMC approach for Bayesian inference. Specifically, we randomly divide the analysis dataset into 10 subsets, each with about 2000 subjects. In each subset, we run two parallel MCMC chains with overdispersed initial values and each chain is run for 5000 iterations. The first 3000 iterations are discarded as burn-in and the inference is based on remaining 2000 iterations

Table 3. ALLHAT data analysis results.

	Mean	2.5%	97.5%	SD
longitudinal submodel				
Intercept	2.688	1.962	3.412	0.393
doxazosin	2.796	2.428	3.143	0.186
Age (years)	0.136	0.111	0.162	0.013
Male	1.200	0.808	1.584	0.209
Race:black	2.388	2.008	2.773	0.196
Antihypertensive drug (Yes)	1.119	0.469	1.760	0.339
σ_1	12.159	11.947	12.364	0.107
σ_2	3.333	3.222	3.441	0.056
ρ	−0.563	−0.584	−0.541	0.011
σ_e	12.416	12.363	12.470	0.028
recurrent event submodel				
Diabetes (Yes)	0.456	0.349	0.555	0.054
MI or Stroke (Yes)	0.663	0.554	0.770	0.054
Coronary revascularization (Yes)	0.777	0.654	0.908	0.065
Association parameter	0.016	0.011	0.020	0.002

from each chain. While the model inference takes more than 3 weeks when using regular MCMC, the computation time is reduced to 26 h by using EP-MCMC. We employ piecewise constant baseline hazard function. To have enough observations in each piecewise interval, we construct intervals by every 1/4th quantile of the CVD event time. Good mixing and convergence properties of the MCMC chains are observed in all trace plots and $\hat{R}$ is smaller than 1.1 for all parameters.

Table 3 displays the inference results from the proposed joint model. Here, 95% CI denotes 95% equal-tail credible intervals. The results from the longitudinal submodel suggest that the systolic blood pressure for subjects in doxazosin group is on average 2.796 units (95% CI [2.428, 3.143]) higher than those in the chlorthalidone group. Other risk factors including older age, male gender, black race, and baseline anti-hypertensive drug use are significantly associated with higher systolic blood pressure. Specifically, 1 year increase in age increases the systolic blood pressure by 0.136 units (95% CI [0.111, 0.162]); males tend to have 1.2 units (95% CI [0.808, 1.584]) higher systolic blood pressure than females; black subjects have 2.388 units (95% CI [2.008, 2.773]) higher systolic blood pressure than white subjects and others; subjects who took anti-hypertensive drugs at baseline have 1.119 units (95% CI [0.469, 1.760]) higher systolic blood pressure than those who did not. These results are consistent with the primary ALLHAT publication [30]. For the recurrent event submodel, we find that subjects with diabetes history have significantly higher risk of having cardiovascular disease events (RR = 1.578; 95% CI [1.418, 1.742]). Similar effects can be found in subjects with MI or stroke (RR = 1.941; 95% CI [1.740, 2.160]), and subjects with history of coronary revascularization (RR = 2.175; 95% CI [1.923, 2.479]). Moreover, the association parameter ν is statistically significant, indicating that high longitudinal systolic blood pressure measurements are associated with increased risk of recurrent CVD event (RR = 1.016; 95% CI [1.011, 1.020]).

To assess the predictive performance, we compute the time-dependent AUC and BS of our proposed model (Model 1) and a simple recurrent event model without considering the longitudinal data (Model 2). 10-fold internal cross-validation is conducted following procedures in Section 3.4 to avoid overfitting. We present the estimated AUC and BS under various prediction scenarios (different prediction times t and prediction time windows Δt)

Table 4. Time-dependent AUC and BS for proposed joint model (Model 1) and a recurrent event model (Model 2).

		Model 1		Model 2	
t (year)	Δt (month)	AUC	BS	AUC	BS
1	3	0.665	0.015	0.556	0.015
	6	0.676	0.027	0.556	0.028
	9	0.675	0.038	0.557	0.040
	12	0.668	0.049	0.558	0.051
2	3	0.660	0.015	0.572	0.015
	6	0.653	0.028	0.566	0.029
	9	0.658	0.041	0.559	0.052
	12	0.651	0.041	0.553	0.052
3	3	0.706	0.017	0.652	0.017
	6	0.695	0.034	0.603	0.035
	9	0.676	0.049	0.604	0.052
	12	0.658	0.065	0.598	0.070
4	3	0.744	0.020	0.639	0.021
	6	0.696	0.067	0.616	0.069
	9	0.685	0.059	0.576	0.063
	12	0.643	0.077	0.576	0.080

Notes: AUC: area under the ROC curve; BS: Brier score. Notation: t is prediction time; Δt is prediction time window.

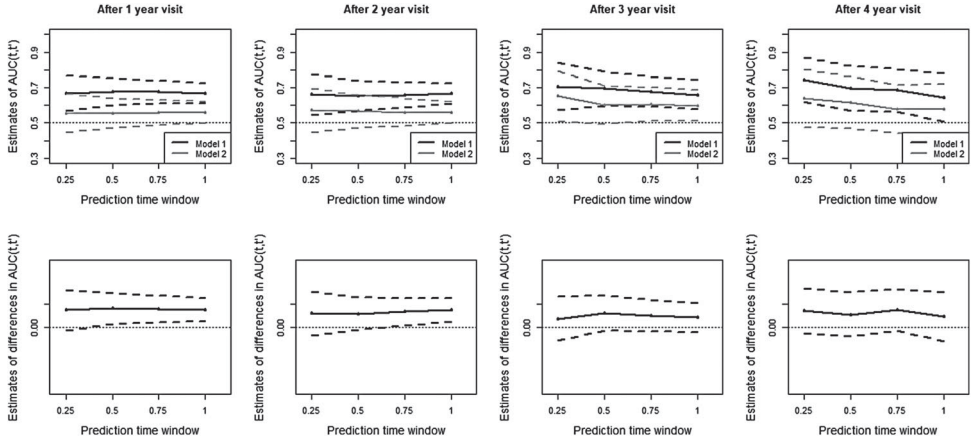

Figure 2. Comparison of AUC (upper panels) from the proposed joint model (Model 1, black solid line) and a recurrent event model (Model 2, gray solid line) and the difference of AUC (lower panels).

in Table 4. Our proposed model has higher AUC and lower BS than the recurrent event model in all prediction scenarios, suggesting that using longitudinal data improves the risk prediction of having a new recurrent event.

Using the testing method in Blanche et al. [17], we obtain the confidence region of AUC for all prediction scenarios in Table 4 and display in the upper panels of Figure 2. We also compute the difference of AUC from Models 1 and 2 and display in the lower panels of Figure 2. We find that in all prediction scenarios, AUC from our proposed model (Model 1; black solid line) is higher than the one from the recurrent event model (Model 2; gray solid line), and the difference of AUC is always above 0. For prediction after year 1 ($t = 1$ year), the estimated difference of AUC from Models 1 and 2 at the prediction windows

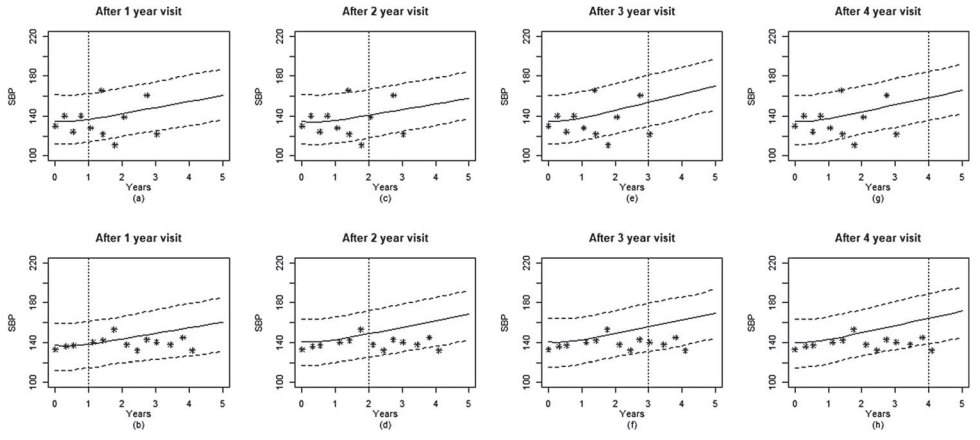

Figure 3. Predicted SBP measurements for Subject 127 (upper panels) and Subject 15 (lower panels) in ALLHAT study. Dashed lines are the 2.5% and 97.5% percentiles. The dotted vertical line represents the prediction time *t*.

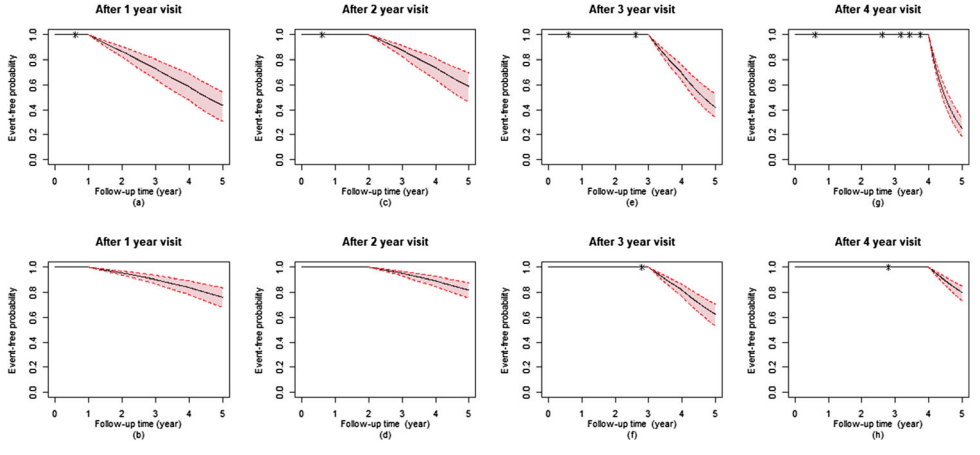

Figure 4. Predicted probability of being CVD event-free for Subject 127 (upper panels) and Subject 15 (lower panels) in the ALLHAT study. Dashed lines are the 2.5% and 97.5% percentiles.

of 0.5, 0.75, and 1 year is significant at 0.05 level (first plot in the lower panels). Similarly, for prediction after year 2 ($t = 2$ year), there is significant difference in estimated AUC between Models 1 and 2 at the prediction windows of 0.75 and 1 year (second plot in the lower panels). The differences of estimated AUC for predictions after year 3 and year 4 are not significant.

Next, we illustrate the joint model providing personalized predictions of both future longitudinal SBP trajectory, and the probability of having a new CVD event in given prediction time window. Figure 3 displays the estimated systolic blood pressure (SBP) of subject 127 (upper panels) and subject 15 (lower panels) at different prediction times *t* from year 1 to year 5. Figure 3 demonstrates how the predicted SBP are updated over time for these two subjects. From left to right on Figure 3, by using more follow-up data, predictions are

closer to the true observed values and the 95% uncertainty band is narrower. It also suggests that subject 127 has higher SBP than subject 15. Figure 4 displays the predicted probability of being free of a new CVD event. Based on the data up to year 1, subject 127 has lower event-free probability (higher risk of having a new CVD event) than subject 15 (the first column). This is because he/she had an CVD event occurred at month 9, which significantly increases the risk of a new CVD event in the near future. Based on the data up to year 2, the prediction is dynamically updated and the predicted probability of CVD event-free increases for both subjects (the second column) because of no event occurred between year 1 and year 2. However, for subject 127, the predicted probability of CVD event-free after years 3 and 4 decreases sharply due to multiple CVD event occurrences (the third and fourth columns). Clinicians may closely monitor this subject and administer personalized treatments based on the prediction.

6. Discussion

In this article, we propose a joint model of a longitudinal outcome process and a recurrent event process. Two processes are correlated via a shared latent function with an association parameter quantifying the strength of correlation. The simulation study suggests that Bayesian methods using both regular MCMC and embarrassingly parallel MCMC (EP-MCMC) provide valid inference and prediction results while the latter one grants significant computational efficiency. We apply our proposed joint model to the ALLHAT study. Our model identifies multiple risk factors for systolic blood pressure (SBP) and risk of recurrent cardiovascular disease (CVD) events (including CHD, stroke, and heart failure) and the findings are consistent with the primary ALLHAT publication. Our model can efficiently utilize the longitudinal SBP measurements, as well as the recurrent CVD events history to make correct predictions for new subjects. When new measurements are available, predictions can be dynamically updated and become more accurate and efficient. In comparison to a recurrent event model, our model has significantly better performance in predicting the risk of having a new recurrent event, suggesting that using longitudinal SBP information improves the prediction. To address the computation issue associated with large sample size of the ALLHAT study, we employ EP-MCMC method which significantly reduces the computing time.

There are some future directions that we would like to pursue. One limitation of our model is that death is treated as non-informative censoring. This assumption might be strong because death may be related to the SBP measures and the recurrent CVD events. In the future work, we will extend our model to account for the correlated death process. Another limitation is that the longitudinal trajectory is modeled via parametric linear mixed effects submodel, which may not be ideal under strong departure from the prespecified parametric trends. In such situations, nonparametric methods is more effective and preferable [31], e.g. principal analysis by conditional expectation (PACE). The occurrence of CVD events are usually correlated to multiple longitudinal measures (e.g. serum cholesterol level and serum creatinine). Multivariate Functional Principal Component Analysis (MFPCA) method [32,33] is a promising way to incorporate multiple longitudinal variables in predicting the recurrence of CVD events. By leveraging other longitudinal outcomes, we may be able to improve the predictive accuracy. Moreover, genome-wide

association study (GWAS) data have been conducted in many CVD studies, incorporating genetic information in our prediction framework may improve the predictive accuracy.

Acknowledgments

The authors acknowledge the Texas Advanced Computing Center (TACC) for providing high-performing computing resources.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Sheng Luo's research was supported in part by National Institutes of Health [grant numbers R01NS091307 and R56AG064803].

Notes on contributors

Xuehan Ren is a Manager in the Biostatistics Department at Gilead Sciences in Foster City, CA, USA.

Jue Wang is Senior Statistical Scientist at Genetech Inc.

Sheng Luo is Professor of Biostatistics and Bioinformatics, Duke University.

ORCID

Sheng Luo http://orcid.org/0000-0003-4214-5809

References

[1] Thayer JF, Yamamoto SS, Brosschot JF. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int J Cardiol. 2010;141(2):122–131.
[2] Benjamin EJ, Virani SS, Callaway CW, et al. Heart disease and stroke statistics-2018 update: a report from the American Heart Association. Circulation. 2018;137(12):e67–e492.
[3] Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. Int J Epidemiol. 2015;44(1):324–333.
[4] Trial PHA. Major outcomes in high-Risk hypertensive patients. J Am Med Assoc. 2002;288(23):2981–2997.
[5] Kannel WB. Blood pressure as a cardiovascular risk factor: prevention and treatment. J Am Med Assoc. 1996;275(20):1571–1576.
[6] Kannel WB, McGee DL. Diabetes and cardiovascular disease: the Framingham study. J Am Med Assoc. 1979;241(19):2035–2038.
[7] Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. Biostatistics. 2000;1(4):465–480.
[8] Han J, Slate EH, Peña EA. Parametric latent class joint model for a longitudinal biomarker and recurrent events. Stat Med. 2007;26(29):5285–5302.
[9] Crowther MJ, Abrams KR, Lambert PC. Joint modeling of longitudinal and survival data. Stata J. 2013;13(1):165–184.
[10] Brown ER, Ibrahim JG. Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. Biometrics. 2003;59(3):686–693.
[11] Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. J R Stat Soc: Ser C (Appl Stat). 2009;58(1):65–81.

[12] Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-Event data. Biometrics. 2011;67(3):819–829.
[13] Król A, Ferrer L, Pignon JP, et al. Joint model for left-censored longitudinal data, recurrent events and terminal event: predictive abilities of tumor burden for cancer evolution with application to the FFCD 2000–05 trial. Biometrics. 2016;72(3):907–916.
[14] Musoro J, Struijk G, Geskus R, et al. Dynamic prediction of recurrent events data by landmarking with application to a follow-up study of patients after kidney transplant. Stat Methods Med Res. 2018;27(3):832–845.
[15] Iadecola C, Gorelick PB. Hypertension, angiotensin, and stroke: beyond blood pressure. Stroke. 2004;35(2):348–350.
[16] Staessen JA, Thijs L, Fagard R, et al. Predicting cardiovascular risk using conventional vs ambulatory blood pressure in older patients with systolic hypertension. J Am Med Assoc. 1999;282(6):539–546.
[17] Blanche P, Proust-Lima C, Loubère L, et al. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. Biometrics. 2015;71(1):102–113.
[18] Damen JA, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. Br Med J. 2016;353:i2416.
[19] Giorda CB, Avogaro A, Maggini M, et al. Recurrence of cardiovascular events in patients with type 2 diabetes: epidemiology and risk factors. Diabetes Care. 2008;31(11):2154–2159.
[20] Lawless J, Zhan M. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. Can J Stat. 1998;26(4):549–565.
[21] Gilks WR, Best N, Tan K. Adaptive rejection metropolis sampling within Gibbs sampling. J R Stat Soc Ser C (Appl Stat). 1995;44(4):455–472.
[22] Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res. 2014;15(1):1593–1623.
[23] Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci. 1992;7(4):457–472.
[24] Welling M, Teh YW. Bayesian learning via stochastic gradient Langevin dynamics. In: Proceedings of the 28th international conference on machine learning (ICML-11); Bellevue, WA, USA; 2011. p. 681–688.
[25] Quiroz M, Kohn R, Villani M, et al. Speeding up MCMC by efficient data subsampling. J Am Stat Assoc. 2019;114(526):831–843.
[26] Wang X, Guo F, Heller KA, et al. Parallelizing MCMC with random partition trees. In: Advances in neural information processing systems; Montreal, QC, Canada; 2015. p. 451–459.
[27] Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000;56(2):337–344.
[28] Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics. 2005;61(1):92–105.
[29] Proust-Lima C, Séne M, Taylor JM, et al. Joint latent class models for longitudinal and time-to-event data: a review. Stat Methods Med Res. 2014;23(1):74–90.
[30] Kaplan NM, Vidt DG. Major cardiovascular events in hypertensive patients randomized to doxazosin versus chlorthalidone. Curr Hypertens Rep. 2000;2(5):431–431.
[31] Yao F. Functional principal component analysis for longitudinal and survival data. Stat Sin. 2007;17(3):965–983.
[32] Happ C, Greven S. Multivariate functional principal component analysis for data observed on different (dimensional) domains. J Am Stat Assoc. 2018;113(522):649–659.
[33] Li C, Xiao L, Luo S. Fast covariance estimation for multivariate sparse functional data. Stat. 2019;OnlineFirst.

Appendix

We can derive $\pi_i(t'|t)$ as follows:

$$
\begin{aligned}
\pi_i(t'|t) &= \int\int P(T_{i,n_i+1} \le t' | T_{i,n_i+1} > t, \boldsymbol{u}_i, v_i, \boldsymbol{y}_i(t), \boldsymbol{\theta}) \\
&\quad \times P(\boldsymbol{u}_i | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}, v_i)\, \mathrm{d}\boldsymbol{u}_i P(v_i | T_{i,n_i+1} > t, \boldsymbol{\theta})\, \mathrm{d}v_i \\
&= \int\int P(T_{i,n_i+1} \le t' | T_{i,n_i+1} > t, \boldsymbol{u}_i, v_i, \boldsymbol{\theta}) \\
&\quad \times P(\boldsymbol{u}_i | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}, v_i)\, \mathrm{d}\boldsymbol{u}_i P(v_i | T_{i,n_i+1} > t, \boldsymbol{\theta})\, \mathrm{d}v_i \\
&= \int\int \frac{P(t < T_{i,n_i+1} \le t' | \boldsymbol{u}_i, v_i, \boldsymbol{\theta})}{P(T_{i,n_i+1} > t | \boldsymbol{u}_i, v_i, \boldsymbol{\theta})} \\
&\quad \times P(\boldsymbol{u}_i | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}, v_i)\, \mathrm{d}\boldsymbol{u}_i P(v_i | T_{i,n_i+1} > t, \boldsymbol{\theta})\, \mathrm{d}v_i \\
&= \int\int \frac{P(T_{i,n_i+1} > t | \boldsymbol{u}_i, v_i, \boldsymbol{\theta}) - P(T_{i,n_i+1} > t' | \boldsymbol{u}_i, \boldsymbol{\theta})}{P(T_{i,n_i+1} > t | \boldsymbol{u}_i, v_i, \boldsymbol{\theta})} \\
&\quad \cdot P(\boldsymbol{u}_i | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}, v_i)\, \mathrm{d}\boldsymbol{u}_i P(v_i | T_{i,n_i+1} > t, \boldsymbol{\theta})\, \mathrm{d}v_i \\
&= 1 - \int\int \frac{P(T_{i,n_i+1} > t' | \boldsymbol{u}_i, v_i, \boldsymbol{\theta})}{P(T_{i,n_i+1} > t | \boldsymbol{u}_i, v_i, \boldsymbol{\theta})} P(\boldsymbol{u}_i | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}, v_i)\, \mathrm{d}\boldsymbol{u}_i P(v_i | T_{i,n_i+1} > t, \boldsymbol{\theta})\, \mathrm{d}v_i \\
&= 1 - \int\int \frac{S(t' | \boldsymbol{u}_i, v_i, \boldsymbol{\theta})}{S(t | \boldsymbol{u}_i, v_i, \boldsymbol{\theta})} P(\boldsymbol{u}_i | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}, v_i)\, \mathrm{d}\boldsymbol{u}_i P(v_i | T_{i,n_i+1} > t, \boldsymbol{\theta})\, \mathrm{d}v_i \\
&= 1 - \int\int \exp\left[-\int_t^{t'} r_i(s | \boldsymbol{u}_i, v_i, \boldsymbol{\theta})\, \mathrm{d}s \right] P(\boldsymbol{u}_i | T_{i,n_i+1} > t, \boldsymbol{y}_i(t), \boldsymbol{\theta}, v_i) \\
&\quad \times \mathrm{d}\boldsymbol{u}_i P(v_i | T_{i,n_i+1} > t, \boldsymbol{\theta})\, \mathrm{d}v_i \\
&\approx \frac{1}{M} \sum_{m=1}^{M} 1 - \exp\left[-\int_t^{t'} r_i^{(m)}(s | \boldsymbol{u}_i^{(m)}, v_i^{(m)}, \boldsymbol{\theta}^{(m)})\, \mathrm{d}s \right].
\end{aligned}
$$

Here $\boldsymbol{\theta}^{(m)}$ is the mth sample ($m = 1, \ldots, M$, where M is the number of post burn-in samples) of parameter vector $\boldsymbol{\theta}$ and $\boldsymbol{u}_i^{(m)}$, $v_i^{(m)}$ is the mth sample of random effects $\boldsymbol{u}_i$ and v_i respectively.