To appear in the *Journal of Statistical Computation and Simulation*
Vol. 00, No. 00, Month 20XX, 1–20

Joint Models for Multiple Longitudinal Processes and Time-to-event Outcome

Lili Yang, Menggang Yu, & Sujuan Gao

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

Joint models are statistical tools for estimating the association between time-to-event and longitudinal outcomes. One challenge to the application of joint models is its computational complexity. Common estimation methods for joint models include a two-stage method, Bayesian and maximum-likelihood methods. In this work, we consider joint models of a time-to-event outcome and multiple longitudinal processes and develop a maximum-likelihood estimation method using the expectation-maximization (EM) algorithm. We assess the performance of the proposed method via simulations and apply the methodology to a data set to determine the association between longitudinal systolic and diastolic blood pressure (BP) measures and time to coronary artery disease (CAD).

Keywords: joint models, EM algorithm, simulation, multiple longitudinal outcomes, time-to-event outcome

1. Introduction

Prospective cohort studies or clinical trials with time-to-event as the primary outcome usually collect many longitudinal variables. Longitudinal studies of Alzheimer's disease, for example collect repeated measures of height, weight, blood pressure (BP) measures and many other variables in order to determine disease etiology [1]. Clinical trials on cardiovascular diseases routinely monitor BPs at regular intervals to ensure patient safety [2]. In addition, the increasing use of electronic medical records (EMR) in many health care systems makes the collection of many longitudinal laboratory measures and time to medical events automatic and straightforward. Separate modeling of the longitudinal processes and the survival outcome may not fully discover potential disease mechanisms. Appropriate statistical methods are needed to utilize the richness of these data in order to identify potential relationships between the longitudinal measures and disease risk.

Many epidemiologic studies have mostly adopted the Cox model [3] using baseline exposure measures. Such an approach implicitly assumes that the exposure variables stay constant over the length of the study, which is unlikely to be true in studies over an extensive period of time. Cox model with observed longitudinal measures as time-dependent covariates [4–6] incorporates changes in exposure levels over the follow-up period. However, this model assumes that the longitudinal outcomes are continuously measured without errors. This assumption may not be realistic when the longitudinal measures are intermittently collected. Furthermore, measurement errors in the longitudinal measurements were not considered in this modeling framework. Lastly, the time-dependent Cox model lacks the flexibility to use various functional forms of the underlying longitudinal processes.

To overcome these difficulties, joint models of longitudinal and survival outcomes were proposed by Faucett and Thomas [7] and Wulfsohn and Tsiatis [8]. They used a linear growth curve model for the longitudinal process and a Cox model with the current value of the longitudinal process as time-dependent covariate. Many extensions to these earlier joint models have been proposed. Henderson et al. modeled the hazard as a function of the history and rate of change of a biomarker [9]. Brown and Ibrahim extended the linear

growth curve model to flexible non-parametric subject-specific random-effects models [10, 11]. Yu et al. considered a survival-cure model for the time-to-event outcome [12]. Huang et al. and Elashoff et al. extended the Cox model to competing risks models [13, 14]. Chi and Ibrahim extended the joint models from a single survival outcome to multiple survival outcomes [15]. Njeru Njagi et al. considered combining conjugate and normal random effects of longitudinal and time-to-event outcomes in joint models to improve model fit [16]. Qiu et al. considered a generalized linear mixed model for the longitudinal outcome and a discrete survival model with frailty to predict event probabilities [17]. Comprehensive reviews of joint models have been published [12, 18–20].

When multiple longitudinal measures are available, extension to the joint model framework needs to appropriately account for potential correlations among the longitudinal measures. Simultaneous modeling of multiple longitudinal outcomes in joint models offers a number of advantages over separate modeling of each longitudinal outcome [11, 15, 21–23]. First, for correlated longitudinal outcomes it is more relevant to estimate the adjusted association of each longitudinal outcome with the event risk [21]. Second, Fieuws et al. showed that accounting for the correlation between longitudinal measures may substantially enhance the predictive ability of joint models [24]. In addition, two studies found that joint models of multiple longitudinal outcomes are more efficient compared with separate modeling of each outcome in some settings [25, 26].

There are three general types of estimation methods in joint models of longitudinal and survival outcomes: a two-stage approach, Bayesian Markov Chain Monte Carlo (MCMC) method, and maximum-likelihood approach. In the two-stage approach, parameter estimation is conducted separately for the longitudinal model and the survival model. Specifically, at the first stage, parameter estimates and predictions are obtained from the longitudinal models without consideration of the survival outcomes. At the second stage, predicted longitudinal values are used as true exposure levels in a time-dependent Cox model. Although the two-stage approach is computationally simple, it can incur bias and loss of efficiency by ignoring the time-to-event information when modeling the longitudinal process [7, 27, 28], as the survival process in this setting essentially produces non-ignorable missing data for the longitudinal outcomes. The two-stage approach has been discussed by many authors [27–32]. Alternatively, both the Bayesian MCMC approach and the maximum-likelihood approach incorporate both types of outcomes into a joint likelihood function and simultaneously estimate model parameters. The Bayesian MCMC approach has been used for joint models of multiple longitudinal and time-to-event outcomes [10, 11, 15, 21, 33]. Bayesian MCMC approach can be relatively easier to implement. On the other hand, maximum likelihood based approach enables rigorous study of asymptotic properties. When sample size is small, the large sample approximation of maximum-likelihood approach may not be accurate. But Bayesian approach may not provide valid quantitation either because the results may reply too much on correct or tight prior specification. In the situation of many parameters in complex models, autocorrelation and convergence can be a challenge with the MCMC approach. To the best of our knowledge, the maximum-likelihood method has only been applied to the joint models with a single longitudinal outcome [8, 13, 18, 34, 35]. In particular, Rizopoulos developed an R package (JM) using the EM algorithm [36] for joint models of a time-to-event outcome and a single longitudinal outcome [37, 38].

In this work, we develop a maximum-likelihood approach using the EM algorithm for parameter estimation in joint models of multiple longitudinal processes and a time-to-event outcome. Commonly used for maximum-likelihood estimation, the EM algorithm offers computational advantages over direct likelihood maximization especially in complex likelihood functions involving random effects [39]. Direct likelihood maximization

method can be useful for joint model estimations with few random effects, simple correlation structure of random effects, or small number of unknown parameters. However, in many joint models including ours, often times, the model structure may be too complex to use direct likelihood maximization. We chose the EM approach over direct likelihood maximization based on the following reasons. For complex joint models, the EM approach can break down the optimization of the whole likelihood function into several M-steps, which makes the optimization more feasible and stable. The EM algorithm increases the likelihood function as iteration continues, ensuring numerical stability. Additional efficiency can be gained when some parameters have closed-form solutions in the M-step. Finally, predicted values are calculated as part of the E-step reducing the need for further computation.

The remainder of this article is organized as follows: in Section 2 we describe data from a primary care patient cohort with extensive EMR data as a motivating example. In Section 3 we describe the joint model framework as well as the joint likelihood function. Estimation method using the EM algorithm and asymptotic inferences of parameter estimates are described in Section 4. In Section 5 we report results from simulation studies. In Section 6 we illustrate the proposed method using data from the primary care patient cohort. Conclusions and discussions are presented in Section 7.

2. A Primary Care Patient Cohort

A primary care patient cohort was assembled in 1991 as part of a depression screening study in primary care clinics in Wishard Health Service. From 1991 to 1993, patients age 60 years or older in the Wishard Health Service were consented for depression screening during their regular clinical visits to their primary care physicians. A total of 4,413 primary care patients were initially contacted, of whom 115 refused; 57 were not eligible due to severe cognitive impairment; 284 were not eligible because they were non-English speaking, in prison, in a nursing home, or had a hearing impairment; 3,957 patients were enrolled in the study. Details about the study have been published [40, 41].

EMR data are available for all enrolled patients and the information includes diagnosis of medical conditions, BP measures, laboratory test measures and medication order and dispensing. One of the research interests using data from this cohort is to examine new risk factors for coronary artery disease (CAD) in elderly population. It is well known from the results of prospective cohort studies that high baseline BP is a risk factor for CAD in middle-aged populations [42–44], but few studies have determined the relationship between longitudinal BP measures and the risk of CAD. Even fewer focused on the elderly population who have declining BP with increasing age. Therefore, it is necessary to apply joint models to determine the association between the longitudinal BP measures and risk of CAD in this elderly cohort.

Among the 3,957 patients enrolled, 2,654 (797 males and 1857 females) were free of CAD at enrollment. For patients with incident CAD events, the date of diagnosis was used as the event time; for patients without CAD, the last outpatient clinic visit before December 31, 2010 was used as the censoring time. Systolic and diastolic BP measured in sitting position from outpatient clinic visits were also collected during follow-up for up to 20 years. Since it has been shown that males have significantly increased CAD risk than females [45, 46], we focus our analysis on the 797 male patients in the cohort where 28% had incident CAD during the follow-up period from enrollment to December 31, 2010. Mean age of patients included in the analysis sample at baseline was 68 (SD=7.4) years, 519 (65.1%) were black, 254 (31.9%) were smokers, and 268 (33.6%) had history of diabetes at baseline. The frequency of BP measurements varied from patient to patient with a mean frequency of 20.5 (SD=20). For computational convenience annualized systolic

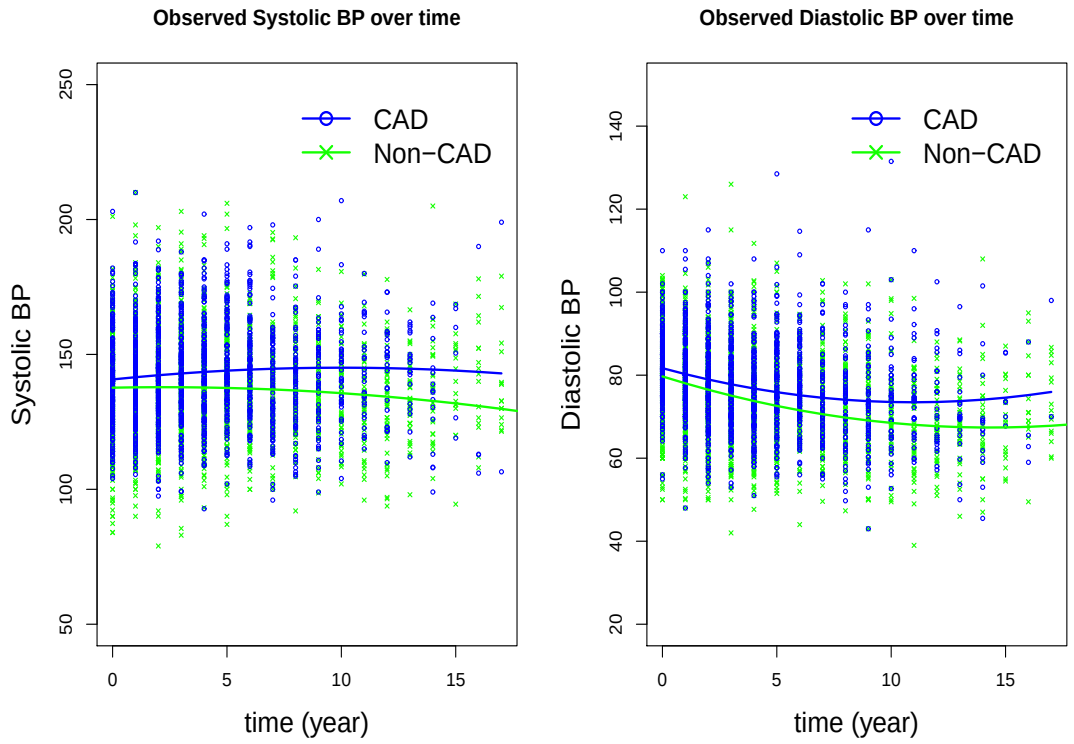

Figure 1. Observed annualized longitudinal systolic and diastolic BP measures over time and fitted population mean curves for the CAD and non-CAD group.

and diastolic BP measures during the study period were derived for each participant. First, we aligned the first BP measures for each patient and set the time point as t=0 (year). Then we continued to create yearly time points for each subject with the event or censoring time as the last time point and calculated the mean BP measures within each yearly interval for each subject. On average, there were about 5.3 (SD=4.4) annualized BP measures per subject. Figure 1 plots the annualized longitudinal systolic and diastolic BP measures over time by CAD status. The blue and green curves represent fitted population mean BP profiles for CAD and non-CAD groups respectively, using linear mixed-effects models with fixed quadratic time effect. It can be seen that the population mean systolic and diastolic BPs were higher over time for the CAD group than that for the non-CAD group, indicating a potential association between the risk of CAD and longitudinal systolic and diastolic BP measures. The figure also shows that the differences in BP measures at baseline between the CAD and non-CAD groups were negligible. Thus analyses relying on baseline BP may not be able to detect any relationship between BP measures and risk of CAD.

3. Joint Models

In this section, we introduce joint models for multiple longitudinal processes and a time-to-event outcome by defining the notations and formulation of the longitudinal and sur-

vival models. Specifically, we consider multivariate mixed-effects models for the multiple longitudinal outcomes and a Cox model for the time-to-event outcome with true underlying functions of the longitudinal measures as time-dependent covariates. We then derive the likelihood function of the joint models.

3.1 *Longitudinal Models*

Let $y_l(t_{ij})$ denote the observed measurement of the l-th longitudinal outcome for subject i at time points t_{ij}, where $i = 1, ..., n$, $j = 1, ..., n_i$, $l = 1, ..., L$. n denotes the number of subjects; n_i is the number of longitudinal repeat measures for each subject; L is the number of longitudinal outcomes in the model. The corresponding longitudinal trajectory is modeled using the following model

$$\begin{aligned} \mathbf{y}_l(\mathbf{t}_i) &= \mathbf{y}_l^*(\mathbf{t}_i) + \boldsymbol{\epsilon}_{il}, \\ &= \mathbf{X}_l^T(\mathbf{t}_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(\mathbf{t}_i)\mathbf{b}_{il} + \boldsymbol{\epsilon}_{il} \end{aligned} \tag{1}$$

where $\mathbf{y}_l^*(\mathbf{t}_i) = (y_l(t_{i1}), y_l(t_{i2}), ..., y_l(t_{in_i}))^T$ is the corresponding true underlying longitudinal measures of the l-th biomarker for the i-th subject; $\mathbf{X}_l^T(\mathbf{t}_i)$ is the design matrix of fixed effects, including time effects and baseline covariates; $\boldsymbol{\beta}_l$ is the corresponding vector of the fixed effects; $\mathbf{Z}_l^T(\mathbf{t}_i)$ is the design matrix for the random effects, $\mathbf{b}_{il}$, distributed as $\mathbf{b}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, ..., \mathbf{b}_{iL})^T \sim N(\mathbf{0}, \mathbf{D})$; $\boldsymbol{\epsilon}_{il}$ is the corresponding measurement error term such that $\boldsymbol{\epsilon}_{il} \sim^{iid} N(0, \sigma_l^2 I_{n_i})$. It is worth noting that the correlations among the multiple longitudinal processes and the within-subject correlation for each longitudinal biomarker are represented in the variance-covariance matrix of random effects $\mathbf{D}$. We assume that the measurement errors of different longitudinal outcomes are independent of each other, and they are also independent of the random effects $\mathbf{b}_i$.

3.2 *The Survival Model*

Let T_i^* and C_i be the true event time and censoring time respectively for subject i. We define the observed event time $T_i = \min(T_i^*, C_i)$ and the event indicator $\delta_i = I(T_i^* \leq C_i)$. Assuming that the hazard function depends on some functions of the true longitudinal measures $\mathcal{F}(\mathbf{y}_{il}^*(t))$ and baseline covariates $\mathbf{w_i}$, the hazard function can be written as

$$h(t_i) = h_0(t) \exp\left\{ \boldsymbol{\gamma}^T \mathbf{w_i} + \sum_{l=1}^{L} \alpha_l \mathcal{F}(\mathbf{y}_l^*(t_i)) \right\}, \tag{2}$$

where $h_0(t)$ denotes the baseline hazard function, and α_l and $\boldsymbol{\gamma}$ are coefficients for the function of lth biomarker and baseline risk factors. The baseline hazard function can be a parametric function or a flexible piecewise constant function. In this work, α_l, $l = 1, 2, ..., L$, are of primary interest. The correlation between the multiple longitudinal biomarkers and the time-to-event outcome is induced by the shared random effects through $y_l^*(t_i)$ or b_{il} in the longitudinal and survival models.

The function $\mathcal{F}(\cdot)$ can be chosen as different functional forms depending on the interest of the study. For example, if the focus is the association between longitudinal values and event risk, $\mathcal{F}(\cdot)$ can be an identity function; if the change in the longitudinal measures is of interest, $\mathcal{F}(\cdot)$ can be chosen as derivative function with respect to time t; for studies interested in the cumulative history of the longitudinal measures over time and event risk, $\mathcal{F}(\cdot)$ can be an integration function of $\mathbf{y}_l^*(t_i)$ over time t. Depending on the choices, random effects b_i may affect the hazard function in a non-linear fashion. This is in

contrast with the frailty type of joint models where the random effects are linear in the exponential term of the hazard.

3.3 *Joint Likelihood Function*

Under the conditional independence assumption between the random effects $\mathbf{b}_i$ and measurement errors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, ..., \epsilon_{il})^T$ of the longitudinal outcomes, the kernel of the joint likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} p(T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}|\boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} \int p(T_i, \delta_i|\boldsymbol{\theta}, \mathbf{b}_i) p(\mathbf{y}_{i1}|\boldsymbol{\theta}, \mathbf{b}_i) \cdots p(\mathbf{y}_{iL}|\boldsymbol{\theta}, \mathbf{b}_i) p(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i, \end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\theta}_{h_0}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{D})^T$ is the vector containing all parameters in the models with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_l)^T$, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_l)^T$. Under the assumed models (1) and (2) for the longitudinal and survival outcomes, there are three components:

$$\begin{aligned} p(T_i, \delta_i|\boldsymbol{\theta}, b_i) = &\left\{ h_0(T_i) \exp\left(\mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il} \Big) \right) \right\}^{\delta_i} \\ &\exp\left\{ - \int_0^{T_i} h_0(u) \exp\left(\mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_l^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(u)\mathbf{b}_{il} \Big) \right) du \right\}, \end{aligned}$$

$$\begin{aligned} &p(\mathbf{y}_{i1}|\boldsymbol{\theta}, \mathbf{b}_i) \cdots p(\mathbf{y}_{iL}|\boldsymbol{\theta}, \mathbf{b}_i) \\ &= \prod_{l=1}^{L} \left(\frac{1}{\sqrt{2\pi\sigma_l^2}} \right)^{n_i} \exp\left\{ -\frac{1}{2\sigma_l^2} \sum_{j=1}^{n_i} \Big(\mathbf{y}_l(t_{ij}) - \Big(\mathbf{X}_l^T(t_{ij})\boldsymbol{\beta}_l + \mathbf{Z}_l^T(t_{ij})\mathbf{b}_{il} \Big) \Big)^2 \right\}, \end{aligned}$$

and

$$p(\mathbf{b}_i|\boldsymbol{\theta}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{k/2} |\mathbf{D}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right),$$

where k is the dimension of the $\mathbf{D}$ matrix.

4. Estimation Method

In this section, we present the maximum-likelihood method based on the EM algorithm for parameter estimation and inference.

4.1 *Implementing the EM Algorithm*

The complete log likelihood function, given the random effects $\mathbf{b}_i$ is:

$$\log L_C = \sum_{i=1}^{n} \Big\{ \log p(T_i, \delta_i|\mathbf{b}_i;\boldsymbol{\theta}) + \log p(\mathbf{y}_{i1}|\mathbf{b}_i;\boldsymbol{\theta}) + ... + \log p(\mathbf{y}_{iL}|\mathbf{b}_i;\boldsymbol{\theta}) + \log p(\mathbf{b}_i;\boldsymbol{\theta}) \Big\},$$

where L_C denotes the complete joint likelihood function.

In the E-step the expected complete log-likelihood function given the conditional distribution of random effects is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^m) = \sum_{i=1}^{n} \int \Big\{ \log p(T_i, \delta_i|\mathbf{b}_i;\boldsymbol{\theta}) + \log p(\mathbf{y}_{i1}|\mathbf{b}_i;\boldsymbol{\theta}) + ... + \log p(\mathbf{y}_{iL}|\mathbf{b}_i;\boldsymbol{\theta}) + \log p(\mathbf{b}_i;\boldsymbol{\theta}) \Big\} p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m) d\mathbf{b}_i.$$

where

$$p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m) = \frac{p(\mathbf{b}_i, T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m)}{p(T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m)}.$$

For the M-step, closed-form expressions are available for the variance of residuals of each longitudinal model and variance-covariance matrix of the random effects, whereas the fixed effects for each longitudinal model and parameters in the survival model have to be estimated numerically. The key steps are:

(1) Estimation of the variance of residuals of each longitudinal model by

$$\begin{aligned}\hat{\sigma_l}^2 = \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} &\Big(\mathbf{y}_{il} - \mathbf{X}_{il}^T\boldsymbol{\beta}_l^m\Big)^T \Big(\mathbf{y}_{il} - \mathbf{X}_{il}^T\boldsymbol{\beta}_l^m - 2\mathbf{Z}_{il}^T \mathrm{E}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m)\Big) \\ &+\mathrm{Tr}\Big(\mathbf{Z}_{il}^T\mathbf{Z}_{il}\mathrm{Var}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m)\Big) \\ &+\mathrm{E}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m)^T\mathbf{Z}_{il}^T\mathbf{Z}_{il}\mathrm{E}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m)\end{aligned}$$

where $l = 1, ..., L$, Tr represents the trace function of a matrix, and E denotes the expectation function.

(2) Estimation of variance-covariance matrix of random effects by

$$\begin{aligned}\hat{\mathbf{D}} = \frac{1}{n}\sum_{i=1}^{n} &\mathrm{Var}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m) \\ &+\mathrm{E}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m)\mathrm{E}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m)^T\end{aligned}$$

(3) Since the fixed effect coefficients of each longitudinal model, $\boldsymbol{\beta}_l$, are involved in both the longitudinal and survival models, there is no closed form solution. The one-step Newton-Raphson algorithm can be implemented to update $\boldsymbol{\beta}_l$:

$$\hat{\boldsymbol{\beta}}_l^{m+1} = \hat{\boldsymbol{\beta}}_l^m - \Big(\partial S(\hat{\boldsymbol{\beta}}_l^m)/\partial\boldsymbol{\beta}_l\Big)^{-1} S(\hat{\boldsymbol{\beta}}_l^m), l = 1, ..., L,$$

where the score functions are

$$S(\boldsymbol{\beta}_l) = \sum_{i=1}^{n} \frac{1}{\sigma_l^2} \mathbf{X}_{il}^T \Big(\mathbf{y}_{il} - \mathbf{X}_{il}^T \boldsymbol{\beta}_l - \mathbf{Z}_{il}^T \mathrm{E}(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) \Big)$$
$$+ \delta_i \alpha_l \frac{\partial \mathcal{F}\Big(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il} \Big)}{\partial \boldsymbol{\beta}_l}$$
$$- \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} h_0(u) \frac{\partial \exp\Big(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i \Big) \Big)}{\partial \boldsymbol{\beta}_l}$$
$$p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i,$$

for $l = 1, ..., L$. The derivatives $\partial S(\hat{\boldsymbol{\beta}}_l^m)/\partial \boldsymbol{\beta}_l$ can be calculated by numerical approximations.

(4) Parameters in the survival model can be similarly updated using the Newton-Raphson algorithm. The baseline hazard function is estimated using piecewise constant function. Score equations used in the Newton-Raphson algorithm are:

$$S(\gamma) = \sum_{i=1}^{n} \mathbf{w}_i \Bigg\{ \delta_i - \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} h_0(u) \exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i \Big) \right)$$
$$p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i \Bigg\}.$$

$$S(\alpha_l) = \sum_{i=1}^{n} \delta_i \int \mathcal{F}\Big(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il} \Big) d\mathbf{b}_i$$
$$- \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} h_0(u) \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i \Big)$$
$$\exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i \Big) \right) p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i,$$

where $l = 1, ..., L$, and

$$S(\boldsymbol{\theta}_{h_0}) = \sum_{i=1}^{n} \delta_i \frac{1}{h_0(T_i; \boldsymbol{\theta}_{h_0})} \frac{\partial h_0(T_i; \boldsymbol{\theta}_{h_0})}{\partial \boldsymbol{\theta}_{h_0}^T}$$
$$- \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} \frac{\partial h_0(T_i; \boldsymbol{\theta}_{h_0})}{\partial \boldsymbol{\theta}_{h_0}^T} \exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i \Big) \right)$$
$$p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i.$$

For computation of the expected likelihood function, a pseudo-adaptive Gaussian-Hermit quadrature rule [35] can be used to approximate the integrals. The E-step and M-step iterate until a pre-specified convergence criterion is met.

4.2 *Inferences and Goodness-of-fit*

After convergence of the EM algorithm, standard errors of estimated parameters can be calculated from the observed inverse Hessian matrix, which can be derived from the score functions using numeric approximations [47]. Using the properties of consistency and asymptotic normality of maximum-likelihood estimates, a 95% confidence interval (CI) of a parameter θ can be derived as:

$$\hat{\theta} \pm 1.96 \times s.e.(\hat{\theta}).$$

For hypothesis tests involving multiple parameters, the Wald test or Score test can be used. The Likelihood Ratio Test is a commonly used statistic to compare nested models. For non-nested models, the Akaike's Information Criterion (AIC) [48] is often used for model selection:

$$\text{AIC} = -2l(\hat{\boldsymbol{\theta}}) + 2k.$$

where k denotes the number of parameters in the model. Alternatively, one can adopt Bayesian Information Criterion (BIC) [49].

5. Simulation Study

We performed Monte Carlo (MC) simulations to assess the performance of the proposed method and compare with the two-stage method. We simulated data from joint models with two correlated normally distributed longitudinal variables and a time-to-event variable. For each MC data set, longitudinal data were simulated for 500 subjects each with 10 equally spaced bivariate longitudinal measures over a 5-year period. We considered similar fixed and random model structures for the two longitudinal outcomes, where the fixed effects included time effect, and one binary baseline covariate, and the random effects included random intercept and random slope. The correlation between the two longitudinal outcomes is represented by the correlation between the two random intercepts. We assumed it is an unstructured variance-covariance matrix when fitting the model. The longitudinal models are:

$$y_1(t_{ij}) = y_1^*(t_{ij}) + \epsilon_1(t_{ij}) = \beta_{01} + \beta_{11}t_{ij} + \beta_{21}w_i + b_{01i} + b_{11i}t_i + \epsilon_1(t_{ij}),$$

$$y_2(t_{ij}) = y_2^*(t_{ij}) + \epsilon_2(t_{ij}) = \beta_{02} + \beta_{12}t_{ij} + \beta_{22}w_i + b_{02i} + b_{12i}t_i + \epsilon_2(t_{ij}),$$

where $\epsilon_1(t_{ij}) \sim N(0, \sigma_1^2)$, $\epsilon_2(t_{ij}) \sim N(0, \sigma_2^2)$ and

$$(b_{01i}, b_{11i}, b_{02i}, b_{12i})^T \sim N\left(\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{01}^2 & 0 & \rho\sigma_{01}\sigma_{02} & 0\\ 0 & \sigma_{11}^2 & 0 & 0\\ \rho\sigma_{01}\sigma_{02} & 0 & \sigma_{02}^2 & 0\\ 0 & 0 & 0 & \sigma_{12}^2\end{pmatrix}\right).$$

The time-to-event endpoint was simulated from a Cox model with a Weibull baseline hazard function, $h_0(t) = abt^{b-1}$, where a and b are the rate and shape parameters respectively. The Cox model is assumed to depend on the current values of the two longitudinal

outcomes and can be expressed as

$$h(t) = abt^{b-1} \exp(\alpha_1 y_1^*(t) + \alpha_2 y_2^*(t)).$$

To simulate the event times, we first simulated a survival probability, s_i, from Uniform(0,1) for each subject and then solved for T_i^* using two R functions [50], *integrate*() and *uniroot*(), from the following equation:

$$s_i - \exp\left\{ - \int_0^{T_i^*} abu^{b-1} \exp(\alpha_1 y_1^*(u) + \alpha_2 y_2^*(u)) du \right\} = 0.$$

Censoring times were independently simulated from another uniform distribution. Overall, the censoring percentage is about 30%. Because of censoring, there were about 6 repeated bivariate measurements per subject. In the actual model fitting, we used a more flexible piecewise constant baseline hazard function instead of the parametric Weibull baseline risk function.

We considered two simulation scenarios (large variances of residual errors and random effects, and small variances of residual errors and random effects) with parameter values specified as below:

- Small variance scenario : $\sigma_1 = \sigma_2 = 0.1, \sigma_{01} = \sigma_{02} = 0.1, \sigma_{11} = \sigma_{12} = 0.04$;
- Large variance scenario : $\sigma_1 = \sigma_2 = 0.5, \sigma_{01} = \sigma_{02} = 0.5, \sigma_{11} = \sigma_{12} = 0.2$.

Other parameters were set as follows: $\beta_{01} = 0.2$, $\beta_{11} = 0.5$, $\beta_{21} = 0.2$, $\beta_{02} = 1$, $\beta_{12} = 0.2$, $\beta_{22} = 0.5$, $\rho = 0.5$, $a = 0.005$, $b = 1.1$, $\alpha_1 = 1$, and $\alpha_2 = 1.5$.

The parameter estimation using the two-stage method was achieved from the following two steps. First, a model for bivariate longitudinal outcomes was fitted to obtain the parameter estimates in the longitudinal models. The parameter estimation was implemented using the 'lme' function in the 'nlme' R package. In addition to the parameter estimation, we also obtained predicted values for the longitudinal measures at each time point up to the event or censoring time. Second, a time-dependent Cox model using the predicted longitudinal data as time-dependent variables and other baseline variables as independent variables was fitted to obtain the parameter estimation in the survival model. The 'coxph' function in the 'survival' R package was used for this model.

Parameter estimates from the two-stage method were used as the initial parameter values in the EM approach.

Among the 500 replicates of each scenario, the convergence of EM algorithm in this joint model simulation looks good. On average, under the small variance scenario, it takes about 3 minutes to complete the model fit for each MC data set; the EM algorithm always converged given that we allows a large the number of iterations; the average number of iterations required for convergence is 68 (SD=50).

Simulation results are presented in Tables 1 and 2. We reported relative biases, empirical standard errors, model-based standard errors, and coverage probabilities of the 95% CIs based on 500 MC data sets. For the variance-covariance matrix of the random effects, $\mathbf{D}$, we reported the results of its cholesky decomposition, where

$$chol(\mathbf{D}) = \begin{pmatrix} D_{11} & 0 & D_{13} & 0 \\ 0 & D_{22} & 0 & 0 \\ 0 & 0 & D_{33} & 0 \\ 0 & 0 & 0 & D_{44} \end{pmatrix}.$$

In the small variance scenario, both the two-stage approach and the EM algorithm

generally performed well: estimated parameters had small relative bias (defined as $(\boldsymbol{\theta}_{\text{true}} - \hat{\boldsymbol{\theta}})/\boldsymbol{\theta}_{\text{true}}$), and coverage of the 95% CIs were close to nominal levels. However, in the large variance scenario, the two-stage approach performed poorly whereas the EM algorithm maintained good performance with small relative bias and good coverage probabilities of the 95% CIs. In particular, on the two parameters representing the association between the longitudinal measures and the event risk, α_1 and α_2, the two-stage approach generally underestimated the strength of the associations.

6. Data Application

We applied joint models to the aforementioned primary care patient cohort data using 797 male patients. Our focus was to determine whether and how longitudinal BP measures were associated with the time to CAD. For convenience we centered patients' baseline age at 60 years. We fitted four different sets of joint models using the proposed EM algorithm. The best set of models were determined using the AIC. The 4 sets of joint models are as follows.

Joint models 1 consider the following models

$$\begin{aligned} y_l(t_{ij}) &= y_l^*(t_{ij}) + \epsilon_{ijl} \\ &= \beta_{0l} + \beta_{1l}t_{ij} + \beta_{2l}\text{age}_i + \beta_{3l}\text{race}_i + b_{0li} + b_{1li}t_{ij} + \epsilon_{ijl}, l = 1, 2 \end{aligned} \tag{3}$$

for the observed longitudinal systolic and diastolic BP measures respectively. The random effect vector $\mathbf{b}_i = (b_{01i}, b_{11i}, b_{02i}, b_{12i})^T$ is normally distributed with mean zero and an unstructured variance-covariance matrix of $\mathbf{D}$; ϵ_{ij1} and ϵ_{ij2} are independently distributed as $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ respectively. The normality assumption of longitudinal systolic and diastolic BP measures was checked by Q-Q plot using the transformed residuals of error based on the fitted joint model [51]. The hazard function satisfies

$$h(t) = h_0(t) \exp\left\{\gamma_1\text{age}_i + \gamma_2\text{smoke}_i + \gamma_3\text{race}_i + \gamma_4\text{diabetes}_i + \alpha_1 y_{i1}^*(t) + \alpha_2 y_{i2}^*(t)\right\}, \tag{4}$$

where the piecewise constant function $h_0(t)$ consists of 7 equally spaced intervals with 6 interior knots based on percentiles of the observed event time points. The above hazard function depends on the current values of systolic and diastolic BP measures and some baseline risk factors.

Joint models 2 assume the same longitudinal models (3) as in the joint models 1, but use the slopes of systolic and diastolic BP measures in the hazard function instead, that is we assume that the risk for CAD at time t depends on the slope of the true trajectory of systolic and diastolic BP at the same time point as previous research has demonstrated that changes in BP are important predictors of health outcomes [52].

$$h(t) = h_0(t) \exp\left\{\gamma_1\text{age}_i + \gamma_2\text{smoke}_i + \gamma_3\text{race}_i + \gamma_4\text{diabetes}_i + \alpha_1 y_{i1}^{*'}(t) + \alpha_2 y_{i2}^{*'}(t)\right\} \tag{5}$$

Joint models 3 assume the same hazard model (4) in the joint model 1, but include a quadratic fixed time effect for both systolic and diastolic BP measures,

$$\begin{aligned} y_l(t_{ij}) &= y_l^*(t_{ij}) + \epsilon_{ijl} \\ &= \beta_{0l} + \beta_{1l}t_{ij} + \underline{\beta_{2l}t_{ij}^2} + \beta_{3l}\text{age}_i + \beta_{4l}\text{race}_i + b_{0li} + b_{1li}t_{ij} + \epsilon_{ijl}, l = 1, 2. \end{aligned} \tag{6}$$

Joint models 4 assumes (5) and (6).

The parameter estimation steps of the two-stage method described in the simulation study section were applied for the parameter estimation using the two-stage method.

In the implementation of the EM algorithm, we used 3 pseudo-adaptive Gaussian-Hermite quadrature points for numerical integration over the random effects and 7 Gaussian-Kronrod quadrature points for the integration in the survival function. Models were compared according to AIC: smaller AIC indicates better model fit. Among the 4 joint models considered, Joint models 3 was the best fitting (AIC=65786) followed by Joint models 4 (AIC=65798), Joint models 2 (AIC=65885) and Joint models 1 (AIC=65898). Here we present parameter estimates from joint models 3 in Tables 3.

It can be seen that systolic BP measures are significantly associated with the risk of developing CAD. Each 10 unit increase of systolic BP is associated with 1.23-fold increase (95% CI:[1.05, 1.5]) in patient's risk of developing CAD. In addition we observe that diastolic BP measures are not significantly associated with the risk of developing CAD once systolic BPs were adjusted in the model. The fitted model also identified several other risk factors for CAD, i.e. participants with older age, being Caucasian and smokers have higher risk of CAD. The fitted longitudinal quadratic growth models suggested that there is a quadratic increasing-then-decreasing trend for systolic BP measures, whereas a decreasing-then-increasing quadratic trend for diastolic BP measures was seen. In Figure2 we plotted subject-specific fitted curves under fitted Joint model 3 for 4 CAD and 4 non-CAD participants, randomly selected from the study population. It can be seen that the quadratic longitudinal models fit the data relatively well.

In this joint model, we assume that the correlation between systolic and diastolic BP is taken into consideration in the variance-covariance matrix of random effects. The estimated correlation between the random intercepts of systolic and diastolic BP is 0.785 (95% CI [0.718, 0.851]), and the estimated correlation between the random slopes of systolic and diastolic BP is not significant, indicating that the systolic and diastolic BP is highly correlated through their random intercepts.

As a comparison, we also fitted two separate single longitudinal measure joint models, one using systolic BP only and the other diastolic BP only, while adjusting for the same covariates as in (6) above. The two separate joint models showed that both systolic and diastolic BP were significantly associated with CAD risk. Our joint models 3 takes the correlation between the two BP measures into consideration and our results indicate that systolic BP had higher impact on the risk of CAD than diastolic BP in this elderly population.

We also analyzed the data using alternative methods including the two-stage approach, Cox model with baseline BP measures, and Cox model with time-dependent BP measures. In the two Cox models we adjusted for the same baseline risk factors as in hazard model (4). Given our research focus was on whether and how longitudinal BP measures were associated with the time to CAD, the Coxs model using baseline only does not fully utilize the longitudinal BP data and is inadequate in addressing the research question. The time-dependent Coxs model ignores measurement errors in the longitudinal measures and is expected to underestimate the variance of the association parameter. Figure 3 plots the estimated parameter, $\hat{\alpha}_1$, for systolic BP and their corresponding 95% CIs from 4 different methods: the EM algorithm (Joint models 3), the two-stage method (Joint models 3), the Cox model with time-dependent covariates and the Cox model with baseline BP measures. Without knowing the true underlying model, these comparisons do not suggest the effectiveness of our proposed model. But they illustrate differences in conclusions resulted from using different models and methods.

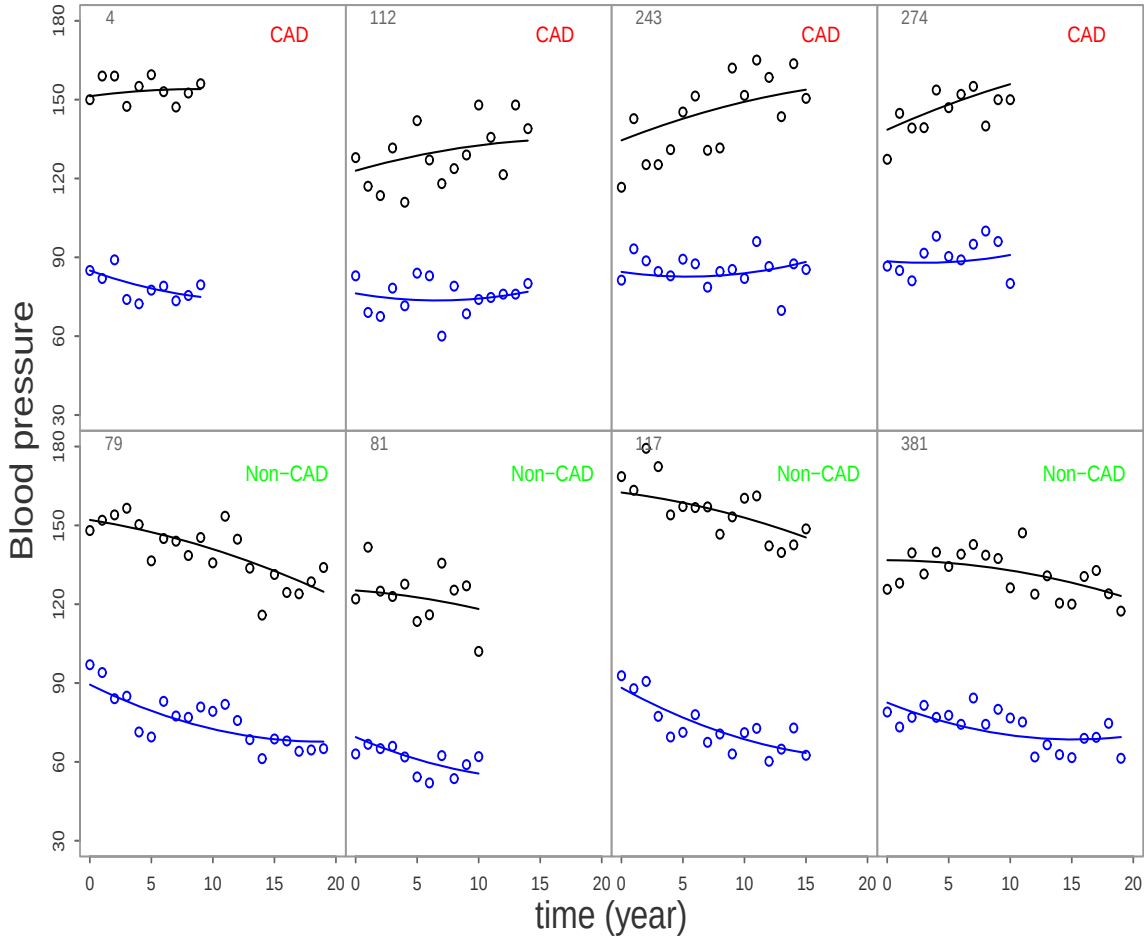

Figure 2. Fitted subject-specific longitudinal BP curves for randomly selected 4 CAD and 4 non-CAD subjects based on fitted Joint models 3. The black dots and black solid curves represent the observed systolic BP overtime and fitted subject-specific curves respectively. The blue dots and blue solid curves represent the observed diastolic BP overtime and fitted subject-specific curves respectively.

7. Conclusion

We developed a maximum-likelihood method using the EM algorithm for parameter estimation of joint models for multiple longitudinal processes and a time-to-event outcome. Simulation studies indicated adequate performance of the EM based estimation approach which performed better than the two-stage estimation approach. We also applied the proposed method to data from a primary care patient cohort using EMR data for longitudinal systolic and diastolic BP and investigating their associations with the risk of CAD.

Compared to joint models with univariate longitudinal data, joint models with multivariate longitudinal outcomes can account for correlations among the longitudinal measures such as systolic and diastolic BP. On the other hand, the including of multiple longitudinal outcomes leads to a more complex model due to additional random effects. A pseudo-adaptive Gaussian-Hermit quadrature rule [35] was used to improve the computing performance of EM algorithm where multiple random effects exist in the model. In addition, simultaneous modeling of multiple longitudinal outcomes in joint models offers a number of advantages over separate modeling of each longitudinal outcome including more robust parameter estimates and improved predictive accuracy [53].

Our current work focused on joint models with normally distributed longitudinal out-

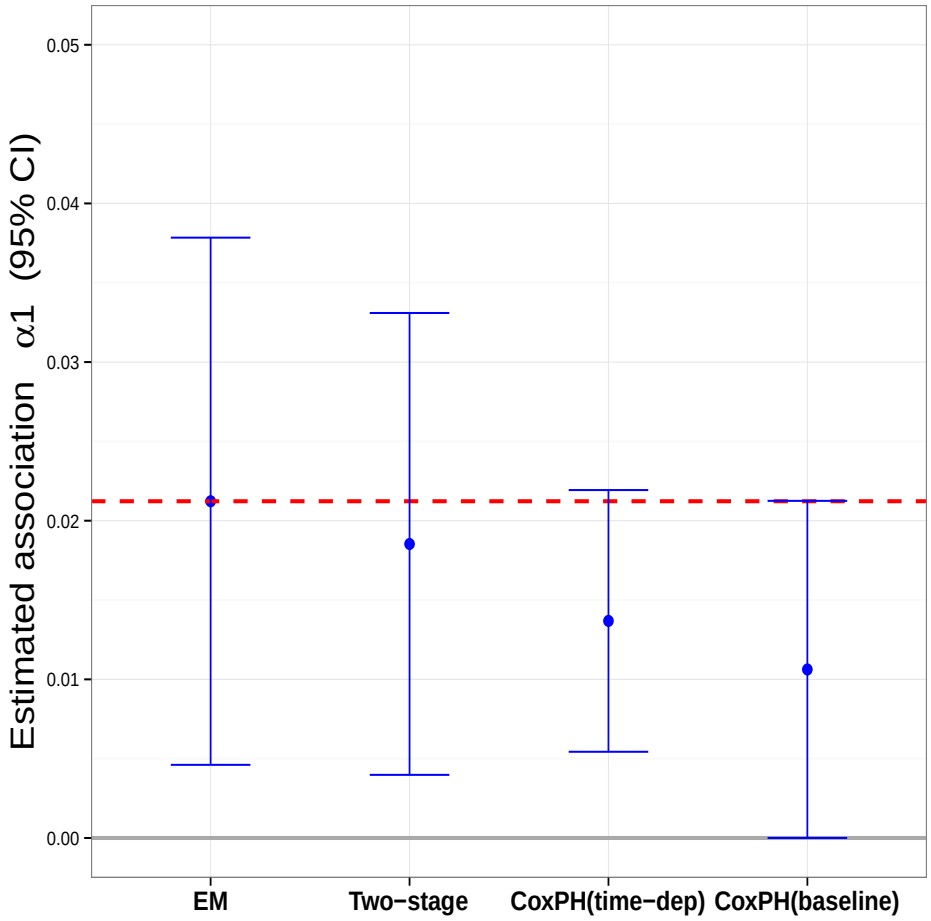

Figure 3. Comparison of estimated association ($\hat{\alpha_1}$) between the longitudinal systolic BP and risk of CAD from four methods. The blue solid dots are estimated $\hat{\alpha_1}$ from the four methods. The upper and lower bars are 95% CI of parameter estimates. The red dashed line denotes the estimate from the EM algorithm.

comes. It is worth noting that the proposed methodology can be extended to joint models with other distributions for the longitudinal outcomes such as binary, Poisson and others. The proposed EM algorithm can be used for estimation from joint models with mixed types of longitudinal outcomes. Other potential extensions include compete-risk models or semi-compete-risk models to take informative censoring into consideration. Another area for further research is on predictive accuracy based on the proposed joint models.

The methodology for joint models of multiple longitudinal processes and time-to-event outcome is applicable to many clinical and epidemiological studies where the association between longitudinal measures and time-to-event outcome is often of interest. The joint model framework provides a platform for exploring various features of the longitudinal measures related to disease risk, extending the traditional approach that relies on baseline measures only in cohort studies. With the increasing use of EMR in routine clinical practices, joint models can become a powerful tool for identifying longitudinal risk factors for disease risk and may offer insights for potential disease mechanisms that are otherwise not available using traditional approaches.

Acknowledgment

References

[1] Yang L, Gao S. Bivariate random change point models for longitudinal outcomes. Statistics in Medicine. 2012;32(6):1038–1053.
[2] Rothwell PM, Howard SC, Dolan E, O'Brien E, Dobson JE, Dahlöf B, Sever PS, Poulter NR. Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. Lancet. 2010;375:895–905.
[3] Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society, Series B. 1972;34(2):187–220.
[4] Andersen PK, Gill RD. Cox regression model for counting processes: A large sample study. Annals of Statistics. 1982;10:1100–1120.
[5] Fleming TR, Harrington DP. Counting processes and survival analysis. John Wiley and Sons; 1991.
[6] Andersen PK, Borgan O, Gill RD, Keiding N. Statistical models based on counting processes. New York: Springer; 1993.
[7] Faucett CJ, Thomas DC. Simultaneously modeling censored survival data and repeatedly measured covariates: A gibbs sampling approach. Statistics in Medicine. 1996;15:1663–1685.
[8] Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics. 1997;53:330–339.
[9] Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. Biostatistics. 2000;4:465–480.
[10] Brown ER, Ibrahim JG. A bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics. 2003;59:221–228.
[11] Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. Biometrics. 2005;61:64–73.
[12] Yu M, Law N, Taylor J, Sandler H. Joint longitudinal-survival-cure models and their application to prostate cancer. Statistica Sinica. 2004;14:835–862.
[13] Huang Y, Dagne G, Wu L. Bayesian inference on joint models of hiv dynamics for time-toevent and longitudinal data with skewness and covariate measurement errors. Statistics in Medicine. 2011;30:2930–2946.
[14] Elashoff R, Li G, Li N. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. Biometrics. 2008;64:762–771.
[15] Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. Biometrics. 2006;62:432–445.
[16] Njeru Njagi E, Molenberghs G, Verbeke G, Kenward MG, Dendale P, Willekens K. A flexible joint-modelling framework for longitudinal and time-to-event data with overdispersion. Statistical Methods in Medical Research; 2013.
[17] Qiu F, Stein CM, Elston RC, et al. Joint modeling of longitudinal data and discrete-time survival outcome. Statistical methods in medical research; 2013.
[18] Tsiatis AA, Davidian M. An overview of joint modeling of longitudinal and time-to-event data. Statistica Sinica. 2004;14:793–818.
[19] Sousa I. A review of joint modeling of longitudinal measurements and time-to-event. REVSTAT. 2011;9:57–81.
[20] Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: A review. Statistical methods in medical research. 2014; 23(1):74–90.
[21] Rizopoulos D, Ghosh P. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Statistics in Medicine. 2011;30:1366–1380.
[22] Xu J, Zeger SL. Joint analysis of longitudinal data comprising repeated measures and times to events. Journal of the Royal Statistical Society Series C. 2001;50:375–387.
[23] Song X, Davidian M, Tsiatis A. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics. 2002;3:511–528.
[24] Fieuws S, Verbeke G, Maes B, Vanrenterghem Y. Predicting renal graft failure using multi-

variate longitudinal profiles. Biostatistics. 2008;9:419–431.
[25] McCulloch C. Joint modelling of mixed outcome types using latent variables. Statistical Methods in Medical Research. 2008;17:53–73.
[26] Gueorguiva R, Sanacora G. Joint analysis of repeatedly observed continuous and ordinal measures of disease severity. Statistics in Medicine. 2006;25:1307–1322.
[27] Albert PS, Shih JH. On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. Biometrics. 2010;66:983–987.
[28] Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biometrical Journal. 2011;53:750–763.
[29] Self S, Pawitan Y. Modeling a marker of disease progression and onset of disease. In: Aids epidemiology. Springer; 1992. p. 231–255.
[30] Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error: applications to survival and cd4 counts in patients with aids. Journal of the American Statistical Association. 1995;90:27–37.
[31] Dafni UG, Tsiatis AA. Evaluating surrogate markers of clinical outcome whenmeasured with error. Biometrics. 1998;54:1445–1462.
[32] Ye W, Lin X, Taylor J. A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. Statistics and its Interface. 2006;1:33–45.
[33] He B, Luo S. Joint modeling of multivariate longitudinal measurements and survival data with applications to parkinson's disease. Statistical methods in medical research; 2013.
[34] Tseng YK, Hsieh F, Wang JL. Joint modelling of accelerated failure time and longitudinal data. Biometrika. 2005;92:587–603.
[35] Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. Computational Statistics and Data Analysis. 2012;56:491–501.
[36] Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from imcplete data via em algorithm. Journal of the Royal Statistical Society, Series B (Methodological). 1977; 39:1–38.
[37] Rizopoulos D. Jm : An R package for the joint modelling of longitudinal and time-to-event data. Journal of Statistical Software. 2010;35:1–33.
[38] Rizopoulos D. Joint models for longitudinal and time-to-event data, with applications in r. Chapman and Hall/CRC Biostatistics Series; 2012.
[39] Couvreur C. The EM algorithm: A guided tour. In: Computer intensive methods in control and signal processing. Springer; 1997. p. 209–222.
[40] Callahan CM, Hui SL, Nienaber NA, Musick BS, Tierney WM. Longitudinal study of depression and health services use among elderly primary care patients. Journal of the American Geriatrics Society. 1994;42:833–838.
[41] Callahan CM, Hendrie HC, Dittus RS, Brater DC, Hui SL, Tierney WM. Improving treatment of late life depression in primary care: a randomized clinical trial. Journal of the American Geriatrics Society. 1994;42:839–846.
[42] Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. American Heart Journal. 1991;121:293–298.
[43] Wilson PWF, Agostino RBD, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97:1837–1847.
[44] Stamler J, Stamler R, Neaton JD. Blood pressure, systolic and diastolic, and cardiovascular risks. us population data. Archives of Internal Medicine. 1993;153(5):598–615.
[45] Hochman J, Tamis JE, Thompson TD. Sex, clinical presentation, and outcome in patients with acute coronary syndromes. New England Journal of Medicine. 1999;341:226–232.
[46] Vaccarino V, Parsons L, Every NR, Barron HV, Krumholz HM. Sex-based differences in early mortality after myocardial infarction. national registry of myocardial infarction 2 participants. New England Journal of Medicine. 1999;341:217–225.
[47] Press W, Teukolsky S, Vetterling W, Flannery B. Numerical recipes: The art of scientific computing. 3rd ed. New York: Cambridge University Press; 2007.
[48] Akaike H. Factor analysis and aic. Psychometrika. 1987;52:317–332.
[49] Schwarz GE. Estimating the dimension of a model. Annals of Statistics. 1978;6(2):461–464.

[50] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN; 2007.
[51] Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. Vol. 998. John Wiley & Sons; 2012.
[52] Gao S, Hendrie HC, Wang C, Stump TE, Stewart JC, Kesterson J, Clark DO, Callahan CM. Redefined blood pressure variability measure and its association with mortality in elderly primary care patients. Hypertension. 2014;64(1):45–52.
[53] Yang L, Yu M, Gao S. Prediction of coronary artery disease risk based on multiple longitudinal biomarkers. Statistics in medicine. 2015;.

Appendix. The M-Step for the Joint Models

In the m-th iteration, the expected complete log-likelihood function given the conditional distribution of random effects is

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^m) = \sum_{i=1}^{n} \int \Big\{ & \log p(T_i, \delta_i|\mathbf{b}_i;\boldsymbol{\theta}) + \log p(\mathbf{y}_{i1}|\mathbf{b}_i;\boldsymbol{\theta}) + ... + \log p(\mathbf{y}_{iL}|\mathbf{b}_i;\boldsymbol{\theta}) \\
& + \log p(\mathbf{b}_i;\boldsymbol{\theta}) \Big\} p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m) d\mathbf{b}_i \\
= \sum_{i=1}^{n} \int \Bigg\{ & \delta_i \left(\log(h_0(T_i)) + \mathbf{w}_i^T\gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il}) \right) \\
& - \int_0^{T_i} h_0(u) \exp\left(\mathbf{w}_i^T\gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_l^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(u)\mathbf{b}_{il}\Big) \right) du \\
& + \sum_{l=1}^{L} \left(-\frac{1}{2} n_i \log\left(2\pi\sigma_l^2\right) - \frac{1}{2\sigma_l^2} \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l - \mathbf{Z}_{il}^T\mathbf{b}_{il}\right)^T \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l - \mathbf{Z}_{il}^T\mathbf{b}_{il}\right) \right) \\
& + \frac{k}{2} \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\mathbf{b}_i^T\mathbf{D}^{-1}\mathbf{b}_i \Bigg\} p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL};\boldsymbol{\theta}^m) d\mathbf{b}_i
\end{aligned}
$$

For the M-step, closed-form expressions are available for the variance of residuals of each longitudinal model and variance-covariance matrix of the random effects, whereas the fixed effects for each longitudinal model and parameters in the survival model have to be estimated numerically.

(1) The variance of residuals of each longitudinal model
Taking derivatives with respect to σ_l^2 for the expected complete log-likelihood function

and setting to zero we get

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^m)}{\partial \sigma_l^2} &= \sum_{i=1}^{n} \int \frac{\partial}{\partial \sigma_l^2} \Big(-\frac{1}{2} n_i \log\left(2\pi\sigma_l^2\right) \\ &\quad -\frac{1}{2\sigma_l^2} \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l - \mathbf{Z}_{il}^T\mathbf{b}_{il}\right)^T \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l - \mathbf{Z}_{il}^T\mathbf{b}_{il}\right) \Big) \\ &\quad p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\ &= -\frac{\sum_{i=1}^n n_i}{2\sigma_l^2} + \frac{1}{2\sigma_l^4} \sum_{i=1}^{n} \int \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l - \mathbf{Z}_{il}^T\mathbf{b}_{il}\right)^T \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l - \mathbf{Z}_{il}^T\mathbf{b}_{il}\right) \\ &\quad p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\ &= 0\end{aligned}$$

Solving the above equation we get

$$\begin{aligned}\hat{\sigma_l}^2 &= \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^{n} \int \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l - \mathbf{Z}_{il}^T\mathbf{b}_{il}\right)^T \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l - \mathbf{Z}_{il}^T\mathbf{b}_{il}\right) \\ &\quad p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\ &= \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^{n} \int \Big(\left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l\right)^T \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l\right) - \mathbf{b}_{il}^T\mathbf{Z}_{il} \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l\right) \\ &\quad - \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l\right)^T \mathbf{Z}_{il}^T\mathbf{b}_{il} + \left(\mathbf{Z}_{il}^T\mathbf{b}_{il}\right)^T \mathbf{Z}_{il}^T\mathbf{b}_{il} \Big) p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\ &= \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^{n} \int \Big(\left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l\right)^T \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l\right) - 2 \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l\right)^T \mathbf{Z}_{il}^T\mathbf{b}_{il} \\ &\quad + \left(\mathbf{Z}_{il}^T\mathbf{b}_{il}\right)^T \mathbf{Z}_{il}^T\mathbf{b}_{il} \Big) p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\ &= \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^{n} \int \Big(\left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l\right)^T \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T\beta_l - 2\mathbf{Z}_{il}^T\mathbf{b}_{il}\right) + \mathbf{b}_{il}^T\mathbf{Z}_{il}\mathbf{Z}_{il}^T\mathbf{b}_{il} \Big) \\ &\quad p(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\ &= \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^{n} \Big(\mathbf{y}_{il} - \mathbf{X}_{il}^T\boldsymbol{\beta}_l^m \Big)^T \Big(\mathbf{y}_{il} - \mathbf{X}_{il}^T\boldsymbol{\beta}_l^m - 2\mathbf{Z}_{il}^T \mathrm{E}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) \Big) \\ &\quad + \mathrm{Tr}\Big(\mathbf{Z}_{il}^T\mathbf{Z}_{il} \mathrm{Var}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) \Big) \\ &\quad + \mathrm{E}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m)^T \mathbf{Z}_{il}^T\mathbf{Z}_{il} \mathrm{E}(\mathbf{b}_i|T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m)\end{aligned}$$

where $l = 1, ..., L$, Tr represents the trace function of a matrix, and E denotes the expectation function.

(2) The variance-covariance matrix of random effects
Taking derivatives with respect to $\mathbf{D}^{-1}$ for the expected complete log-likelihood function and setting to zero we get

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^m)}{\partial \mathbf{D}^{-1}} &= \sum_{i=1}^{n} \int \frac{\partial}{\partial \mathbf{D}^{-1}} \left(-\frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right) p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\
&= \sum_{i=1}^{n} \int \frac{\partial}{\partial \mathbf{D}^{-1}} \left(\frac{1}{2} \log |\mathbf{D^{-1}}| - \frac{1}{2} \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right) p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\
&= \sum_{i=1}^{n} \int \left(\frac{1}{2} \mathbf{D} - \frac{1}{2} \mathbf{b}_i^T \mathbf{b}_i \right) p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\
&= \frac{n}{2} \mathbf{D} - \frac{1}{2} E \left(\mathbf{b}_i^T \mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m \right) \\
&= 0
\end{aligned}$$

Solving the above equation we get

$$\begin{aligned}
\hat{\mathbf{D}} = \frac{1}{n} \sum_{i=1}^{n} & \mathrm{Var}(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) \\
& + \mathrm{E}(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) \mathrm{E}(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m)^T
\end{aligned}$$

(3) The score functions for the fixed effect vector of each longitudinal model can be derived by taking derivatives with respect to $\boldsymbol{\beta}_l$ for the expected complete log-likelihood function.

$$\begin{aligned}
S(\boldsymbol{\beta}_l) = \sum_{i=1}^{n} \int \frac{\partial}{\partial \boldsymbol{\beta}_l} \Bigg\{ & \delta_i \left(\sum_{l=1}^{L} \alpha_l \mathcal{F}(\mathbf{X}_l^T(T_i) \boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i) \mathbf{b}_{il}) \right) \\
& - \int_0^{T_i} h_0(u) \exp \left(\mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_l^T(u) \boldsymbol{\beta}_l + \mathbf{Z}_l^T(u) \mathbf{b}_{il} \Big) \right) du \\
& + \sum_{l=1}^{L} \left(-\frac{1}{2} n_i \log \left(2\pi \sigma_l^2 \right) - \frac{1}{2\sigma_l^2} \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T \beta_l - \mathbf{Z}_{il}^T \mathbf{b}_{il} \right)^T \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T \beta_l - \mathbf{Z}_{il}^T \mathbf{b}_{il} \right) \right) \Bigg\} \\
& p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\
= \sum_{i=1}^{n} & \frac{1}{\sigma_l^2} \mathbf{X}_{il}^T \Big(\mathbf{y}_{il} - \mathbf{X}_{il}^T \boldsymbol{\beta}_l - \mathbf{Z}_{il}^T \mathrm{E}(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) \Big) \\
& + \delta_i \alpha_l \frac{\partial \mathcal{F}\Big(\mathbf{X}_l^T(T_i) \boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i) \mathbf{b}_{il} \Big)}{\partial \boldsymbol{\beta}_l} \\
& - \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} h_0(u) \frac{\partial \exp \left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u) \boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u) \mathbf{b}_i \Big) \right)}{\partial \boldsymbol{\beta}_l} \\
& p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i,
\end{aligned}$$

for $l = 1, ..., L$.

(4) The score functions of parameters in the survival model are derived as follows.

$$
\begin{aligned}
S(\gamma) = \sum_{i=1}^{n} \int \frac{\partial}{\partial \gamma} \Bigg\{ \delta_i \left(\log(h_0(T_i)) + \mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il}) \right) \\
- \int_0^{T_i} h_0(u) \exp\left(\mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_l^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(u)\mathbf{b}_{il} \Big) \right) du \Bigg\} \\
p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\
= \sum_{i=1}^{n} \mathbf{w}_i \Bigg\{ \delta_i - \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} h_0(u) \exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i \Big) \right) \\
p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i \Bigg\}.
\end{aligned}
$$

$$
\begin{aligned}
S(\alpha_l) = \sum_{i=1}^{n} \int \frac{\partial}{\partial \alpha_l} \Bigg\{ \delta_i \left(\log(h_0(T_i)) + \mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il}) \right) \\
- \int_0^{T_i} h_0(u) \exp\left(\mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_l^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(u)\mathbf{b}_{il} \Big) \right) du \Bigg\} \\
p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\
= \sum_{i=1}^{n} \delta_i \int \mathcal{F}\Big(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il} \Big) p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\
- \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} h_0(u) \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i \Big) \\
\exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i \Big) \right) p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i,
\end{aligned}
$$

where $l = 1, ..., L$.

$$
\begin{aligned}
S(\boldsymbol{\theta}_{h_0}) = \sum_{i=1}^{n} \int \frac{\partial}{\partial \boldsymbol{\theta}_{h_0}} \Bigg\{ \delta_i \left(\log(h_0(T_i)) + \mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il}) \right) \\
- \int_0^{T_i} h_0(u) \exp\left(\mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_l^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(u)\mathbf{b}_{il} \Big) \right) du \Bigg\} \\
p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) d\mathbf{b}_i \\
= \sum_{i=1}^{n} \delta_i \frac{1}{h_0(T_i; \boldsymbol{\theta}_{h_0})} \frac{\partial h_0(T_i; \boldsymbol{\theta}_{h_0})}{\partial \boldsymbol{\theta}_{h_0}^T} \\
- \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} \frac{\partial h_0(T_i; \boldsymbol{\theta}_{h_0})}{\partial \boldsymbol{\theta}_{h_0}^T} \exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i \Big) \right) \\
p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i.
\end{aligned}
$$

Table 1. Simulation results for comparing the EM algorithm and the two-stage approach under a small variance scenario.

		EM				Two-stage			
Parameter	True	Relative Bias	Model SE	Emp. SE	95%CI Cov.	Relative Bias	Model SE	Emp. SE	95%CI Cov.
β_{01}	0.20	0.000	0.008	0.008	95.4%	-0.004	0.008	0.007	95.2%
β_{11}	0.50	-0.000	0.003	0.003	94.2%	0.003	0.003	0.003	92.6%
β_{21}	0.20	-0.000	0.011	0.010	96.8%	0.002	0.011	0.010	97%
$\log(\sigma_1)$	-2.30	0.000	0.017	0.017	94.8%	0.000	0.017	0.017	94.8%
β_{02}	1.00	0.001	0.008	0.008	95%	-0.001	0.008	0.008	95.4%
β_{12}	0.20	0.001	0.003	0.003	94.4%	0.010	0.003	0.003	91.6%
β_{22}	0.50	-0.001	0.011	0.010	95%	0.000	0.011	0.010	95.2%
$\log(\sigma_2)$	-2.30	0.000	0.017	0.016	96.6%	0.000	0.017	0.017	94.8%
α_1	1.00	-0.104	0.201	0.198	92.6%	-0.000	0.393	0.378	95.4%
α_2	1.50	-0.004	0.221	0.223	94.2%	-0.007	0.243	0.242	94.6%
D_{11}	-2.30	-0.003	0.050	0.049	94.8%	-0.002	0.044	0.043	96.6%
D_{22}	-3.22	-0.004	0.079	0.082	95.6%	-0.001	0.069	0.070	95.4%
D_{13}	0.05	-0.003	0.007	0.007	94.4%	-0.001	0.006	0.006	91.4%
D_{33}	-2.45	-0.007	0.064	0.065	95.4%	-0.002	0.044	0.054	89%
D_{44}	-3.22	-0.009	0.082	0.083	96%	-0.003	0.069	0.069	94.4%

Table 2. Simulation results for comparing the EM algorithm and the two-stage approach under a large variance scenario.

		EM				Two-stage			
Parameter	True	Relative Bias	Model SE	Emp. SE	95%CI Cov.	Relative Bias	Model SE	Emp. SE	95%CI Cov.
β_{01}	0.20	-0.062	0.038	0.037	93.4%	-0.109	0.038	0.037	91.4%
β_{11}	0.50	-0.014	0.017	0.017	94.6%	0.034	0.015	0.015	78.8%
β_{21}	0.20	0.041	0.052	0.055	94.6%	0.077	0.052	0.054	93.8%
$\log(\sigma_1)$	-0.69	0.003	0.017	0.018	95%	0.005	0.017	0.018	93.8%
β_{02}	1.00	-0.016	0.038	0.041	91.8%	-0.032	0.037	0.040	83%
β_{12}	0.20	-0.041	0.017	0.016	93%	0.135	0.015	0.014	55.6%
β_{22}	0.50	0.019	0.052	0.054	92.4%	0.040	0.051	0.054	91.6%
$\log(\sigma_2)$	-0.69	0.001	0.017	0.017	96.2%	0.005	0.017	0.018	93.8%
α_1	1.00	-0.035	0.106	0.103	95.6%	0.122	0.093	0.090	71.6%
α_2	1.50	0.010	0.127	0.123	95.6%	0.114	0.103	0.101	59.2%
D_{11}	-0.69	-0.016	0.050	0.054	94.2%	-0.036	0.046	0.049	92%
D_{22}	-1.61	-0.033	0.079	0.081	91%	-0.037	0.072	0.074	86.6%
D_{13}	0.25	0.014	0.033	0.033	95%	0.109	0.026	0.029	78%
D_{33}	-0.84	-0.027	0.064	0.062	96.6%	-0.023	0.046	0.050	91%
D_{44}	-1.61	-0.070	0.085	0.083	80.6%	-0.082	0.077	0.076	65%

Table 3. Parameter estimates, standard errors and 95%CI for **Joint Models 3**. α_1 and α_2 are the association estimates between the risk of CAD and **current value** of systolic and diastolic BP at event time point, respectively. λ_i $i = 1, ..., 7$ denote the baseline hazards of the 7 piecewise constant intervals.

Parameter	Estimate	Std Err	lower 95%CI	upper 95%CI
Longitudinal systolic BP				
Intercept	135.53	0.80	133.95	137.10
time	0.26	0.16	-0.06	0.57
time^2	-0.03	0.01	-0.05	-0.01
Age	0.01	0.06	-0.10	0.11
Race	4.40	0.82	2.78	6.01
$\log(\sigma_1)$	2.47	0.01	2.45	2.50
Longitudinal diastolic BP				
Intercept	79.42	0.34	78.75	80.09
time	-1.64	0.09	-1.82	-1.46
time^2	0.06	0.01	0.05	0.07
Age	-0.13	0.02	-0.18	-0.09
Race	2.74	0.32	2.11	3.36
$\log(\sigma_1)$	1.94	0.01	1.92	1.97
Time-to-CAD				
Age	0.06	0.01	0.04	0.08
Smoking History	0.35	0.15	0.06	0.65
Race	-0.49	0.15	-0.78	-0.19
Diabetes	-0.00	0.14	-0.28	0.28
α_1	0.021	0.008	0.005	0.038
α_2	0.011	0.014	-0.017	0.039
$\log(\lambda_1)$	-7.73	0.80	-9.30	-6.17
$\log(\lambda_2)$	-8.17	0.80	-9.74	-6.61
$\log(\lambda_3)$	-7.85	0.79	-9.41	-6.29
$\log(\lambda_4)$	-7.19	0.78	-8.72	-5.66
$\log(\lambda_5)$	-6.49	0.77	-8.00	-4.98
$\log(\lambda_6)$	-6.61	0.76	-8.10	-5.12
$\log(\lambda_7)$	-6.03	0.75	-7.50	-4.55