ORIGINAL ARTICLE

Semiparametric modeling of biomarker trajectory and variability with correlated measurement errors

Renwen Luo[1] | **Chuoxin Ma**[1] | **Jianxin Pan**[1]

[1]Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, Zhuhai, 519087, China

Correspondence
Jianxin Pan, Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, Zhuhai, 519087, China
Email: jianxinpan@uic.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 12271047; Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, Grant/Award Number: 2022B1212010006

The prognostic significance of biomarker variability in predicting associated disease risk is well-established. However, prevailing methods that assess the relationship between biomarker variability and time to event often overlook within-subject correlation in longitudinal measurement errors, resulting in biased parameter estimates and erroneous statistical inference. Additionally, these methods typically assumes that biomarker trajectory can be represented as a linear combination of spline basis functions with normally distributed random effects. This not only leads to significant computational demands due to the necessity of high-dimensional integration over the random effects but also limits the applicability because of the normality restriction imposed on the random effects. This paper addresses these limitations by incorporating correlated longitudinal measurement errors and proposing a novel semiparametric multiplicative random effects model. This model does not assume normality for the random effects and eliminates the need for integration with respect to them. The biomarker variability is incorporated as a covariate within a Cox model for time-to-event data, thus facilitating a joint modeling strategy. We demonstrate the asymptotic properties of the proposed estimators and validate their performance through simulation studies. The methodology is applied to assess the impact of systolic blood pressure variability on cardiovascular mortality using data from the Atherosclerosis Risk in Communities study.

KEYWORDS
ARIC study, correlated errors, joint modeling, semiparametric model, variability

1 | INTRODUCTION

This paper is motivated by the Atherosclerosis Risk in Communities (ARIC) study, which aims to assess the relationship between the levels of and/or changes in biomarkers and the risk of developing cardiovascular disease events. Cardiovascular death (CVD) is one of the main events in the ARIC study. Certain biomarkers, such as systolic blood pressure (SBP), are known to play a critical role in evaluating CVD risk, and elevated SBP levels are typically associated with higher CVD risk.[1] In medical science, however, the long-held assumption that the average level of blood pressure alone fully explains all blood pressure-related risks of vascular events has been called into question. Increasing evidence indicates that visit-to-visit variability in blood pressure, particularly SBP, also plays a significant role. Rothwell, Rothwell et al and Rothwell et al published their results on *The Lancet* and highlighted the limitations of predicting death risk solely based on factors like blood pressure.[1,2,3] Their findings suggest that patients with normal average blood pressure levels but high visit-to-visit variability may face a higher risk of stroke or death compared to those with higher but relatively stable blood pressure levels. This inspires us to investigate whether the visit-to-visit variability of SBP contributes to the CVD of patients in the ARIC study.

To assess the impact of visit-to-visit variability on time-to-event outcomes (e.g., time to cardiovascular death in the ARIC study), three primary methods based on survival analysis have been employed. The first method calculated sample standard deviation, coefficient of variation, or successive variation based on the longitudinal repeated measurements and then incorporated them as covariates in the survival model.[4,5] The second method used the first-order derivative of the biomarker trajectory at the current time point as a measure of variability.[6] The third method treated the residual variance of the biomarker trajectory or its logarithm as a frailty factor in a survival model to account for the association between the biomarker variability and event risk, which can be significantly affected by the regression dilution bias.[7,8] Each of these three methods has its limitations. The variability measures in the first approach may yield unreliable estimates, particularly when the small number of repeated measurements is small.[9] The second method may generate negative values for variability, which contradicts the inherently positive nature of variability measures. While the absolute value of the derivative of the biomarker trajectory may serve as a potential measure of variability, it is unsuitable for characterizing cumulative variability over a time period. The term “cumulative variability” refers to the aggregation of past variability values. It is typically of greater scientific interest since it encompasses both historical information and current information that may influence current event risk but “variability" only contains current information. We discuss this point in Section S14 of the supporting information. The third method, as pointed out by Barrett et al, does not distinguish between true variability in a subject’s repeated measurements and measurement errors.[8]

Wang et al, motivated by the concept of roughness measures in smoothing splines, recently introduced a new measure that characterizes the biomarker variability.[10] However, their study was limited to the current biomarker level and its variability, which may not sufficiently capture the association structure between these covariates and current risk. As noted by Rizopoulos et al and Mauff et al, the association between past values of time-dependent covariates and event risk may change based on the time gap between the current moment and when those covariates were measured.[11,12] To address this limitation, Wang et al proposed a new model with cumulative biomarker trajectory and cumulative variability measure as covariates. This model is inspired by the idea of weighted cumulative exposure that assigns greater weights to values of time-dependent covariates measured closer to the current time point.[13,14] In the rest of this paper, we call methods proposed by Wang and colleagues as Wang’s methods.[10,13]

Although the existing methods typically assume that longitudinal measurement errors within the same subject are mutually independent,[15] our analysis for the ARIC study in Section 5 shows that these measurement errors can be correlated to each other. Furthermore, our simulations, as detailed in Section 4.2 of this paper and Section S1 of

the supporting information, suggest that ignoring the correlation in measurement errors may lead classical estimation methods like Conditional Score, Corrected Score and Wang's methods to produce biased parameter estimates, with significantly reduced coverage probabilities for those parameters describing the effects of biomarker trajectory and its variability on event risk. [16,17].

Since the true biomarker trajectories are usually unknown in real applications, existing methods typically assumes that they can be approximated as a linear combination of spline basis functions with random coefficients, which are typically referred to random effects. [15,10,13] Expectation-Maximization (EM) algorithm and Bayesian algorithm are commonly used for parameter estimation [10,8]. However, these algorithms typically assume normally distributed random effects, an assumption that may not always hold in real-world applications. Additionally, the computational cost associated with these algorithms can become prohibitively expensive due to the potentially high-dimensional integration over the random effects. [18]

In ARIC study, it is observed that younger hypertensive patients can effectively control their SBP level with antihypertensive medication. However, as patients age, factors such as vascular stiffening contribute to a rise in SBP despite continued adherence to medication. [19] This trend is evident in Figure 1, which plots the estimated population SBP trajectory (normalized by 50) from study enrollment through a 15-year follow-up period (normalized by 3650 days). The trajectory remains stable and linear for the first 7 years, becoming increasingly nonlinear and rising from years 7 to 15. This trend is also present in other datasets, such as Acquired Immunodeficiency Syndrome (AIDS) data depicted in Figure 3 of Yao and Primary Biliary Cirrhosis (PBC) data shown in Figure 3 of Ding and Wang [20,21]. This trend motivates us to develop a new model which captures population trajectory approximately linear over initial time period but nonlinear as time progresses.

[Figure 1 here.]

In this paper, we propose new joint models to investigate the effects of cumulative biomarker trajectory and cumulative variability on current event risk. Our models incorporate correlation in measurement errors, ensuring that the parameters describing these effects can be unbiasedly estimated. We introduce a new semiparametric multiplicative random effects (SMRE) model for captures population trajectory approximately linear over initial time period but nonlinear as time progresses. The proposed SMRE model allows for the construction of an asymptotically unbiased estimating equation, which does not impose distributional assumption on the random effects and does not involve integration with respect to them. Therefore, solving this equation for parameter estimation is computationally efficient. We rigorously establish the consistency and asymptotic normality of the proposed parameter estimators.

This paper is organized as follows. Section 2 delineates the model formulation for the proposed joint models. Section 3 provides estimating procedures and theoretical properties of the proposed parameter estimators. Simulation studies are conducted in Section 4 to evaluate performance of the proposed methodology, which is applied to analyze the ARIC data in Section 5. This paper is concluded with a discussion in Section 6.

2 | MODEL FORMULATION

For each subject i, suppose longitudinal outcomes $y_{i1}, \ldots, y_{in_i}$ are observed at time points $t_{i1}, \ldots, t_{in_i}$. Suppose the event of interest is observed at T_i, with $\delta_i = 1$ implying failure while $\delta_i = 0$ indicating censoring. Then T_i is determined as the minimum of the time to event T_i^* and the right censoring time C_i, specifically $T_i = \min\{T_i^*, C_i\}$. We assume that all of observation times t_{ij} and the event times T_i fall within the study period $[0, \varphi]$, and the censoring times C_i are noninformative and observation times t_{ij} are independent of the longitudinal process, time to event and censoring.

2.1 | Related literature

In medical research,[15,18,22] the biomarkers are often measured longitudinally. In particular, y_{ij}, the biomarker for the i-th subject at time t_{ij}, is typically assumed to follow the model:

$$y_{ij} = m_i(t_{ij}) + \epsilon_{ij}, \quad m_i(t) = \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta} + f_i(t), \quad i = 1, \ldots, n, \; j = 1, \ldots, n_i, \tag{1}$$

where $m_i(\cdot)$ representing the true biomarker trajectory, regression parameter $\boldsymbol{\beta}$ captures the relationship between the baseline covariate $\boldsymbol{x}_i$ and the true trajectory $m_i(\cdot)$, $f_i(\cdot)$ is a smooth function that describes the shape of the subject-specific biomarker trajectory, ϵ_{ij} represents the measurement error at time t_{ij}, n denotes sample size and n_i denotes number of longitudinal measurements for the i-th subject.

Wang et al, motivated by the idea of roughness term in smoothing spline, proposed a roughness measure $\pi_i(t) = \left[\int_0^t \{m_i''(s)\}^2 ds\right]^{1/2}$ to describe the cumulative variability over time period $[0, t]$ for subject i, based on the biomarker trajectory $m_i(\cdot)$.[10] To assign greater weights to values of $m_i''(s)$ measured closer to current time t, Wang et al proposed a weighted roughness measure

$$\pi_i(t, \sigma) = \left[\int_0^t \omega_\sigma(t - s)\{m_i''(s)\}^2 ds\right]^{1/2}$$

to describe the cumulative variability over time period $[0, t]$.[13] Here, $\omega_\sigma(t - s) = \phi\{(t - s)/\sigma\}/[\sigma\{\Phi(t/\sigma) - \Phi(0)\}]$ represents a Gaussian weight function with a weight parameter σ, which controls the shape of the weight function, where $\phi(\cdot)$ and $\Phi(\cdot)$ represents density and cumulative distribution function for standard normal distribution, respectively. This weight function is designed to assign greater weight to the past values of $\{m_i''(s)\}^2$ as the value of s approaches the current time t. Specifically, according to Mauff et al, the value of $\pi_i(t, \sigma)$ is mainly determined by values of $\{m_i''(s)\}^2$ for $s \in [t - 2\sigma, t]$, provided that $t > 2\sigma$. And the value of $\pi_i(t, \sigma)$ is mainly determined by values of $\{m_i''(s)\}^2$ for $s \in [0, t]$, provided that $t \leq 2\sigma$. The word "cumulative" essentially means that $\pi_i(t, \sigma)$ cumulates the past values of $\{m_i''(s)\}^2$ for $s \in [0, t]$. Similarly, Wang et al considered the weighted biomarker trajectory

$$\eta_i(t, \sigma) = \int_0^t \omega_\sigma(t - s) m_i(s) ds$$

to describe the cumulative biomarker trajectory for the i-th subject over time period $[0, t]$.[13] Detailed interpretation of the cumulative biomarker trajectory and the cumulative variability can be found in Section S13 of the supporting information. Then, they incorporated $\eta_i(t, \sigma)$ and $\pi_i(t, \sigma)$ into Cox model:

$$\lambda_i(t) = \lambda_0(t) \exp\left\{\boldsymbol{z}_i^{\mathsf{T}}\boldsymbol{\gamma} + \alpha_1\eta_i(t, \sigma_1) + \alpha_2\pi_i(t, \sigma_2)\right\}, \quad i = 1, \ldots, n \tag{2}$$

where $\lambda_0(t)$ is baseline hazard function, $\lambda_i(t) = \lim_{dt \downarrow 0} P\{T_i^* \in (t, t + dt] | T_i^* \geq t, \{m_i(s) : s \in [0, t]\}, \boldsymbol{x}_i, \boldsymbol{z}_i\}/dt$ is hazard function for the i-th subject with time to event T_i^*, σ_1 and σ_2 are two positive weight parameters, $\boldsymbol{z}_i$ represents baseline covariates, which may or may not be the same as $\boldsymbol{x}_i$, and regression parameter $\boldsymbol{\gamma}$ describes the relationship between the baseline covariates $\boldsymbol{z}_i$ and the hazard function $\lambda_i(t)$. In model (2), statistical inference for parameters α_1 and α_2 is of particular interest since α_1 and α_2 respectively describes the effects of the cumulative biomarker trajectory and cumulative variability on the time to event.

2.2 | Modeling of correlated measurement errors

For modeling of correlation in the measurement errors within the same subject, we assume in the longitudinal submodel (1) that

$$\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{in_i})^{\mathsf{T}} \sim N(\mathbf{0}_{n_i}, \boldsymbol{\Sigma}_i(\boldsymbol{\zeta}_1)) \tag{3}$$

where $N(\mathbf{0}_{n_i}, \boldsymbol{\Sigma}_i(\boldsymbol{\zeta}_1))$ denotes a multivariate normal distribution with zero mean and covariance $\boldsymbol{\Sigma}_i(\boldsymbol{\zeta}_1)$, implying $\boldsymbol{\Sigma}_i$ is determined by parameter $\boldsymbol{\zeta}_1$. For example, if $\boldsymbol{\Sigma}_i$ is assumed to have the compound symmetric (CS) structure $\boldsymbol{\Sigma}_i = \sigma_\epsilon^2(\rho \boldsymbol{J}_i + (1-\rho)\boldsymbol{I}_{n_i})$, where $\boldsymbol{J}_{n_i}$ is $n_i \times n_i$ matrix with all elements equal to 1, $\boldsymbol{I}_{n_i}$ is $n_i \times n_i$ identity matrix, and variance correlation parameters σ_ϵ^2 and ρ satisfy $\sigma_\epsilon^2 > 0$ and $\rho \in (-1, 1)$, then $\boldsymbol{\zeta}_1 = (\rho, \sigma_\epsilon^2)$. Note, neglecting the correlation in measurement errors within the same subject is equivalent to assuming $\boldsymbol{\Sigma}_i = \sigma_\epsilon^2 \boldsymbol{I}_{n_i}$. That is, assuming that $\boldsymbol{\Sigma}_i$ is of independence structure. As discussed in Section 1, to avoid biased parameter estimates in the survival submodel, it is essential to consider a parametric model for $\boldsymbol{\Sigma}_i$, accounting for within-subject correlation in measurement errors. In Section 4.2, we will show the impact of neglecting correlation in measurement errors.

2.3 | Semiparametric multiplicative random effects model

To capture population biomarker trajectory is approximately linear over initial time period and becomes nonlinear as time progresses, we propose a SMRE model for $f_i(t)$,

$$f_i(t) = b_{i0} + b_{i1}t + b_{i2}\mu(t), \tag{4}$$

where $\boldsymbol{b}_i = (b_{i0}, b_{i1}, b_{i2})^{\mathsf{T}}$ is a vector of random effects and we do not impose distribution assumption on $\boldsymbol{b}_i$, $\mu(t)$ is a smooth function over $[0, \varphi]$ which describes the shape of $f_i(t)$ when t is large. This is justified by the following assumption on $\mu(t)$:

$$\lim_{t \to 0^+} \frac{|\mu(t)|}{t} = 0, \tag{5}$$

Suppose $\phi = (\phi_0, \phi_1)^{\mathsf{T}} = (E(b_{i0}), E(b_{i1}))^{\mathsf{T}}$, we assume $E(b_{i2}) = 1$ for identifiability of $\mu(t)$, then the shape of the population trajectory $E\{f_i(t)\} = \phi_0 + \phi_1 t + \mu(t)$ is primarily influenced by $\phi_0 + \phi_1 t$ when t is small and is determined by both the shapes of $\phi_0 + \phi_1 t$ and $\mu(t)$ when t is large, provided that ϕ_0 or ϕ_1 is not equal to zero. Therefore, SMRE model (4) can capture the population longitudinal trajectory which is approximately linear when t is small and is nonlinear when t is large. Note, when $\mu(t) = 0$, the proposed SMRE model (4) reduces to the classical linear mixed model.[23]

There is another way to interpret the SMRE model (4) with assumption (5). In practical applications, we often lack the knowledge for $f_i(t)$ and need to approximate or estimate it. This is typically done using regression spline techniques. Let $(1, t, t^2, ..., t^d, (t-u_1)_+^d, ..., (t-u_l)_+^d)^{\mathsf{T}}$ be a vector of regression spline basis functions, with order $d+1$ and l interior knots $\{u_j\}_{j=1}^l$, where $a_+ = \max\{a, 0\}$ for any scalar a. According to the property of the regression spline techniques, we assume there exists a random effect vector $\boldsymbol{a}_i = (a_{i0}, ..., a_{i(d+l)})^{\mathsf{T}}$, such that $f_i(t)$ can be well approximated as $f_i(t) \approx a_{i0} + a_{i1}t + g_i(t)$, where $g_i(t) = \sum_{k=2}^d a_{ik}t^k + \sum_{k=1}^l a_{i(d+k)}(t-u_k)_+^d$. Clearly, $\lim_{t\to 0^+} |g_i(t)|/t = 0$. When t is close to zero, $a_{i0} + a_{i1}t$ dominates $g_i(t)$ provided that a_{i0} or a_{i1} is not equal to zero, but when t is large, the shape of $f_i(t)$ is determined by both $a_{i0} + a_{i1}t$ and $g_i(t)$. As argued in Ding and Wang, after conducting the functional principal component analysis (FPCA) for $g_i(t)$, if the first eigenfunction of $g_i(t)$ explains a large percentage

of the variation of $g_i(t)$, $g_i(t)$ can be roughly approximated by $g_i(t) \approx \mu(t) + c_{i1}\varphi_{i1}(t)$, where $\mu(t)$ is the mean function of $g_i(t)$, $\varphi_{i1}(t)$ is the first eigenfunction and c_{i1} is the first eigenvalue.[21] If further the first eigenfunction $\varphi_{i1}(t)$ has a similar shape as the mean function $\mu(t)$ multiplied by a scale factor C, that is, we can find a constant C such that $\varphi_{i1}(t) \approx C\mu(t)$, then $g_i(t) \approx (1 + c_{i1}C)\mu(t)$. This leads to $f_i(t) \approx a_{i0} + a_{i1}t + (1 + c_{i1}C)\mu(t)$. Let $b_{i0} = a_{i0}$, $b_{i1} = a_{i1}$, and $b_{i2} = (1 + c_{i1}C)$, then we have $f_i(t) \approx b_{i0} + b_{i1}t + b_{i2}\mu(t)$, which inspires us to consider the SMRE model (4) with assumption (5). Additionally, if $b_{i0} = b_{i1} = 0$, then the proposed SMRE model simplifies to the nonparametric multiplicative random effects (NMRE) model proposed by Ding and Wang.[21] In their model, they assume that $f_i(t) = b_{i2}\mu(t)$ to characterize the longitudinal trajectories from different subjects as being proportional or approximately parallel, to each other. Therefore, the proposed SMRE model can be viewed as a generalization of the NMRE model.

To assess the applicability of the SMRE model (4) with assumption (5) to the practical data, one should assess whether $g_i(t) \approx b_{i2}\mu(t)$ holds. That is, one should verify: (i) whether the first eigenfunction $\varphi_{i1}(t)$ explains a large percentage of the variation of $g_i(t)$; (ii) whether the first eigenfunction $\varphi_{i1}(t)$ has a similar shape as the mean function $\mu(t)$ multiplied by a scale factor C. To do these, note that longitudinal model $y_{ij} = \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta} + a_{i0} + a_{i1}t_{ij} + g_i(t) + \epsilon_{ij}$ can be fitted by standard statistical method designed for fitting the linear mixed model. After fitting this model, one can obtain estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and estimator $\tilde{\boldsymbol{a}}_i$ of $\boldsymbol{a}_i$. Then, one can treat $y_{ij} - \boldsymbol{x}_i^{\mathsf{T}}\tilde{\boldsymbol{\beta}} - \tilde{a}_{i0} - \tilde{a}_{i1}t_{ij}$ as constructed response variables with mean $g_i(t_{ij})$ and conduct the FPCA to ascertain whether $\varphi_{i1}(t)$ explains a large percentage of the variation of $g_i(t)$ and whether $\varphi_{i1}(t)$ has a similar shape as the mean function $\mu(t)$ multiplied by a scale factor C. In Section S2 of the supporting information, we conduct simulation to evaluate the cost of misspecifying the SMRE model (4) with assumption (5) on $f_i(t)$.

Under the survival submodel (2) and the SMRE model (4), we have $\pi_i(t, \sigma_2) = |b_{i2}| \left[\int_0^t \omega_{\sigma_2}(t-s)\{\mu''(s)\}^2 ds\right]^{1/2}$, which indicates that the weighted roughness measure permits a form only depending on b_{i2} and $\mu''(\cdot)$. The use of the squared root in $\pi_2(t, \sigma_2)$ is due to the fact that $\mathbb{E}[\exp\{\pi_i(t, \sigma_2)^2\}]$ may not exist (i.e., $\mathbb{E}\{\exp(b_{i2}^2)\}$ does not exist even when b_{i2} is normally distributed). Therefore, if we remove the square root in model (2), the expectation of the hazard function $\lambda_i(t)$ in the survival submodel (2) may not exist.

Given the random effects and covariates, the longitudinal process and event time are assumed to be mutually independent. Let $\boldsymbol{Y}_i = (y_{i1}, ..., y_{in_i})^{\mathsf{T}}$ and define the covariance of the random effects $\boldsymbol{b}_i$ as $\boldsymbol{D}_b(\boldsymbol{\zeta}_2)$, implying that $\boldsymbol{D}_b$ is parameterized by $\boldsymbol{\zeta}_2$. The observed data $\{\boldsymbol{Y}_i, \boldsymbol{x}_i, T_i, \delta_i, \boldsymbol{z}_i\}_{i=1}^n$ are assumed to be independent and identically distributed. We aim to estimate $(\boldsymbol{\theta}^{\mathsf{T}}, \mu(\cdot))^{\mathsf{T}} = (\boldsymbol{\beta}^{\mathsf{T}}, \boldsymbol{\phi}^{\mathsf{T}}, \boldsymbol{\zeta}_1^{\mathsf{T}}, \boldsymbol{\zeta}_2^{\mathsf{T}}, \boldsymbol{\gamma}^{\mathsf{T}}, \alpha_1, \boldsymbol{\alpha}_2, \mu(\cdot))^{\mathsf{T}}$ in the proposed joint models (1), (2), (3) and (4).

2.4 | Reparameterization of parameters

In real application, one could simplify the model by setting $\boldsymbol{D}_b = \sigma_b^2\boldsymbol{I}_3$ (that is, $\boldsymbol{\zeta}_2 = \sigma_b^2 > 0$), where $\boldsymbol{D}_b$ is the covariance matrix of random effects vector $\boldsymbol{b}_i$. The covariance matrix $\boldsymbol{\Sigma}_i$ of measurement errors $\boldsymbol{\epsilon}_i$ is often parameterized by two parameters: ρ and σ_ϵ^2 (that is, $\boldsymbol{\zeta}_1 = (\rho, \sigma_\epsilon^2)$), where ρ is correlation parameter and σ_ϵ^2 is variance parameter. An example for $\boldsymbol{\Sigma}_i$ with a compound symmetry structure can be found in Section 2.2. Sometimes the estimates $\hat{\rho}$, $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_b^2$ may be not well defined. That is, sometimes it is possible that some of $\hat{\rho} \notin (-1, 1)$, $\hat{\sigma}_\epsilon^2 < 0$ and $\hat{\sigma}_b^2 < 0$ may happen. In such a case, we suggest reparameterize ρ, σ_ϵ^2 and σ_b^2 as $\varrho = \log\{(1+\rho)/(1-\rho)\}$, $\varsigma_\epsilon = \log(\sigma_\epsilon^2)$ and $\varsigma_b = \log(\sigma_b^2)$, respectively. After estimating ϱ, ς_ϵ and ς_b, $\hat{\rho}$, $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_b^2$ can be obtained by corresponding transformation and their standard errors can be obtained by the delta method. Such reparameterization is used in our simulation studies and real data analysis.

3 | ESTIMATING PROCEDURES

There are two main challenges in developing the estimation method for the proposed joint models (1), (2), (3), and (4). First, the random effects $\boldsymbol{b}_i$ are not subject to a specific distributional assumption, making the joint likelihood function of the longitudinal data $\boldsymbol{Y}_i$ and the time-to-event data (T_i, δ_i) mathematically inexpressible, $i = 1, \ldots, n$. Consequently, traditional maximum likelihood estimates or Bayesian maximum posterior estimates become unattainable. Second, there is a challenge of simultaneously estimating the nonparametric component $\mu(\cdot)$ and its second derivative $\mu''(\cdot)$ as specified in the proposed joint models. To address these issues, we initially construct a working likelihood function to substitute for the joint likelihood function, followed by the derivation of an asymptotically unbiased estimating equation for the parameter $\boldsymbol{\theta}$, based on this working likelihood function alongside appropriately estimator $\hat{\mu}(\cdot)$ and $\hat{\mu}''(\cdot)$ respectively for $\mu(\cdot)$ and $\mu''(\cdot)$.

3.1 | Working likelihood functions

We begin with constructing working likelihood function for the longitudinal submodels (1), (3) and (4). For subject i, let $\boldsymbol{X}_i = (\boldsymbol{x}_i, ..., \boldsymbol{x}_i)^{\mathsf{T}}$ be an $n_i \times p_{\boldsymbol{\beta}}$ dimensional baseline covariate matrix, $\boldsymbol{G}_i$ be an $n_i \times 2$ dimensional matrix with j-th row $(1, t_{ij})$, $\boldsymbol{\mu}_i = (\mu(t_{i1}), ..., \mu(t_{in_i}))^{\mathsf{T}}$ and $\boldsymbol{E}_i = (\boldsymbol{G}_i, \boldsymbol{\mu}_i)$, where $p_{\boldsymbol{\beta}}$ is the dimension of $\boldsymbol{\beta}$. Quasi likelihood method, which does not impose distribution assumption on the random effect $\boldsymbol{b}_i$, is adopted.[24] Specifically, we use the following working log-likelihood function:

$$L_L(\boldsymbol{\theta}_1) = (2n)^{-1} \sum_{i=1}^{n} \left\{ -\log(|\boldsymbol{\Phi}_i|) - \boldsymbol{r}_i^{\mathsf{T}} \boldsymbol{\Phi}_i^{-1} \boldsymbol{r}_i \right\}, \tag{6}$$

where $\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{G}_i\phi - \boldsymbol{\mu}_i$ is a vector of residuals, $\boldsymbol{\Phi}_i = \boldsymbol{E}_i \boldsymbol{D}_b \boldsymbol{E}_i^{\mathsf{T}} + \boldsymbol{\Sigma}_i$ is a covariance matrix of $\boldsymbol{Y}_i$ and $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^{\mathsf{T}}, \phi^{\mathsf{T}}, \zeta_1^{\mathsf{T}}, \zeta_2^{\mathsf{T}})^{\mathsf{T}}$ is a vector of parameters involved in $L_L(\boldsymbol{\theta}_1)$. Although (6) represents a normal-type log-likelihood function, it can be established that the expectation of $\partial L_L(\boldsymbol{\theta}_1)/\partial \boldsymbol{\theta}_1$, evaluated at the true values of $\boldsymbol{\theta}_1$ and $\mu(\cdot)$, is zero, regardless of distribution of $\boldsymbol{b}_i$. Consequently, if $\mu(\cdot)$ is evaluated at its true value, $\partial L_L(\boldsymbol{\theta}_1)/\partial \boldsymbol{\theta}_1 = 0$ is an unbiased estimating equation for $\boldsymbol{\theta}_1$ and the working log-likelihood function (6) does not necessitate a distributional assumption on the random effects $\boldsymbol{b}_i$. See McCullagh for more details of the quasi likelihood method.[24]

We now pay attention to constructing working likelihood function for the survival submodel (2). Define $N_i(t) = I(T_i \leq t, \delta_i = 1)$ and $N_i^c(t) = I(T_i \leq t, \delta_i = 0)$ as counting processes corresponding to the observed survival outcomes T_i. Define $\mathcal{F}_i(t)$ as a right continuous filtration generated by $\{N_i(s), N_i^c(s), 0 \leq s \leq t\}$, $\boldsymbol{x}_i$, $\boldsymbol{z}_i$ and $\boldsymbol{b}_i$. Then, by direct calculation, the survival submodel (2) can also be written as $\mathbb{E}\{dN_i(t)|\mathcal{F}_i(t-)\} = \lambda_0(t)Q_i(t)$, where $\mathcal{F}_i(t-)$ is the left limitation of $\mathcal{F}_i(t)$,

$$Q_i(t) = \exp\left\{ \boldsymbol{z}_i^{\mathsf{T}}\boldsymbol{\gamma} + \alpha_1 \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{H}_i^{\mathsf{T}}(t)\boldsymbol{b}_i \right\} I(T_i \geq t), \tag{7}$$

$$\boldsymbol{H}_i(t) = \left(\alpha_1, \alpha_1 \int_0^t \omega_{\sigma_1}(t-s) s ds, \alpha_1 \int_0^t \omega_{\sigma_1}(t-s)\mu(s) ds + \alpha_2 \text{sign}(b_{i2}) \left[\int_0^t \omega_{\sigma_2}(t-s)\{\mu''(s)\}^2 ds \right]^{1/2} \right)^{\mathsf{T}}, \tag{8}$$

and $\text{sign}(b_{i2})$ is the sign of b_{i2}. That is, $\text{sign}(b_{i2}) = 1$ if $b_{i2} > 0$, $\text{sign}(b_{i2}) = -1$ if $b_{i2} < 0$ and $\text{sign}(b_{i2}) = 0$ if $b_{i2} = 0$.

We here consider an extension of the Corrected Score approach.[17] Recall that subject i has n_i longitudinal observations. If $n_i > 3$, we can estimate $\boldsymbol{b}_i$ based on all the longitudinal observations of subject i and calculate its weighted least square estimator $\hat{\boldsymbol{b}}_i = (\boldsymbol{E}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{E}_i)^{-1} \boldsymbol{E}_i^{\mathsf{T}} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$. Using standard results from the linear re-

gression, conditional on $\boldsymbol{b}_i$, the weighted least square estimator $\hat{\boldsymbol{b}}_i$ is normally distributed with mean $\boldsymbol{b}_i$ and variance $\boldsymbol{V}_i^{-1} = (\boldsymbol{E}_i^{\mathsf{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{E}_i)^{-1}$ as $\boldsymbol{\epsilon}_i$ follows from $N(\boldsymbol{0}_{n_i}, \boldsymbol{\Sigma}_i(\zeta_1))$. Note, $Q_i(t)$ in (7) is a term used to construct the partial likelihood function for Cox regression models.[25] Since $\boldsymbol{b}_i$ is unknown, $Q_i(t)$ is not available. We propose to use

$$\tilde{Q}_i(t) = \exp\left\{\boldsymbol{z}_i^{\mathsf{T}}\boldsymbol{\gamma} + \alpha_1\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{H}_i^{\mathsf{T}}(t)\hat{\boldsymbol{b}}_i - \frac{1}{2}\boldsymbol{H}_i^{\mathsf{T}}(t)\boldsymbol{V}_i^{-1}\boldsymbol{H}_i(t)\right\} I(T_i \geq t, n_i > 3),$$

instead, where $\tilde{Q}_i(t)$ is inspired by the following log-normal distribution property. For any Gaussian random variable X with mean μ_X and variance σ_X^2, we have that $\mathbb{E}\{\exp(X)\} = \exp(\mu_X + \sigma_X^2/2)$. Therefore, $\exp(X - \sigma_X^2/2)$ can be used as an unbiased estimate for $\exp(\mu_X)$. We then see that given $\mathcal{F}_i(t-)$, $\mathbb{E}\left\{\tilde{Q}_i(t)|\mathcal{F}_i(t-)\right\} = Q_i(t)I(n_i > 3)$. Therefore, a natural approach is to construct the working likelihood function by replacing $Q_i(t)$ in the partial likelihood function with $\tilde{Q}_i(t)$. Under our model setting, this idea requires $\text{sign}(b_{i2})$, which is involved in $\boldsymbol{H}_i(t)$ given by (8), to be known. When noise level is low or moderate, that is, when the variances of measurement errors ϵ_{ij}'s are of small or moderate scale, estimating $\text{sign}(b_{i2})$ by $\text{sign}(\hat{b}_{i2})$ is a good choice. This is because $\hat{\boldsymbol{b}}_i$ follows a multivariate normal distribution $N(\boldsymbol{b}_i, \boldsymbol{V}_i^{-1})$ given $\boldsymbol{b}_i$. As the variances of ϵ_{ij} decrease, $\hat{\boldsymbol{b}}_i$ becomes closer to $\boldsymbol{b}_i$ with high probability, making $\text{sign}(\hat{b}_{i2})$ a reasonable estimator of $\text{sign}(b_{i2})$. Here we focus on situation where the noise level is low or moderate and assume that $\text{sign}(b_{i2})$ is known from now on. A simulation in Section S12.2 of the supporting information evaluates the performance of estimating $\text{sign}(b_{i2})$ by $\text{sign}(\hat{b}_{i2})$. With the aforementioned arguments, we now consider the working log-likelihood function for the survival submodel (2), which is defined as

$$L_S(\boldsymbol{\theta}_2) = n^{-1}\sum_{i=1}^{n}\int_0^{\varphi}\left[\boldsymbol{z}_i^{\mathsf{T}}\boldsymbol{\gamma} + \alpha_1\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{H}_i^{\mathsf{T}}(t)\hat{\boldsymbol{b}}_i - \frac{1}{2}\boldsymbol{H}_i^{\mathsf{T}}(t)\boldsymbol{V}_i^{-1}\boldsymbol{H}_i(t) - \log\{\tilde{S}^{(0)}(t)\}\right] d\tilde{N}_i(t), \tag{9}$$

where $\tilde{N}_i(t) = I(n_i > 3)N_i(t)$, $\tilde{S}^{(0)}(t) = n^{-1}\sum_{i=1}^{n}\tilde{Q}_i(t)$, $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}^{\mathsf{T}}, \boldsymbol{\zeta}_1^{\mathsf{T}}, \boldsymbol{\gamma}^{\mathsf{T}}, \alpha_1, \alpha_2)^{\mathsf{T}}$ is a vector of parameters involved in $L_S(\boldsymbol{\theta}_2)$. Following similar argument in Wang, it can be established that the expectation of $\partial L_S(\boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}_2$, evaluated at the true values of $\boldsymbol{\theta}_2$ and $\mu(\cdot)$, is zero, regardless of distribution of $\boldsymbol{b}_i$.[17] Consequently, if $\mu(\cdot)$ is evaluated at its true value, $\partial L_S(\boldsymbol{\theta}_2)/\partial\boldsymbol{\theta}_2 = 0$ is an unbiased estimating equation for $\boldsymbol{\theta}_2$ and the working log-likelihood function (9) does not necessitate a distributional assumption on the random effects $\boldsymbol{b}_i$. See Wang for more details[17].

The working log-likelihood function $L_J(\boldsymbol{\theta})$ for the proposed joint models is then naturally constructed as the summation of $L_L(\boldsymbol{\theta}_1)$ and $L_S(\boldsymbol{\theta}_2)$ respectively defined in (6) and (9). That is,

$$L_J(\boldsymbol{\theta}) = L_L(\boldsymbol{\theta}_1) + L_S(\boldsymbol{\theta}_2). \tag{10}$$

3.2 | Asymptotically unbiased estimating equations

We can obtain the estimating equations for the unknown parameter $\boldsymbol{\theta}$ based on the derivative of working log-likelihood functions $L_J(\boldsymbol{\theta})$. First, we need to introduce the notations. Define $\tilde{\boldsymbol{S}}_{\boldsymbol{\beta}}^{(1)}(t)$, $\tilde{\boldsymbol{S}}_{\boldsymbol{\zeta}_1}^{(1)}(t)$, $\tilde{\boldsymbol{S}}_{\boldsymbol{\gamma}}^{(1)}(t)$, $\tilde{S}_{\alpha_1}^{(1)}(t)$, $\tilde{S}_{\alpha_2}^{(1)}(t)$ as the first-order partial derivatives of $\tilde{S}^{(0)}(t)$ with respect to $\boldsymbol{\beta}$, $\boldsymbol{\zeta}_1$, $\boldsymbol{\gamma}$, α_1 and α_2, respectively. Let $\boldsymbol{H}_{i\alpha_1}(t)$ and $\boldsymbol{H}_{i\alpha_2}(t)$ be the first-order partial derivatives of $\boldsymbol{H}_i(t)$ with respect to α_1 and α_2, respectively. Define $\partial\log(|\boldsymbol{\Phi}_i|)/\partial\boldsymbol{\zeta}_1$, $\partial\boldsymbol{\Phi}_i^{-1}/\partial\boldsymbol{\zeta}_1$, $\partial\hat{\boldsymbol{b}}_i/\partial\boldsymbol{\zeta}_1$ and $\partial\boldsymbol{V}_i^{-1}/\partial\boldsymbol{\zeta}_1$ as first-order partial derivatives of $\log(|\boldsymbol{\Phi}_i|)$, $\boldsymbol{\Phi}_i^{-1}$, $\hat{\boldsymbol{b}}_i$ and $\boldsymbol{V}_i^{-1}$ with respect to $\boldsymbol{\zeta}_1$, respectively. Explicit formulas for these derivatives can be found in Section S3 of the supporting information. The estimating equation for $\boldsymbol{\theta}$ is given by

$$\boldsymbol{U}(\boldsymbol{\theta}) = \left\{\boldsymbol{U}_{\boldsymbol{\beta}}^{\mathsf{T}}(\boldsymbol{\theta}), \boldsymbol{U}_{\phi}^{\mathsf{T}}(\boldsymbol{\theta}), \boldsymbol{U}_{\boldsymbol{\zeta}_1}^{\mathsf{T}}(\boldsymbol{\theta}), \boldsymbol{U}_{\boldsymbol{\zeta}_2}^{\mathsf{T}}(\boldsymbol{\theta}), \boldsymbol{U}_{\boldsymbol{\gamma}}^{\mathsf{T}}(\boldsymbol{\theta}), U_{\alpha_1}^{\mathsf{T}}(\boldsymbol{\theta}), U_{\alpha_2}^{\mathsf{T}}(\boldsymbol{\theta})\right\}^{\mathsf{T}} = \boldsymbol{0},$$

where

$$
\begin{aligned}
\boldsymbol{U}_{\beta}(\boldsymbol{\theta}) &= n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{\Phi}_i^{-1}\boldsymbol{r}_i + n^{-1}\sum_{i=1}^{n}\int_0^{\varphi}\left\{\alpha_1\boldsymbol{x}_i - \boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{E}_i\boldsymbol{V}_i^{-1}\boldsymbol{H}_i(t) - \frac{\tilde{\boldsymbol{S}}_{\beta}^{(1)}(t)}{\tilde{S}^{(0)}(t)}\right\}d\tilde{N}_i(t),\\
\boldsymbol{U}_{\phi}(\boldsymbol{\theta}) &= n^{-1}\sum_{i=1}^{n}\boldsymbol{G}_i^{\mathsf{T}}\boldsymbol{\Phi}_i^{-1}\boldsymbol{r}_i,\\
\boldsymbol{U}_{\zeta_1}(\boldsymbol{\theta}) &= (2n)^{-1}\sum_{i=1}^{n}\left\{-\frac{\partial\log(|\boldsymbol{\Phi}_i|)}{\partial\boldsymbol{\zeta}_1} - \boldsymbol{r}_i^{\mathsf{T}}\frac{\partial\boldsymbol{\Phi}_i^{-1}}{\partial\boldsymbol{\zeta}_1}\boldsymbol{r}_i\right\}\\
&\quad + n^{-1}\sum_{i=1}^{n}\int_0^{\varphi}\left[\left\{\frac{\partial\hat{\boldsymbol{b}}_i}{\partial\boldsymbol{\zeta}_1} - \frac{1}{2}\frac{\partial\boldsymbol{V}_i^{-1}}{\partial\boldsymbol{\zeta}_1}\boldsymbol{H}_i(t)\right\}^{\mathsf{T}}\boldsymbol{H}_i(t) - \frac{\tilde{\boldsymbol{S}}_{\zeta_1}^{(1)}(t)}{\tilde{S}^{(0)}(t)}\right]d\tilde{N}_i(t),\\
\boldsymbol{U}_{\zeta_2}(\boldsymbol{\theta}) &= (2n)^{-1}\sum_{i=1}^{n}\left\{-\frac{\partial\log(|\boldsymbol{\Phi}_i|)}{\partial\boldsymbol{\zeta}_2} - \boldsymbol{r}_i^{\mathsf{T}}\frac{\partial\boldsymbol{\Phi}_i^{-1}}{\partial\boldsymbol{\zeta}_2}\boldsymbol{r}_i\right\},\\
\boldsymbol{U}_{\gamma}(\boldsymbol{\theta}) &= n^{-1}\sum_{i=1}^{n}\int_0^{\varphi}\left\{\boldsymbol{z}_i - \frac{\tilde{\boldsymbol{S}}_{\gamma}^{(1)}(t)}{\tilde{S}^{(0)}(t)}\right\}d\tilde{N}_i(t),\\
U_{\alpha_1}(\boldsymbol{\theta}) &= n^{-1}\sum_{i=1}^{n}\int_0^{\varphi}\left[\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{H}_{i\alpha_1}^{\mathsf{T}}(t)\left\{\hat{\boldsymbol{b}}_i - \boldsymbol{V}_i^{-1}\boldsymbol{H}_i(t)\right\} - \frac{\tilde{S}_{\alpha_1}^{(1)}(t)}{\tilde{S}^{(0)}(t)}\right]d\tilde{N}_i(t),\\
U_{\alpha_2}(\boldsymbol{\theta}) &= n^{-1}\sum_{i=1}^{n}\int_0^{\varphi}\left[\boldsymbol{H}_{i\alpha_2}^{\mathsf{T}}(t)\left\{\hat{\boldsymbol{b}}_i - \boldsymbol{V}_i^{-1}\boldsymbol{H}_i(t)\right\} - \frac{\tilde{S}_{\alpha_2}^{(1)}(t)}{\tilde{S}^{(0)}(t)}\right]d\tilde{N}_i(t)
\end{aligned}
$$

Given that $\mu(\cdot)$ is evaluated at its true value and the parametric form of $\boldsymbol{\Sigma}_i$ is correctly specified (e.g., both the true and assumed forms are compound symmetry), the above estimating equations can be shown to be unbiased for $\boldsymbol{\theta}$. However, if the parametric form of $\boldsymbol{\Sigma}_i$ is incorrectly specified (e.g., true form is compound symmetry but we assume that $\boldsymbol{\Sigma}_i$ follows an independence structure), the estimating equations $U_{\alpha_1}(\boldsymbol{\theta}) = 0$ and $U_{\alpha_2}(\boldsymbol{\theta}) = 0$ for α_1 and α_2 in the survival submodel are biased, as shown in Proposition S3 in Section S6 of the supporting information. Since assuming $\boldsymbol{\Sigma}_i$ follows the independence structure is equivalent to assuming that measurement errors within the i-subject are mutually independent, Proposition S3 essentially highlights the cost of neglecting the correlation in longitudinal measurement errors. In fact, the proof of Proposition S3 reveals that such neglect can lead to more biased estimating equations for α_1 and α_2, particularly when the signal strengths of measurement errors and correlation are moderate to high. Otherwise, the biases may be small. Simulation studies in Section 4.2 and real data analysis in Section 5 evaluate the cost of neglecting the correlation in longitudinal measurements. More detailed discussion can be found in Section S15 of the supporting information.

We now discuss how to estimate $\mu(\cdot)$ and $\mu''(\cdot)$. The regression spline is adopted due to its simple mathematical expressions and good theoretical property. Let

$$
\begin{aligned}
\boldsymbol{F}(t) &= (t^2, ..., t^d, (t-u_1)_+^d, ..., (t-u_l)_+^d)^{\mathsf{T}},\\
\boldsymbol{F}''(t) &= (2, ..., d(d-1)t^{d-2}, d(d-1)(t-u_1)_+^{d-2}, ..., d(d-1)(t-u_l)_+^{d-2})^{\mathsf{T}},
\end{aligned}
\tag{11}
$$

where $\{u_j\}_{j=1}^{l}$ are spline knots, $d+1$ is order of the spline, $a_+ = \max\{a, 0\}$ for any scalar a. Then, according to the assumption (5), $\mu(\cdot)$ and $\mu''(\cdot)$ can be estimated by $\boldsymbol{F}^{\mathsf{T}}(\cdot)\boldsymbol{\xi}$ and $\boldsymbol{F}^{\mathsf{T},''}(\cdot)\boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is spline parameter to be estimated.

To estimate $\mu(\cdot)$ and $\mu''(\cdot)$, we need to estimate $\boldsymbol{\xi}$. Let $\boldsymbol{F}_i = (\boldsymbol{F}(t_{i1}), ..., \boldsymbol{F}(t_{in_i}))^{\mathsf{T}}$. Let

$$\boldsymbol{U}_{\boldsymbol{\xi}}(\boldsymbol{\theta}_1, \boldsymbol{\xi}) = n^{-1} \sum_{i=1}^{n} \boldsymbol{F}_i^{\mathsf{T}} \boldsymbol{\Phi}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \mathbf{G}_i \boldsymbol{\phi} - \boldsymbol{F}_i \boldsymbol{\xi}) = 0.$$

Define $L_L(\boldsymbol{\theta}_1, \boldsymbol{\xi})$ as a function obtained by replacing $\mu(\cdot)$ in $L_L(\boldsymbol{\theta}_1)$ by $\boldsymbol{F}^{\mathsf{T}}(\cdot)\boldsymbol{\xi}$. Let $\boldsymbol{U}_{L,\boldsymbol{\theta}_1}(\boldsymbol{\theta}_1, \boldsymbol{\xi})$ be the derivative of $L_L(\boldsymbol{\theta}_1, \boldsymbol{\xi})$ with respect to $\boldsymbol{\theta}_1$. Suppose $\boldsymbol{\theta}_{10}$ and $\mu_0(\cdot)$ are the true values of $\boldsymbol{\theta}_1$ and $\mu(\cdot)$, respectively. Then, Proposition S4 in Section S7 of the supporting information show the following results. After jointly solving $\boldsymbol{U}_{L,\boldsymbol{\theta}_1}(\boldsymbol{\theta}_1, \boldsymbol{\xi}) = 0$ and $\boldsymbol{U}_{\boldsymbol{\xi}}(\boldsymbol{\theta}_1, \boldsymbol{\xi}) = 0$, there exist solutions $\bar{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\xi}}$ such that $\bar{\boldsymbol{\theta}}_1$ converges to $\boldsymbol{\theta}_{10}$, $\hat{\mu}(\cdot)$ and $\hat{\mu}''(\cdot)$ uniformly converge to $\mu_0(\cdot)$ and $\mu_0''(\cdot)$ over $[0, \varphi]$ in probability, respectively, as the sample size n and the number of knots I diverge to infinity at appropriate rates, where $\hat{\mu}(\cdot) = \boldsymbol{F}^{\mathsf{T}}(\cdot)\hat{\boldsymbol{\xi}}$ and $\hat{\mu}''(\cdot) = \boldsymbol{F}^{\mathsf{T},''}(\cdot)\hat{\boldsymbol{\xi}}$. These results show that $\bar{\boldsymbol{\theta}}_1$ is already a consistent estimator of $\boldsymbol{\theta}_1$, and therefore, it can be set as the initial value of $\boldsymbol{\theta}_1$ in our algorithm. Additionally,

$$\boldsymbol{U}(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}) = 0, \tag{12}$$

is an asymptotically unbiased estimating equation for $\boldsymbol{\theta}$, where $\boldsymbol{U}(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}})$ is obtained by replacing $\mu(\cdot)$ and $\mu''(\cdot)$ in $\boldsymbol{U}(\boldsymbol{\theta})$ by $\hat{\mu}(\cdot)$ and $\hat{\mu}''(\cdot)$, respectively.

Theorem S1 and Theorem S2, which are respectively presented in Section S8 and S9 of the supporting information, show that the solution to (12), defined as $\hat{\boldsymbol{\theta}}$, converges in probability to its true value $\boldsymbol{\theta}_0$ and $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normally distributed. The asymptotic covariance of $\hat{\boldsymbol{\theta}}$ can be estimated by $n^{-1}\hat{\boldsymbol{\Sigma}}_J$, whose explicit mathematical expression is given in Section S10 of the supporting information.

The estimation equation (12) is derived from the SMRE model (4) under assumption (5). The SMRE model enables the estimation equation (12) to avoid imposing any distributional assumptions on the random effects $\boldsymbol{b}_i$, making our proposed method more applicable to practical problems. Additionally, it does not require the integration of random effects, which enhances the computational efficiency of our method.

3.3 | Profile search algorithm

We now discuss how to choose weight parameters σ_1 and σ_2 involved in $\eta_i(t, \sigma_1)$ and $\pi_i(t, \sigma_2)$, respectively. In related literature, Wang et al proposed a modified Newton-Raphson algorithm, while Mauff et al suggested using a Bayesian-based algorithm.[13,12] However, both algorithms are computationally intensive, which highlights the urgency of developing more computationally efficient method. Define $L_J\{\hat{\boldsymbol{\theta}}(s_1, s_2)\}$ as the joint working likelihood function with $\hat{\boldsymbol{\theta}}$ evaluated at $\sigma_1 = s_1, \sigma_2 = s_2$. We propose a profile search algorithm to choose σ_1 and σ_2, which is straightforward and intuitive. Initially, one may start with an initial guess with $\sigma_2 = \sigma_2^{(0)}$, one then search σ_1 over a given grid $\mathcal{A}_1$ such that the working likelihood function $L_J\{\hat{\boldsymbol{\theta}}(\sigma_1, \sigma_2^{(0)})\}$ is maximized. Denote the search result by $\sigma_1^{(1)}$, one then search σ_2 over a given grid $\mathcal{A}_2$ such that $L_J\{\hat{\boldsymbol{\theta}}(\sigma_1^{(1)}, \sigma_2)\}$ is maximized. Iterate between these two steps until convergence, the resulting search results $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are expected to be not far away from σ_{10} and σ_{20}, the true values of σ_1 and σ_2, respectively.

The above profile search algorithm is straightforward to implement, requiring not too much complexity in its setup, and is computationally efficient. In our experience, one or two iterations may be enough for convergence. However, the performance of the proposed profile search algorithm heavily relies on the specific choices of search grids, $\mathcal{A}_1$ and $\mathcal{A}_2$. Inappropriate grids may do not include σ_{10} and σ_{20}. In real application, $\mathcal{A}_1$ and $\mathcal{A}_2$ can be chosen based on prior knowledge from practitioners of related fields. And the number of grid points depends on user's prior knowledge of

the weight parameters and the desired precision of the estimator of weight parameters. Hesterman et al suggested that one may choose sufficient number of grid points such that the maximum distance between grid points is less than the desired precision for parameter estimator [26]. In our context, we aim to estimate weight parameters σ_1 and σ_2. Figure 3 in Wang et al indicates that when true values of weight parameters are small (e.g., less than 5), we should choose small precision for weight parameters, otherwise we may choose large precision for parameters [13]. We recommend that the users can roughly assess the true values of σ_{10} and σ_{20} based on their prior knowledge and then choose small/large precision for $\hat{\sigma}_1$ ($\hat{\sigma}_2$) when σ_{10} (σ_{20}) is small/large, and the users can choose sufficient number of grid points such that the maximum distance between grid points is less than the desired precision for $\hat{\sigma}_1$ ($\hat{\sigma}_2$). In our experience, 5 to 10 points in each dimension is sufficient for convergence in real applications.

3.4 | Model selection

In real application, one can select the most appropriate parametric structure for measurement error covariance $\boldsymbol{\Sigma}_i(\boldsymbol{\zeta}_1)$ among some commonly used candidate structures. For example, under the independence structure, $\boldsymbol{\Sigma}_i(\boldsymbol{\zeta}_1) = \sigma_\epsilon^2 \boldsymbol{I}_{n_i}$ and $\boldsymbol{\zeta}_1$ is simply σ_ϵ^2; under the CS structure, $\boldsymbol{\Sigma}_i(\boldsymbol{\zeta}_1) = \sigma_\epsilon^2(\rho \boldsymbol{J}_{n_i} + (1-\rho)\boldsymbol{I}_{n_i})$ and $\boldsymbol{\zeta}_1 = (\rho, \sigma_\epsilon^2)^{\mathsf{T}}$; under the first-order autoregressive (AR(1)) structure, $\boldsymbol{\Sigma}_i(\boldsymbol{\zeta}_1) = \sigma_\epsilon^2(\rho^{|j-k|})_{1\le j,k\le n_i}$ and $\boldsymbol{\zeta}_1 = (\rho, \sigma_\epsilon^2)^{\mathsf{T}}$. Spline basis functions and the location of interior knots can be set based on equi-quantile or equi-distant criteria. The structure $\mathcal{S}$ of $\boldsymbol{\Sigma}_i$, the degree d of spline basis functions and the number l of interior knots can be selected by minimizing criteria like AIC or BIC constructed based on the longitudinal submodel. These two criterion are defined as

$$\mathrm{AIC}(\mathcal{S}, d, l) = \frac{1}{n}\left[-2L_L\{\bar{\boldsymbol{\theta}}_1(\mathcal{S}, d, l), \hat{\boldsymbol{\xi}}(\mathcal{S}, d, l)\} + 2p_{\bar{\boldsymbol{\theta}}_1}\right], \tag{13}$$

$$\mathrm{BIC}(\mathcal{S}, d, l) = \frac{1}{n}\left[-2L_L\{\bar{\boldsymbol{\theta}}_1(\mathcal{S}, d, l), \hat{\boldsymbol{\xi}}(\mathcal{S}, d, l)\} + (\log n)p_{\bar{\boldsymbol{\theta}}_1}\right], \tag{14}$$

where $p_{\bar{\boldsymbol{\theta}}_1}$ is the dimension of $\bar{\boldsymbol{\theta}}_1$, $L_L\{\bar{\boldsymbol{\theta}}_1(\mathcal{S}, d, l), \hat{\boldsymbol{\xi}}(\mathcal{S}, d, l)\}$ represents the working likelihood function for the longitudinal submodel evaluated at $(\bar{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\xi}})$ with the structure of $\boldsymbol{\Sigma}_i$, the degree of spline basis functions and the number of interior knots fixed at $\mathcal{S}$, d and l, respectively. Minimization of (13) or (14) with respect to $\mathcal{S}$, d and l sometimes requires extensive computation. Alternatively, one may start with initial guesses for d and l, choose $\mathcal{S}$ by minimizing (13) or (14) , choose d and l in turn based on (13) or (14). In our experience, BIC (14) usually selects more parsimonious model. Sometimes $\boldsymbol{\Sigma}_i$ does not possess specific parametric structure, that is, $\boldsymbol{\Sigma}_i$ is unstructured. How to estimate unstructured $\boldsymbol{\Sigma}_i$ is of future research interest and we will briefly discuss this point in the discussion section.

3.5 | Algorithm

For clear exposition, we now present a simplified version algorithm of the proposed estimating procedure to obtain $\hat{\boldsymbol{\theta}}$, algorithm with more details can be found in Section S11 of the supporting information. In the following Algorithm, Newton-Raphson algorithm can used to solve the estimating equations involved in Step 2 and Step 4.

Step 1. Select appropriate $\mathcal{S}$, d and l based on AIC (13) or BIC (14).

Step 2. Solve equation $(\boldsymbol{U}_{L,\boldsymbol{\theta}_1}^{\mathsf{T}}(\boldsymbol{\theta}_1, \boldsymbol{\xi}), \boldsymbol{U}_{\boldsymbol{\xi}}^{\mathsf{T}}(\boldsymbol{\theta}_1, \boldsymbol{\xi}))^{\mathsf{T}} = 0$ to obtain $\bar{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\xi}}$. Then $\hat{\mu}(\cdot) = \boldsymbol{F}^{\mathsf{T}}(\cdot)\hat{\boldsymbol{\xi}}$ and $\hat{\mu}''(\cdot) = \boldsymbol{F}^{\mathsf{T},''}(\cdot)\hat{\boldsymbol{\xi}}$.

Step 3. Search appropriate $\hat{\sigma}_1$ and $\hat{\sigma}_2$ by the proposed profile search algorithm proposed in Section 3.3.

Step 4. Based on $\bar{\boldsymbol{\theta}}_1$, construct $\bar{\boldsymbol{\theta}}$ as the initial value of $\boldsymbol{\theta}$ (the initial value $\boldsymbol{\theta}_1$ can be set as $\bar{\boldsymbol{\theta}}_1$, and the initial values of γ, α_1 and α_2 can be simply set as zeros). Solve equation $\boldsymbol{U}(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}) = 0$ to obtain $\hat{\boldsymbol{\theta}}$.

4 | SIMULATION STUDIES

In this section, we conduct extensive simulation studies to evaluate several key performances of the proposed method: (i) the finite sample properties of the proposed parameter estimator $\hat{\boldsymbol{\theta}}$; (ii) the performance of the profile search algorithm; (iii) the impact of neglecting the correlation in measurement errors within the same subject on the parameter estimates for the survival submodel (2); and (iv) computational time of the proposed algorithm.

4.1 | Finite sample properties of parameter estimators

We consider two choices for $\mu(t)$ in SMRE model (4): (i) $\mu^{\dagger}(t) = t^2(\sin(t)+1)$; (ii) $\mu^{\ddagger}(t) = 0.25t^3 - t^2$. Note, these two functions satisfy the assumption (5). These two functions result in two specific population longitudinal trajectories: (i) $m^{\dagger}(t) = \phi_0 + \phi_1 t + \mu^{\dagger}(t)$; (ii) $m^{\ddagger}(t) = \phi_0 + \phi_1 t + \mu^{\ddagger}(t)$. The random effects $\boldsymbol{b}_i = (b_{i0}, b_{i1}, b_{i2})^{\mathsf{T}}$ are generated from the multivariate normal distribution or multivariate uniform distribution with $E(\boldsymbol{b}_i) = (0.2, 0.7, 1)^{\mathsf{T}}$ and $\boldsymbol{D}_b = 0.12\boldsymbol{I}_3$. This means that $\phi_0 = 0.2$, $\phi_1 = 0.7$ and $\zeta_2 = \sigma_b^2 = 0.12$. Then the population trajectory $m^{\dagger}(t)$ initially increases, decreases and then increases again. Another population trajectory $m^{\ddagger}(t)$ resembles the longitudinal trajectory which is flat at first and then increases. Shapes of the two population trajectories $m^{\dagger}(t)$ and $m^{\ddagger}(t)$ are plotted in Figure 2. Three choices for the weight parameters σ_1 and σ_2 are considered: (i) $\sigma_1 = \sigma_2 = 2$; (ii) $\sigma_1 = \sigma_2 = 4$; (iii) $\sigma_1 = \sigma_2 = 6$. The covariates $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are generated from the uniform distribution on interval $(-0.5, 0.5)$ and the Bernoulli distribution with mean 0.5, respectively. The covariance matrix $\boldsymbol{\Sigma}_i$ of the measurement errors are set to have the compound symmetric structure with $\boldsymbol{\zeta}_1 = (\rho, \sigma_\epsilon^2)^{\mathsf{T}} = (0.5, 0.1)^{\mathsf{T}}$. We set sample size $n = 500$. For each subject, the first longitudinal observation is taken at $t = 0$ and the subsequent longitudinal observations are taken every 0.5 unit until the event time. The generated longitudinal data is unbalanced. We set the search grids $\mathcal{A}_1 = \{\sigma_{10} - 0.3, \sigma_{10}, \sigma_{10} + 0.3\}$ and $\mathcal{A}_2 = \{\sigma_{20} - 0.3, \sigma_{20}, \sigma_{20} + 0.3\}$, where σ_{10} and σ_{20} are the true values of σ_1 and σ_2, respectively. Different baseline hazard functions are set for different settings for the weight parameters and the random effects, which together with detailed censoring generating mechanisms are deferred to the Section S12.1 of the supporting information. Censoring rate is around 0.3. The location and the number of interior knots are determined by the equi-distant criteria and BIC criteria (14).

[Figure 2 here.]

Model parameters are estimated in three ways: (i) the proposed parameter estimators with the true weight parameters σ_{10} and σ_{20} are known (KSP), where KSP means the known weight parameters; (ii) the proposed parameter estimators with the weight parameters searched by the proposed profile search algorithm (PS); (iii) the parameter estimators (WM) with the survival model misspecified as that in Wang et al,

$$\lambda_i(t) = \lambda_0(t) \exp\left(\boldsymbol{z}_i^{\mathsf{T}}\boldsymbol{\gamma} + \alpha_1 m_i(t) + \alpha_2 \left[\int_0^t \{m_i''(s)\}^2 ds\right]^{1/2}\right) \tag{15}$$

where WM means Wang's method.[10] All the simulation results are based on 500 Monte Carlo replications. We calculated the following summary statistics: the average of parameter estimates (Estimates), the Monte Carlo standard deviation (SD), the average of estimated standard errors (SE), the 95% empirical coverage probability (CP), the average of relative mean square errors (RMSE), the average of relative mean weighted integral errors (RMWIE), and the average

of relative mean weighted roughness integral errors (RMWRIE) over 500 replicates, where

$$\text{RMSE} = n^{-1}\sum_{i=1}^{n} n_i^{-1}\sum_{j=1}^{n_i}|\mu(t_{ij}) - \hat{\mu}(t_{ij})|^2\mu(t_{ij})^{-2},$$

$$\text{RMWIE} = n^{-1}\sum_{i=1}^{n}\left|\int_0^{T_i}\omega_{\sigma_1}(T_i - s)\{\mu(s) - \hat{\mu}(s)\}ds\right|^2\left|\int_0^{T_i}\omega_{\sigma_1}(T_i - s)\mu(s)ds\right|^{-2},$$

$$\text{RMWRIE} = n^{-1}\sum_{i=1}^{n}\left|\int_0^{T_i}\omega_{\sigma_2}(T_i - s)[\{\mu''(s)\}^2 - \{\hat{\mu}''(s)\}^2]ds\right|^2\left|\int_0^{T_i}\omega_{\sigma_2}(T_i - s)\{\mu''(s)\}^2ds\right|^{-2}$$

To save space, only simulation results for $\mu^{\dagger}(t)$ with $\sigma_1 = \sigma_2 = 4$ are presented here and the rest are shown in the S12 of the supporting information.

[Tables 1 here.]

Table 1 presents simulation results with the random effects generated from the multivariate normal distribution and the multivariate uniform distribution. With the known weight parameters, the proposed parameter estimators visually achieves consistency. Most parameter estimators have coverage probability close to 0.95 as well as similar SE and SD, indicating asymptotic normality. However, parameter estimator $\hat{\zeta}_2 = \hat{\sigma}_b^2$ shows slightly low coverage probability when the random effects are normally distributed, implying a small downward bias, which is not unexpected. Since the score equation for ζ_2 is derived by taking derivative of the function $L_L(\boldsymbol{\theta}_1)$ with respect to ζ_2 and $L_L(\boldsymbol{\theta}_1)$ may be viewed as an approximation to marginal likelihood of the longitudinal observations, $\hat{\zeta}_2$ can be viewed as an approximation to the maximum likelihood estimators (MLE). In finite sample situations, the MLE of variance component parameters tends to exhibit downward bias due to the lack of consideration for the loss in degrees of freedom resulting from the estimation of mean parameters β and ϕ.[27] Since our work aims to explore the effects of cumulative biomarker trajectory and cumulative variability on the event risk. Therefore, the parameters of primary interest are those that describe these effects. That is, α_1 and α_2. The parameter ζ_2 quantifies the variance of the random effects showing the level of subject heterogeneity but it does not provide information about how the cumulative trajectory and cumulative variability affect event risk. Therefore, ζ_2 is of minor interest and the bias in ζ_2 is acceptable. To obtain unbiased $\hat{\zeta}_2$, the restricted maximum likelihood techniques, which can correct the bias in the variance component estimators, are suggested.[27] The proposed profile search algorithm produces parameter estimates for the survival submodel (2) with similar performance to those estimated with the known weight parameters. Specifically, the estimates, the standard errors and the standard deviations of survival parameter estimators obtained by the profile search algorithm are similar to those obtained with the known σ_1 and σ_2. We also observe that if our model is misspecified as Wang' model (15), the parameter estimates are visually biased, which highlights the necessity for introducing the weight function $\omega_\sigma(\cdot)$ in the survival submodel (2). The averages of RMSE, RMWIE, RMWRIE are 0.0004, 0.0003 and 0.0014 for the normal random effects and 0.0003, 0.0004 and 0.0012 for the uniform random effects, indicating that the proposed estimators $\hat{\mu}(t)$, $\int_0^t \omega_{\sigma_1}(t - s)\hat{\mu}(s)ds$ and $\int_0^t \omega_{\sigma_2}(t - s)\{\hat{\mu}''(s)\}^2ds$ have good performance.

4.2 | Impact of overlooking correlation in longitudinal measurement errors

In this section, we use the same simulation settings as those in Section 4.1, except the true correlation parameter, ρ, is set as 0.8. We apply the proposed method to estimate model parameters, varying ρ from 0 to 0.9 in increments of 0.1. When ρ is not equal to 0.8, the correlation structure is misspecified. As discussed in Section 3.2, neglecting

correlation or misspecifying ρ as 0 may lead to biased estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Since $\rho = 0$ is a typical assumption adopted by existing methods, comparing the simulation results when $\rho = 0.8$ to those when $\rho = 0$ can help us to find the cost of misspecifying ρ as 0. For space consideration, we focus on simulation results for two parameters α_1 and α_2 with random effects being normally distributed, $\sigma_1 = \sigma_2 = 4$ and $\mu(t) = t^2(\sin(t) + 1)$. Simulation results for other cases are arranged in Section S12.4 of the supporting information.

[Figure 3 here.]

Figure 3 demonstrates relevant deviations in confidence bands and coverage probabilities of α_1 and α_2, which are constructed based on parameter estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ with their standard errors, as ρ varies from 0 to 0.9. Assuming ρ to be zero results in visually biased and inconsistent estimators $\hat{\alpha}_1$, and $\hat{\alpha}_2$ (e.g., coverage probabilities when $\rho = 0$ are clearly smaller than confidence level 0.95). However, when correlation structure is correctly specified and ρ is 0.8, these parameter estimators are consistent and asymptotically normally distributed, with coverage probabilities close to 0.95. As ρ approaches its true value 0.8, the performances of parameter estimators are more similar to those when ρ is equal to its true value. These findings highlight the importance of taking the correlation in the measurement errors into account, which is also where our method surpasses Wang's methods, which ignore the correlation.

We also observe that the standard errors of parameter estimators for survival submodel are robust with respect to misspecified correlation structures. To show this, we add a new case where the proposed method is implemented with a working correlation matrix as CS, while the true correlation matrix follows a AR(1) structure with the correlation parameter $\rho = 0.8$. For comparison, results for the case where the proposed method is implemented with both of the working the true correlation matrces follow a CS structure with $\rho = 0.8$ are also reported. Additionally, in order to support our discussion, average values of AIC and BIC across 500 replicates for selecting appropriate correlation parameter for measurement errors are added. These results are shown in Table 2.

[Table 2 here.]

From Table 2, we can see that the standard errors of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ demonstrate robustness against misspecified correlation structures. When the true correlation matrix follows AR(1) structure and the working correlation matrix follows CS structure, the SD and the SE exhibit similarities to the SE and SD where both the working and true correlation matrices follow CS structure. When both the true and working correlation matrices follow the CS structure, the average values of the AIC and the BIC reach their minimums when $\rho = 0.8$, corresponding to its true value. In fact, across 500 replicates, the proportion of correctly selecting the correlation parameter using the AIC and BIC is approximately 91%, demonstrating their effectiveness in selecting the appropriate correlation structure. When the true correlation matrix follows a AR(1) structure with $\rho = 0.8$, while the CS structure is employed as the working correlation, AIC and BIC tend to select CS structure with $\rho = 0.7$ or $\rho = 0.8$ as appropriate model, as they are close to the true correlation structure. In fact, average values of AIC and BIC when $\rho = 0.7$ are similar to those when $\rho = 0.8$. The proportion of selecting the CS structure with $\rho = 0.7$ is approximately 50% and the proportion of selecting the CS structure with $\rho = 0.8$ is approximately 48%. Note that the true value of correlation parameter ρ is 0.8, and each of AR(1) correlation matrix with $\rho = 0$ and CS correlation matrix with $\rho = 0$ is exactly the identity matrix. In Table 2, we observed that the reduction in coverage probability of the parameter estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ for the survival submodel, when the correlation parameter ρ is misspecified as 0 and the true correlation matrix follows a AR(1) structure, is smaller than the reduction when the true correlation matrix follows a CS structure. This is due to the fact that AR(1) correlation matrix with $\rho = 0.8$ is closer to identity matrix then CS correlation matrix with $\rho = 0.8$, and therefore misspecifying correlation parameter ρ as 0 leads to less biased parameter estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ when the true correlation matrix follows AR(1) structure.

4.3 | Computational time

We conduct a comparative analysis of the computational time between our proposed algorithm and the EM algorithm that incorporates a fully exponential Laplace approximation, as utilized by Wang et al [10,13] The simulation settings are identical to those detailed in Section 4.1. However, we have specifically focused on the scenario where the random effects are normally distributed, as this is a prerequisite for Wang's EM algorithm. The sample size for these simulations was $n = 500$ with an approximate censoring rate of 0.3. All algorithms were executed on a workstation equipped with an Intel Xeon Gold 6230R CPU at 2.10GHz with 26 cores and 256 GB of RAM. On average, the computational time required for each replication of our algorithm was 3.63 minutes, whereas implementing the EM algorithm with the fully exponential Laplace approximation without bootstrapping requires 96.06 minutes. These simulation results underscores the computational efficiency of our proposed method. Given that Wang's methods employ the bootstrap approach to estimate the standard errors of parameter estimators, Wang's methods can become computationally demanding when a large number of bootstrap resamples is utilized. This computational intensity is primarily due to the iterative nature of the bootstrap process, which requires repeated sampling and model estimation.

5 | REAL DATA ANALYSIS

In this section, we apply the proposed method to the ARIC study, which was described in the Section 1. It is widely recognized that the value of SBP (in mmHg) serves as a crucial indicator for CVD and the elevated SBP levels are often associated with an increased risk of cardiovascular mortality.[1] In the ARIC study, there are some subjects who take the anti-hypertensive medication and therefore have moderate levels of SBP. However, these subjects still experience CVD, indicating that there may be other factors contributing to their mortality. One potential factor is the variability of SBP, which we will investigate in this analysis. Given the extensive literature on the intricate mechanistic effects of diabetes on the cardiovascular mortality,[28] we narrow our focus to a specific subpopulation within the ARIC study. Specifically, we consider white male participants residing in Washington County and Suburban Minneapolis, who concurrently use the anti-hypertensive medication while being diagnosed with the diabetes. We utilize the available follow-up data, consisting of SBP measurements, time to CVD and corresponding censoring information up until 2016. The final sample comprises 200 subjects, among whom 85 experienced CVD events. The number of the SBP longtiduinal measurements varies from 4 to 6. The covariates considered in our analysis include the cumulative SBP trajectory, the cumulative SBP variability, as well as two baseline covariates: age (measured in years) and smoking status (represented by a binary indicator: 1 for current smokers at baseline examination and 0 otherwise). To facilitate analysis and interpretation, we rescale the variables. Specifically, age is divided by 10, SBP measurements are divided by 50, and all the event times, censoring times and longitudinal observation times are divided by 3650. To verify that whether the SMRE model (4) with assumption (5) is reasonable for the ARIC data, we conduct the FPCA for $g_i(t)$, the shapes of the mean function and first eigenfunction of $g_i(t)$ are plotted in Figure 4 , in which it is easy to see that the first eigenfunction has a similar shape as the mean function multiplied by scale factor 5 such that $\varphi_{i1}(t) \approx 5\mu(t)$. In addition, the first eigenfunction explains 90.142% of the variation of the $g_i(t)$. Therefore, the SMRE model (4) with assumption (5) are reasonable for the ARIC data.

[Figure 4 here]

We then consider the following longitudinal submodel:

$$\text{SBP}_{ij} = m_i(t_{ij}) + \epsilon_{ij}, \quad m_i(t) = \text{Age}_i\beta_1 + \text{Smoke}_i\beta_2 + f_i(t), \quad f_i(t) = b_{i0} + b_{i1}t + b_{i2}\mu(t),$$

and the following survival submodel:

$$\lambda_i(t) = \lambda_0(t) \exp\{\text{Age}_i\gamma_1 + \text{Smoke}_i\gamma_2 + \alpha_1\eta_i(t,\sigma_1) + \alpha_2\pi_i(t,\sigma_2)\}.$$

Here SBP_{ij} denotes the longitudinal measurement of SBP at time point t_{ij} for the j-th measurement on the i-th subject, Age_i represents the age of subject i and Smoke_i indicates whether subject i is a smoker or not. The regression parameters β_1 and β_2 correspond to the covariates Age_i and Smoke_i, respectively, in the longitudinal submodel. Similarly, the regression parameters γ_1 and γ_2 correspond to the covariates Age_i and Smoke_i, respectively, in the survival submodel. The regression parameters α_1 and α_2 describes the relationship between cumulative SBP trajectory and the hazard function, and the relationship between the cumulative SBP variability and hazard function, respectively.

To select appropriate structure for the covariance of measurement errors, we consider CS, AR(1), and independence structures as candidates. We assume $\boldsymbol{D}_b = \sigma_b^2\boldsymbol{I}_3$, which means $\boldsymbol{\zeta}_2$ is simply σ_b^2. After model selection based on the AIC and BIC in (13) and (14), we select a compound symmetric structure for $\boldsymbol{\Sigma}_i$, that is, $\boldsymbol{\Sigma}_i = \sigma_\epsilon^2(\rho\boldsymbol{J}_{n_i} + (1-\rho)\boldsymbol{I}_{n_i})$ and $\boldsymbol{\zeta}_1 = (\rho, \sigma_\epsilon^2)^\top$. The degree of spline basis functions, d, and the number of interior knots, l, are selected as 3 and 1, respectively. The AIC and BIC results for model selection are given Table 3. The location of the single knot is arranged based on the equi-distant criteria. By the profile search algorithm, the weight parameters σ_1 and σ_2 in survival submodel (2) are searched as 0.1 and 0.3, respectively. Parameters to be estimated are $\beta_1, \beta_2, \phi_0, \phi_1, \rho, \sigma_\epsilon^2, \sigma_b^2, \gamma_1, \gamma_2, \alpha_1, \alpha_2$.

[Tables 3-4 here.]

The estimated effects of age, smoke, cumulative SBP trajectory and cumulative SBP variability are summarized in Table 4. We consider a parameter to be statistically significant if the p-value of its estimator is less than the significance level of 0.05. For comparison, we also report the estimation results when ρ is assumed to be 0, which is commonly assumed by existing methods. The results presented in Table 4 provide evidence for a positive and significant association between the cumulative SBP variability and the time to CVD (with a p-value for $\hat{\alpha}_2$ less than 0.05). Specifically, an increase of 1 unit in cumulative SBP will result in the hazard function becoming 1.025 times its previous value, and therefore the increases the risk of CVD. These findings strongly support the notion that the cumulative SBP variability is related to the occurrence of the event. Additionally, we observed a significant and positive within-subject correlation in measurement errors, with $\hat{\rho} = 0.485$. When assuming a zero correlation (i.e., $\rho = 0$), the estimated $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are different from those under $\hat{\rho} = 0.485$, this is because the within-subject correlation in the measurement errors is ignored, leading to biases in $\hat{\alpha}_1$ and $\hat{\alpha}_2$, as predicted in Proposition S3 of the supporting information. Note, the absolute difference relative to the value of $\hat{\alpha}_2$ when accounting for correlation (i.e. relative difference) is (0.032 - 0.025)/0.025 = 0.28. This indicates that neglecting correlation leads to a 28% upward shift in $\hat{\alpha}_2$, compared to the estimator obtained when accounting for correlation. We consider a 28% relative difference to be substantial, suggesting that the association parameter estimator $\hat{\alpha}_2$ is affected by change in the correlation structure. When ignoring correlation in longitudinal measurement errors, the parameter estimates $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_b^2$, the estimated standard errors for $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\phi}_0$, $\hat{\phi}_1$, $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_b^2$ are very different from those estimated by the model taking the within-subject correlation in measurement errors into account. This is due to the fact that neglecting the correlation may lead to efficiency loss,[29] which may obtain inefficient parameter estimators when the covariance matrix $\boldsymbol{\Phi}_i$ is misspecified (i.e., when assuming $\rho = 0$). Specifically, the random effect variance estimator $\hat{\sigma}_b^2$ is even not statistically significant when assuming $\rho = 0$, indicating that the random effect vector $\boldsymbol{b}_i$, which captures the heterogeneity of SBP trajectories in different subjects, is a constant vector. This is obviously a misleading conclusion and suggests that ignoring the within-subject correlation in the measurement errors may lead to misleading inference. Additionally, we have observed that the estimated regression parameter $\hat{\gamma}_1$ corresponding to the covariate of age is statistically significant. This indicates that older subjects

are more likely to experience CVD. Although previous research, such as Mok et al, has shown a positive association between baseline SBP and the risk of myocardial infarction, we find that the cumulative SBP trajectory in our analysis is not statistically associated with the CVD risk. This discrepancy may be attributed to the fact that we focus on a subpopulation who take the anti-hypertensive medication and have the diabetes while Mok et al considered another subpopulation who developed myocardial infarction.[30] Ma and Pan, who studied another subpopulation of patients living in Washington County and Suburban Minneapolis, also found that SBP values were also not significantly related to risk of CVD.[31]

[Figures 5-6 here.]

Figure 5 plots the observed SBP trajectories, the fitted SBP trajectories, the fitted cumulative SBP trajectories and the fitted cumulative SBP variability for randomly selected 5 subjects in the ARIC study. It illustrates that the fitted SBP trajectories obtained using our methods closely match the observed SBP trajectories. Upon visual inspection, we observe that the subject with trajectory depicted in red display the highest SBP variability. This is evident in the fitted cumulative SBP variability, as shown in the bottom right sub-figure of Figure 6. This observation suggests that the weighted roughness measure $\pi_i(t, \sigma_2)$ appropriately captures the variability of the biomarker.

As depicted in Figure 6, the estimated population SBP (EPSBP) trajectory ranges from 2.5 to 2.625, corresponding to a range of 125 to 131.25 (since SBP is divided by 50 in our analysis) for the value of SBP. This range of SBP changes aligns with what is commonly observed in hypertensive patients who are undergoing anti-hypertensive medication.[32] This finding suggests that our method effectively captures the pattern of the ARIC data. Additionally, the estimated population cumulative SBP (EPCSBP) trajectory exhibits a generally flat trend followed by an uptrend at larger time points. Furthermore, the estimated population cumulative SBP variability initially decreases and then increases. This pattern is consistent with the shape of the EPSBP. Overall, our method demonstrates good performance in analyzing the ARIC data.

Under the proposed SMRE model (4) with assumption (5), our proposed method neither imposes distribution assumption on random effects nor involves their integration, which makes our proposed method very applicable for real problems and computationally efficient. Since the integrals involved in the cumulative trajectory and the cumulative variability do not permit analytical expression, the computational burden mainly comes from calculating these integrals via numerical approximation. The computation time for implementing the proposed algorithm for the ARIC data is 5.364 min using a workstation with Intel Xeon Gold 6230R 2.10GHz 26 cores CPU and 256G RAM.

6 | DISCUSSION

In this paper, we propose a novel approach to estimate the association between the event time and the cumulative biomarker variability. Our method takes the correlation in the longitudinal measurement errors within the same subject into account, resulting in unbiased parameter estimators in survival submodel (2). The SMRE model (4) is proposed to capture the population trajectory which is approximately linear when time is small while nonlinear when time is large. It assumes that the biomarker trajectory can be represented as sum of a linear term and a nonlinear term. And the linear term dominates the nonlinear term when time is small. This SMRE model enables the construction of an estimating equation that does not impose any distributional assumption on the random effects and does not require their integration. As a result, the proposed method is both more flexible and computationally efficient compared to existing approaches. As the sample size and the number of knots go to infinity at appropriate rates, consistency and asymptotic normality of the proposed estimators are established.

Biomarker trajectory in longitudinal data often exhibits the pattern captured by the SMRE model (4) and assumption (5). In the early stages, biomarker values tend to linearly increase, decrease, or remain stable at a steady rate. Possible reasons include patients' relatively better health conditions at early stage of disease or the effectiveness of taking medication. However, as the patient's condition worsens, biomarker values often show nonlinear and irregular changes due to declining health. For instance, the frequently studied AIDS data in Figure 3 of Yao indicates that, for those patients taking zalcitabine, the population CD4 trajectory is linearly decreasing in the early stage of AIDS but becomes nonlinear over time.[20] Similarly, the PBC data in Figure 3 of Ding and Wang shows that the population serum bilirubin trajectory is linearly increasing in the early stage but becomes nonlinear over time.[21]

In the construction of the working likelihood function $L_S(\boldsymbol{\theta}_2)$ in (9) for the survival submodel, we only consider data from subjects with greater than three longitudinal measurements . This is because we need to ensure the existence of the weighted least squares estimate for the three-dimensional random effects $\boldsymbol{b}_i = (b_{i0}, b_{i1}, b_{i2})^\top$. However, this does not mean that we should discard data from subjects with fewer than three longitudinal measurements when constructing the working likelihood function $L_L(\boldsymbol{\theta}_1)$ in (6) for the longitudinal submodel, as data from these subjects contribute to the longitudinal working likelihood function.

In Section 3.4, we suggest to select the appropriate covariance structure of $\boldsymbol{\Sigma}_i$ among some parametric candidates. However, it is important to acknowledge that in real-world scenarios, $\boldsymbol{\Sigma}_i$ may exhibit a completely unstructured pattern and may not be accurately captured by any of candidates. In such a case, the modified Cholesky decomposition (MCD) may be a promising approach for modeling $\boldsymbol{\Sigma}_i$, since MCD is widely used to build model for unstructured covariance matrices and can provide flexible and effective representations of the underlying correlation structure.[33] Although we do not delve into this specific modeling approach in this paper, we recognize its potential and leave it as an avenue for future research.

Supporting Information

Supporting information for this article is available online. Section S1 includes proofs and a simulation study which provide evidence that the Conditional Score method and the Corrected Score method may lead to biased parameter estimates in survival submodel when neglecting correlation in longitudinal measurement errors. Section S2 includes a simulation study evaluating the cost of misspecifying the SMRE model (4) with assumption (5). Section S3 gives explicit expression for $\boldsymbol{U}(\boldsymbol{\theta})$. Sections S4-S9 present rigorous proofs for theoretical results with required regularity conditions and useful lemmas. Section S10 gives explicit expression for $\hat{\Sigma}_J$. Section S11 presents a detailed version of algorithm to obtain $\hat{\boldsymbol{\theta}}$. Section S12 presents additional simulation results. Section S13 presents a discussion for interpreting the impact of the Gaussian weight function $\omega_\sigma(t-s)$ on the cumulative trajectory and cumulative variability at current time t. Section S14 discusses why it is not appropriate to describe the cumulative variability using absolute value of derivative of trajectory. Finally, Section S15 gives detailed discussion for the impact of neglecting correlation in longitudinal measurement errors on parameter estimators for the survival submodel.

Acknowledgements

The authors thank the anonymous AE and reviewers for their valuable suggestions. This work is supported by the National Natural Science Foundation of China (12271047), and by the Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College (2022B1212010006).

References

1. Rothwell PM. Limitations of the usual blood-pressure hypothesis and importance of variability, instability, and episodic hypertension. *Lancet.* 2010; 375(9718):938-948.

2. Rothwell PM, Howard SC, Dolan E, et al. Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. *Lancet.* 2010;375(9718):895-905.

3. Rothwell PM, Howard SC, Dolan E, et al. Effects of β blockers and calcium-channel blockers on within-individual variability in blood pressure and risk of stroke. *Lancet Neurol.* 2010;9(5):469-480.

4. Arroyo-Carmona RE, López-Serrano AL, Albarado-Ibañez A, et al. Heart rate variability as early biomarker for the evaluation of diabetes mellitus progress. *J Diabetes Res.* 2016;2016(1):8483537.

5. Carr MJ, Bao Y, Pan J, et al. The predictive ability of blood pressure in elderly trial patients. *J Hypertens.* 2012;30(9):1725-1733.

6. Crowther MJ, Andersson TML, Lambert PC, et al. Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Stat Med.* 2016;35(7);1193-1209.

7. Gao F, Miller JP, Xiong C, et al. A joint-modeling approach to assess the impact of biomarker variability on the risk of developing clinical outcome. *Stat Methods Appt.* 2011;20(1):83-100.

8. Barrett JK, Huille R, Parker R, et al. Estimating the association between blood pressure variability and cardiovascular disease: An application using the ARIC Study. *Stat Med.* 2019;38(10):1855-1868.

9. Howard SC, Rothwell PM. Reproducibility of measures of visit-to-visit variability in blood pressure after transient ischaemic attack or minor stroke. *Cerebrovasc Dis.* 2019;28(4):331-340.

10. Wang C, Shen J, Charalambous C, et al. Modeling biomarker variability in joint analysis of longitudinal and time-to-event data. *Biostat.* 2024;25(2):577-596.

11. Rizopoulos D, Hatfield LA, Carlin BP, et al. Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *J Am Stat Assoc.* 2014;109(508):1385-1397.

12. Mauff K, Steyerberg EW, Nijpels G, et al. Extension of the association structure in joint models to include weighted cumulative effects. *Stat Med.* 2017;36(23):3746-3759.

13. Wang C, Shen J, Charalambous C, et al. Weighted biomarker variability in joint analysis of longitudinal and time to event data. *Ann Appl Stat.* 2024:18(3);2576-2595.

14. Vacek PM. Assessing the effect of intensity when exposure varies over time. *Stat Med.* 1997:16(5);505-513.

15. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin.* 2004:14(3);809-834.

16. Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika.* 2001;88(2):447-458.

17. Wang C. Corrected score estimator for joint modeling of longitudinal and failure time data. *Stat Sin.* 2006;16(1):235-253.

18. Rizopoulos D. *Joint models for longitudinal and time-to-event data: With applications in R*. Boca Raton, FL: CRC Press; 2012.

19. Pinto E. Blood pressure and ageing. *Postgrad Med J.* 2007;83(976):109-114.

20. Yao F. Functional principal component analysis for longitudinal and survival data. *Stat Sin.* 2007;17(3);965-983.

21. Ding J, Wang JL. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics.* 2008;64(2);546-556.

22. Alsefri M, Sudell M, García-Fiñana M, et al. Bayesian joint modelling of longitudinal and time to event data: a methodological review. *BMC Med Res Methodol.* 2020;20(4):1-17.

23. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics.* 1982;38(4):963-974.

24. McCullagh P. Quasi-likelihood functions. *Ann Stat.* 1983;11(1):59-67.

25. Fleming TR, Harrington DP. *Counting processes and survival analysis*. John Wiley & Sons. 2013.

26. Hesterman JY, Caucci L, Kupinski MA, et al. Maximum-likelihood estimation with a contracting-grid search algorithm. *IEEE Trans Nucl Sci.* (2010);57(3):1077-1084.

27. Lindstrom MJ, Bates DM. Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J Am Stat Assoc.* 1988;83(404):1014-1022.

28. Sowers JR, Epstein M, Frohlich ED. Diabetes, hypertension, and cardiovascular disease: an update. *Hypertension.* 2001;37(4):1053-1059.

29. Diggle P. *Analysis of longitudinal data*. Oxford university press. 2002.

30. Mok Y, Sang Y, Ballew SH, et al. American heart association's life's simple 7 at middle age and prognosis after myocardial infarction in later life. *J Am Heart Assoc.* 2018;7(4);e007658.

31. Ma C, Pan J. Multistate analysis of multitype recurrent event and failure time data with event feedbacks in biomarkers. *Scand J Stat.* 2022;49(2):864-885.

32. Arguedas JA, Leiva V, Wright JM. Blood pressure targets in adults with hypertension. *Cochrane Database Syst Rev.* 2020;12.

33. Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika.* 1999;86(3):677-690.

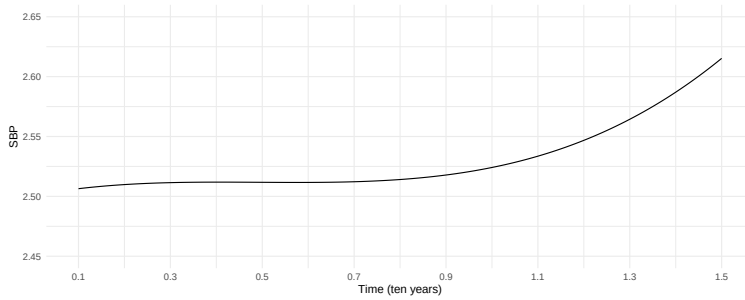

FIGURE 1 Estimated population SBP (EPSBP) for ARIC study.

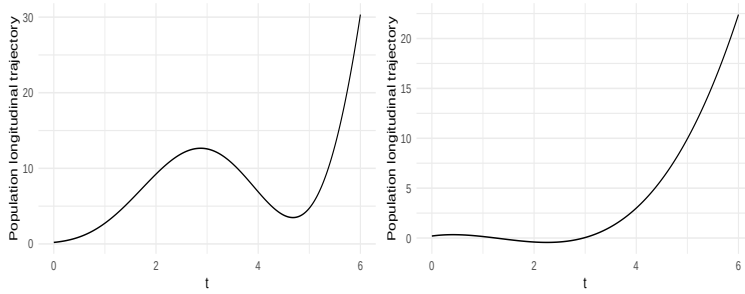

FIGURE 2 Shapes of two population longitudinal trajectories $m^{\dagger}(t)$ (left figure) and $m^{\ddagger}(t)$ (right figure).

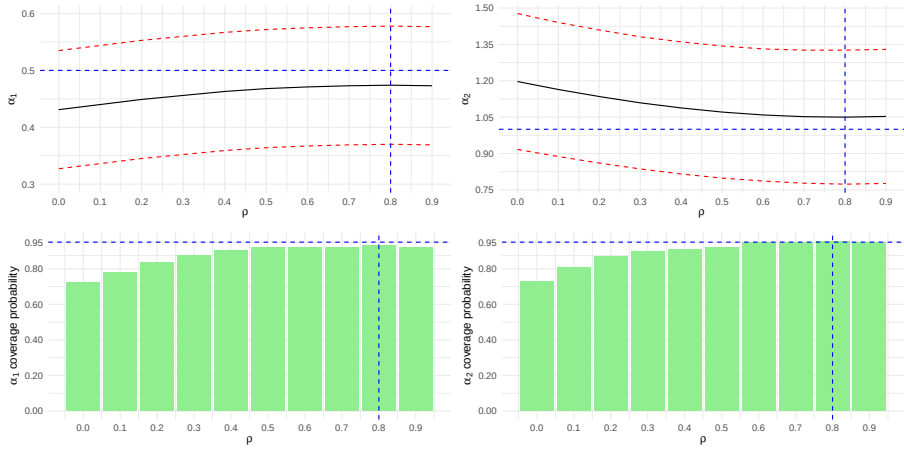

FIGURE 3 Impact of overlooking correlation in longitudinal measurement errors for $\mu(t) = t^2(\sin(t) + 1)$, $\sigma_1 = \sigma_2 = 4$ and normal random effects. Top: averaged estimates of α_1 and α_2 with 95% confidence interval. Bottom: histogram of coverage probability of α_1 and α_2. Black solid line: parameter estimates. Red dashed line: 95% confidence interval. Blue horizontal line: true value of α_1 or α_2. Blue vertical line: true value of correlation parameter ρ.

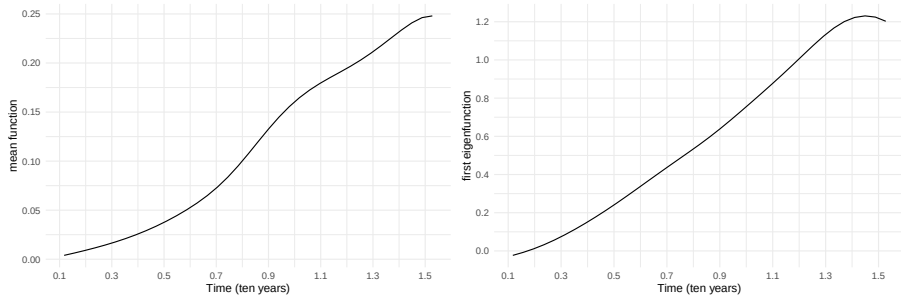

FIGURE 4 The mean function and the first eigenfunction (right) of $g_i(t)$ for the ARIC data. The eigenvalue of the first eigenfunction is 1.149 and the summation of all the eigenvalues are 1.274, which means the first eigenfunction explains 90.142% of the variation of $g_i(t)$.

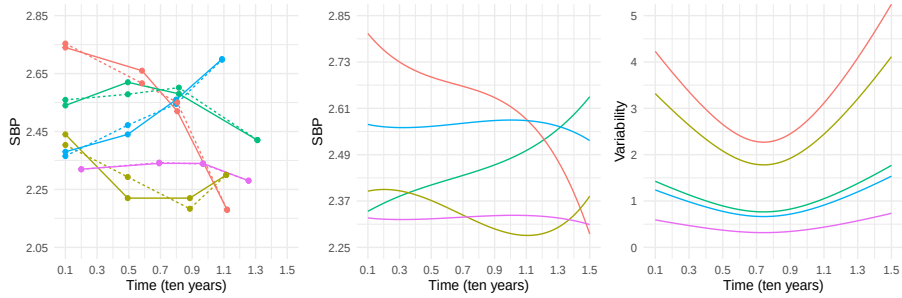

FIGURE 5 SBP trajectories (solid line for observed trajectories and dotted line for fitted trajectories), fitted cumulative SBP trajectories and fitted cumulative SBP variability for randomly selected 5 subjects in the Atherosclerosis Risk in Communities data.

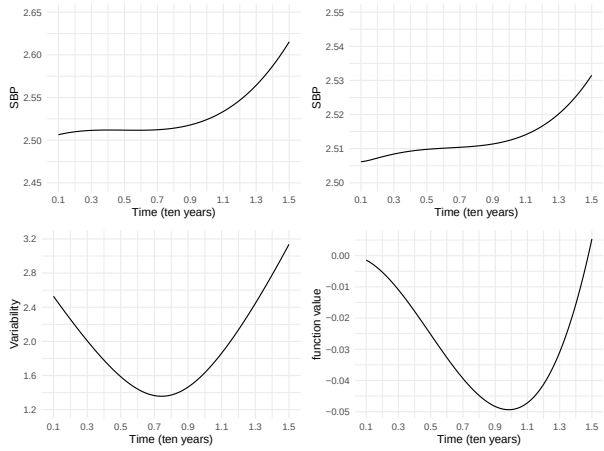

FIGURE 6 EPSBP Trajectory (top left), EPCSBP Trajectory (top right), EPCSBP Variability (bottom left) and $\mu(t)$ of the SMRE model (bottom right) in the ARIC data.

TABLE 1 Simulation Studies for $\mu^{\dagger}(t) = t^2(\sin(t)+1)$ with $\sigma_1 = \sigma_2 = 4$.

Parameter		KSP				PS				WM			
	True value	Estimates	SE	SD	CP	Estimates	SE	SD	CP	Estimates	SE	SD	CP
Normal Random Effects													
β	0.700	0.701	0.069	0.073	0.936	0.701	0.069	0.073	0.936	0.701	0.069	0.073	0.936
ϕ_0	0.200	0.198	0.021	0.020	0.956	0.198	0.021	0.020	0.956	0.198	0.021	0.020	0.956
ϕ_1	0.700	0.707	0.035	0.036	0.936	0.707	0.035	0.036	0.936	0.707	0.035	0.036	0.936
ρ	0.500	0.513	0.069	0.078	0.924	0.513	0.069	0.078	0.924	0.513	0.069	0.078	0.924
σ_ϵ^2	0.100	0.105	0.015	0.020	0.922	0.105	0.015	0.020	0.922	0.105	0.015	0.020	0.922
σ_b^2	0.120	0.115	0.005	0.006	0.906	0.115	0.005	0.006	0.906	0.115	0.005	0.006	0.906
γ	-1.000	-0.996	0.116	0.120	0.944	-0.997	0.116	0.121	0.944	-0.058	0.156	0.375	0.044
α_1	0.500	0.496	0.054	0.055	0.940	0.492	0.053	0.055	0.944	-0.213	0.042	0.114	0.000
α_2	1.000	0.989	0.144	0.139	0.960	0.983	0.144	0.139	0.956	-0.834	0.143	0.357	0.000
Uniform Random Effects													
β	0.700	0.697	0.070	0.072	0.940	0.697	0.070	0.072	0.940	0.697	0.070	0.072	0.940
ϕ_0	0.200	0.199	0.021	0.022	0.934	0.199	0.021	0.022	0.934	0.199	0.021	0.022	0.934
ϕ_1	0.700	0.707	0.035	0.036	0.932	0.707	0.035	0.036	0.932	0.707	0.035	0.036	0.932
ρ	0.500	0.486	0.085	0.098	0.924	0.486	0.085	0.098	0.924	0.486	0.085	0.098	0.924
σ_ϵ^2	0.100	0.100	0.017	0.019	0.920	0.100	0.017	0.019	0.920	0.100	0.017	0.019	0.920
σ_b^2	0.120	0.119	0.009	0.010	0.936	0.119	0.009	0.010	0.936	0.119	0.009	0.010	0.936
γ	-1.000	-0.996	0.120	0.120	0.960	-0.996	0.121	0.121	0.960	-0.376	0.152	0.341	0.166
α_1	0.500	0.505	0.057	0.057	0.950	0.501	0.056	0.057	0.946	-0.164	0.043	0.114	0.000
α_2	1.000	0.958	0.159	0.160	0.946	0.949	0.158	0.160	0.936	-1.015	0.152	0.373	0.000

TABLE 2 Correlation misspecification results for $\mu^{\dagger}(t) = t^2(\sin(t)+1)$ with $\sigma_1 = \sigma_2 = 4$ and the normal random effects. Compound symmetry is used as working correlation structure in both of left and right half of the table. CS: compound symmetry. AR(1): first-order autoregressive. The true value of correlation parameter is $\rho = 0.8$. The proportion of correctly selecting the correlation parameter using the AIC and BIC is approximately 91% when CS is used as true correlation structure

True correlation structure: CS							True correlation structure: AR(1)						
	True value	0.5	1		True value	0.5	1		True value	0.5	1		True value	0.5	1
$\rho = 0$		α_1	α_2	$\rho = 0.5$		α_1	α_2	$\rho = 0$		α_1	α_2	$\rho = 0.5$		α_1	α_2
AIC 23.580	Estimates	0.431	1.197	AIC 23.538	Estimates	0.468	1.071	AIC 25.268	Estimates	0.468	1.086	AIC 25.231	Estimates	0.491	1.034
BIC 5.692	SE	0.053	0.144	BIC 5.650	SE	0.054	0.145	BIC 7.380	SE	0.048	0.138	BIC 7.343	SE	0.048	0.139
	SD	0.053	0.143		SD	0.053	0.139		SD	0.055	0.148		SD	0.055	0.144
	CP	0.724	0.730		CP	0.918	0.920		CP	0.891	0.876		CP	0.923	0.923
$\rho = 0.1$		α_1	α_2	$\rho = 0.6$		α_1	α_2	$\rho = 0.1$		α_1	α_2	$\rho = 0.6$		α_1	α_2
AIC 23.575	Estimates	0.440	1.164	AIC 23.522	Estimates	0.471	1.059	AIC 25.263	Estimates	0.475	1.062	AIC 25.219	Estimates	0.492	1.031
BIC 5.687	SE	0.053	0.144	BIC 5.634	SE	0.054	0.146	BIC 7.375	SE	0.048	0.138	BIC 7.331	SE	0.048	0.139
	SD	0.053	0.141		SD	0.053	0.139		SD	0.055	0.146		SD	0.055	0.144
	CP	0.782	0.806		CP	0.919	0.946		CP	0.902	0.888		CP	0.925	0.928
$\rho = 0.2$		α_1	α_2	$\rho = 0.7$		α_1	α_2	$\rho = 0.2$		α_1	α_2	$\rho = 0.7$		α_1	α_2
AIC 23.568	Estimates	0.449	1.135	AIC 23.502	Estimates	0.473	1.052	AIC 25.257	Estimates	0.481	1.050	AIC 25.207	Estimates	0.493	1.028
BIC 5.680	SE	0.053	0.145	BIC 5.614	SE	0.054	0.146	BIC 7.369	SE	0.048	0.138	BIC 7.319	SE	0.048	0.139
	SD	0.053	0.140		SD	0.053	0.140		SD	0.055	0.145		SD	0.055	0.146
	CP	0.836	0.868		CP	0.922	0.948		CP	0.908	0.899		CP	0.926	0.932
$\rho = 0.3$		α_1	α_2	$\rho = 0.8$		α_1	α_2	$\rho = 0.3$		α_1	α_2	$\rho = 0.8$		α_1	α_2
AIC 23.560	Estimates	0.456	1.109	AIC **23.487**	Estimates	**0.474**	**1.050**	AIC 25.250	Estimates	0.485	1.042	AIC **25.206**	Estimates	0.494	1.026
BIC 5.672	SE	0.053	0.145	BIC **5.598**	SE	**0.054**	**0.146**	BIC 7.362	SE	0.048	0.138	BIC **7.318**	SE	0.048	0.140
	SD	0.053	0.139		SD	**0.053**	**0.141**		SD	0.055	0.145		SD	0.055	0.145
	CP	0.874	0.896		CP	**0.934**	**0.952**		CP	0.914	0.909		CP	0.927	0.936
$\rho = 0.4$		α_1	α_2	$\rho = 0.9$		α_1	α_2	$\rho = 0.4$		α_1	α_2	$\rho = 0.9$		α_1	α_2
AIC 23.551	Estimates	0.463	1.088	AIC 23.542	Estimates	0.473	1.053	AIC 25.242	Estimates	0.489	1.038	AIC 25.300	Estimates	0.492	1.029
BIC 5.662	SE	0.055	0.145	BIC 5.654	SE	0.055	0.146	BIC 7.354	SE	0.048	0.139	BIC 7.4115	SE	0.048	0.140
	SD	0.053	0.139		SD	0.053	0.141		SD	0.055	0.144		SD	0.055	0.146
	CP	0.902	0.908		CP	0.920	0.948		CP	0.919	0.917		CP	0.925	0.933

TABLE 3 Model selection results for ARIC data analysis. d: degree of spline basis functions. l: number of spline knots. $\mathcal{S}$: structure of covariance of measurement errors within the same subject. In: independence structure CS: compound symmetry structure. AR: first-order autoregressive structure.

	$(d, l, \mathcal{S})$					
	(3, 1, In)	(3, 2, In)	(3, 3, In)	(4, 1, In)	(4, 2, In)	(4, 3, In)
AIC	-6.055	-6.042	-6.040	-6.041	-6.028	-5.700
BIC	-5.890	-5.860	-5.842	-5.860	-5.830	-5.486
	(3, 1, CS)	(3, 2, CS)	(3, 3, CS)	(4, 1, CS)	(4, 2, CS)	(4, 3, CS)
AIC	**-6.196**	-6.181	-6.186	-6.183	-6.175	-6.171
BIC	**-6.031**	-6.000	-5.988	-6.002	-5.977	-5.956
	(3, 1, AR)	(3, 2, AR)	(3, 3, AR)	(4, 1, AR)	(4, 2, AR)	(4, 3, AR)
AIC	-6.105	-6.091	-6.093	-6.099	-6.084	-6.041
BIC	-5.940	-5.909	-5.895	-5.910	-5.887	-5.827

TABLE 4 Estimation Results for the Atherosclerosis Risk in Communities data.

	Model with correlated measurement errors			Model assuming independent measurement errors		
Parameter	Estimate	SE	P value	Estimate	SE	P value
β_1	0.004	0.034	0.906	-0.010	0.165	0.952
β_2	-0.034	0.045	0.450	0.012	0.220	0.957
ϕ_0	2.500	0.023	**<0.001**	2.508	0.097	**<0.001**
ϕ_1	0.073	0.017	**<0.001**	0.069	0.098	0.481
ρ	0.485	0.050	**<0.001**	-	-	-
σ_ϵ^2	0.096	0.009	**<0.001**	0.030	0.003	**<0.001**
σ_b^2	0.012	0.004	**0.003**	0.004	0.004	0.317
γ_1	0.614	0.200	**0.002**	0.589	0.188	**0.001**
γ_2	0.218	0.250	0.383	0.089	0.260	0.732
α_1	0.013	0.036	0.361	0.014	0.036	0.697
α_2	0.025	0.009	**0.006**	0.032	0.007	**<0.001**