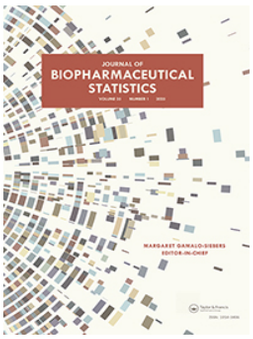

Taylor & Francis
Taylor & Francis Group

Journal of Biopharmaceutical Statistics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lbps20

A Bayesian joint bent-cable model for longitudinal measurements and survival time with heterogeneous random-effects distributions

Oludare Ariyo, Kehinde Olobatuyi & Taban Baghfalaki

To cite this article: Oludare Ariyo, Kehinde Olobatuyi & Taban Baghfalaki (20 Jan 2025): A Bayesian joint bent-cable model for longitudinal measurements and survival time with heterogeneous random-effects distributions, Journal of Biopharmaceutical Statistics, DOI: 10.1080/10543406.2025.2450321

To link to this article: https://doi.org/10.1080/10543406.2025.2450321

Published online: 20 Jan 2025.

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
https://doi.org/10.1080/10543406.2025.2450321

Taylor & Francis
Taylor & Francis Group

A Bayesian joint bent-cable model for longitudinal measurements and survival time with heterogeneous random-effects distributions

Oludare Ariyo [a], Kehinde Olobatuyi [b], and Taban Baghfalaki [c]

[a]School of Mathematics, Statistics and Actuarial Science, University of Essex, Colchester, UK; [b]Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia, Canada; [c]Department of Mathematics, The University of Manchester, Manchester, UK

ABSTRACT

Biomarkers are measured repeatedly in clinical studies until a pre-defined endpoint, such as death from certain causes, is reached. Such repeated measurements may present a dynamic process for understanding when to expect the study's endpoint. Joint modelling is often employed to handle such a model. Typically, shared random effects are assumed to be common to both the longitudinal component and the study's endpoint. These shared random effects usually assume homogeneous and follow a normal distribution. However, identifying homogeneous subgroups is important when the underlying population is heterogeneous. This issue has received little attention in the literature, particularly for multi-phase longitudinal responses. In this paper, we propose a joint modelling approach for longitudinal and survival models using a bent-cable mixed model for longitudinal measurements and a Weibull distribution for the survival component. We also incorporate a finite mixture of normal distribution assumptions to account for the unobserved heterogeneity in the shared random effects model. A Bayesian MCMC is developed for parameter estimation and inferences. The proposed method is evaluated using simulation studies and the Tehran Lipid and Glucose Study dataset.

ARTICLE HISTORY
Received 7 February 2024
Accepted 2 January 2025

KEYWORDS
Bayesian paradigm; bent-cable mixed model; heterogeneity; joint modelling; mixture distributions

1. Introduction

In many longitudinal clinical studies, repeated biomarker measurements are collected until a pre-defined endpoint occurs, such as death by some causes. In a Tehran, Lipid and Glucose Study (TLGS), for example, patients' biomarkers such as total cholesterol, fasting blood sugar, etc., are collected over time wherein death, by cardiovascular disease (CVD) or diabetes or high blood pressure or cancer, etc., are considered as the study's endpoint. These repeatedly measured response variables represent dynamic processes, which also help us understand when to expect the study's endpoint. As such, a joint modelling of the response variable and time-to-event processes is preferable to handle such data instead of separately modelling the variables, as it may lead to inefficient or biased results (Guo and Carlin 2004; Viviani et al. 2014; Wu et al. 2010).

In the last two decades, studies that proposed several extensions of the joint modelling of longitudinal measurements and survival time have become increasingly popular in the literature (Ariyo and Adeleke 2022; Baghfalaki and Ganjali 2015; Baghfalaki et al. 2014a,c; 2017; Ibrahim et al. 2010; Rappl et al. 2023; Rizopoulos 2012; Wulfsohn and Tsiatis 1997; Xu and Zeger 2001) and Alsefri et al. (2020), Papageorgiou et al. (2019), Sousa (2011), and Zhudenkov et al. (2022) gave methodological reviews of the joint modelling of longitudinal and time-to-event data.

CONTACT Oludare Ariyo oludare.ariyo@essex.ac.uk School of Mathematics, Statistics and Actuarial Science, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK

The standard approach for creating joint models assumes that some shared random effects are common to both the longitudinal component and the study's endpoint. To model the response processes for the longitudinal part, linear mixed models (LMMs) are commonly used (Ariyo et al. 2020, 2022; Baghfalaki et al. 2014b, 2017; Pusponegoro et al. 2017; Verbeke et al. 1997). LMMs can be limited when the longitudinal response exhibits multiphase trajectories over time. For instance, the fasting blood sugar concentration or cholesterol level of patients may decrease over time due to the effect of drugs or by improved healthy lifestyle of the patients up to a certain point and then start to increase again, indicating the development of resistance to treatment over time. These multiphase changes may occur gradually instead of abruptly. The bent-cable model offers a more realistic analysis for multi-phase data, capturing gradual changes over time (Chiu et al. 2006). It provides increased flexibility by fitting both linear trends before and after the transition, as well as the smooth intermediate phase (Kneib 2013). Additionally, it offers enhanced interpretability, allowing researchers to clearly identify the transition period and improve predictive accuracy compared to simpler linear models (Lin and Carroll 2001). Therefore, we consider using the bent-cable model for longitudinal response and a Weibull model for time-to-event analysis, linked through shared random effects.

In joint models that combine longitudinal measurements and time-to-event data, it is common to assume that the shared random effects follow a normal distribution. However, in cases where the underlying population is heterogeneous, it becomes crucial to identify homogeneous subsamples. This issue has received little attention in the literature on joint modelling. To address this, Baghfalaki et al. (2017) used a finite mixture of normal distributions as the random effects distribution in longitudinal mixed-effects models. Additionally, Elashoff et al. (2010) discussed heterogeneous random effects with a parameterization of the normal random effects. Another proposed solution is an alternative parameterization for shared parameter models that evaluates the impact of misspecifying the random effects distribution on parameter estimates, as proposed by Rizopoulos et al. (2008). To our knowledge, this has not been addressed, especially in situations where the longitudinal response has multiphase trajectories.

In this paper, we propose the joint modelling of longitudinal and survival models. We assume a bent-cable mixed model for the longitudinal measurement and Weibull distribution for the survival part using a finite mixture of normal distribution assumptions for the unobserved heterogeneity of the shared random effect model. We use a Bayesian approach to fit the model using the Markov Chain Monte Carlo (MCMC) methodology, implemented using JAGS (Plummer 2012; Plummer et al. 2003) and R2jags package (Su and Yajima 2015) as an interface between R platform and JAGS. We perform an extensive simulation study to evaluate the performance of the proposed models and finally we apply the model to Tehran Lipid and Glucose Study data sets.

The paper is organized as follows: Section 2 provides a detailed description of our proposed joint bent-cable model for longitudinal measurements with a heterogeneous random-effects distribution for random effects. We demonstrate the proposed model using simulation studies in Section 3. In Section 4, we apply the proposed model to Tehran Lipid and Glucose Study. Finally, we present some concluding remarks in Section 5.

2. Bent-cable model

2.1. Joint models and estimation

In this section, we present a joint model for longitudinal data that accounts for heterogeneous distribution of a random effect, where the longitudinal data contains a threshold or change point.

Let y_i represents longitudinal response for subject i monitored over time s_{ij}. Here, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n_i$, where n is the total number of subjects. The true event time, denoted by T_i^*, when the study's end point occurs. The censoring time is represented by C_i. The true event T_i is calculated as the minimum of the true event time T_i^* and the censoring time C_i. This represents the estimated survival time for the i^{th} individual. The censoring

indicator δ_i is given by 1 when there is no censoring and 0 otherwise. Therefore, the observed data for the outcome consists of pairs $\{(T_i, \delta_i), i = 1, 2, \ldots, n\}$. The two processes are jointly modelled using a bent-cable mixed model (BCMM) for $\mathbf{y}_i$ and a Weibull proportional hazard for (T_i, δ_i).

2.2. Time-to-event sub-model

For the first part of the joint model, we use a Weibull proportional hazard model (due to its flexibility in modelling various hazard functions), which may be linked to the bent-cable model through shared random effects. As such,

$$\boldsymbol{T}_i \sim \text{Weibull}(\boldsymbol{w}_i^T\varphi + \Phi^T b_i, r), \tag{1}$$

where φ is the p-dimensional vector containing all parameters of the fixed effects, w_i is also the p-dimensional vector of independent variables, $\mathbf{b}_i$ is the shared parameter between the models, Φ is a q-dimensional vector of association parameters, and r is the scalar shape parameter of the Weibull distribution. The proposed hazard function is expressed as

$$h(t_i|\boldsymbol{w}_i, \Phi, \mathbf{b}_i) = h_0(t_i)\exp\{\mathbf{w}_i^T\varphi + \boldsymbol{\phi}^T\mathbf{b}_i\},$$

where $h_0(t_i)$ is the baseline hazard function and the density function for survival time denoted by $S(t_i)$ can be obtained using the relationship; $S(t_i) = \exp[-H(t_i)]$, where $H(t)$ is the cumulative hazard function given by $\int_0^t h(u)du$, which then leads to the expression of the survival function as;

$$S(t_i) = h(t_i|\mathbf{w}_i, \varphi, \mathbf{b}_i) \times \exp\left\{-H_0(t_i)\exp\left\{\mathbf{w}_i^T\varphi + \boldsymbol{\phi}^T\mathbf{b}_i\right\}\right\},$$

can be expressed as (Dobson and Barnett 2018):

$$S(t_i) = \left[rt_i^{r-1}\exp\left\{\mathbf{w}_i^T\varphi + \boldsymbol{\phi}^T\mathbf{b}_i\right\}\right]^{\delta_i}\exp\left\{-\left\{\exp\mathbf{w}_i^T\varphi + \boldsymbol{\phi}^T\mathbf{b}_i\right\}\right\}.$$

2.3. Longitudinal sub-model

We now describe the second part of the joint model, BCMM, to analyze longitudinal data. y_{ij} is assumed to follow a normal distribution;

$$y_{ij}(s)|\mathbf{b}_i \sim N\left(\mu_{ij}, \sigma_\varepsilon^2\right) \quad i = 1, \ldots, n, \tag{2}$$

where μ_{ij} is the bent-cable mixed-effects model given as follows;

$$\mu_{ij} = \mathbf{x}_{ij}^T\eta + \beta_{0i} + \beta_{1i}s_{ij} + \beta_{2i}q(s_{ij}|\tau_{ij}, \gamma_{ij})$$

where $\mathbf{x}_{ij}$ the vector of time-varying covariates, $q_{ij}(s_{ij}, \tau_{ij}, \gamma_{ij}) = \frac{(s_{ij}-\tau_{ij}+\gamma_{ij})^2}{4\gamma_{ij}}I\{|s_{ij} - \tau_{ij}| \le \gamma_{ij}\} + (s_{ij} - \gamma_{ij})I\{s_{ij} > \gamma_{ij}\}$, $I(\cdot)$ is an indicator function, $\tau_{ij} > 0$ is the half-width of bend, γ_{ij} is the change point. The β_{0ij} and β_{1ij} are the incoming random intercept and slope, respectively, β_{2ij} is the change in slope between the incoming and outgoing linear phases and all the random coefficient parameters in the model as expressed below;

$$\begin{aligned} \beta_{0i} &= \beta_0 + b_{1i}, \quad \beta_{1i} = \beta_1 + b_{2i}, \quad \beta_{2i} = \beta_2 + b_{3i}, \\ \tau_i &= \tau + b_{4i}, \quad \gamma_i = \gamma + b_{5i}. \end{aligned} \tag{3}$$

For convenience's sake, let θ be a vector containing all parameters of the fixed effects of the model such that; $\theta = (\beta_0, \beta_1, \beta_2, \tau_i, \gamma_i, \eta)^T$. Also, $\mathbf{b}$ is the vector of all the random effects (assuming homogeneity)

in the model, which is typically assumed to be normally distributed, i.e., $\mathbf{b}_i = (b_1, b_2, b_3, b_4, b_5)^T$, $\mathbf{b}_i \sim N_5(\mathbf{0}, \boldsymbol{D})$, where $\boldsymbol{D}$ is variance – covariance matrix with a dimension of 5.

Here, we will assume heterogeneity by replacing the normal assumption for $\mathbf{b}_i$ in (1) and (2) by a finite mixture q dimensional normal distributions, i.e

$$\mathbf{b}_i \sim \sum_{k=1}^{g} \pi_k N_q(\mu_k, \boldsymbol{D}),$$

where $\pi_k = Pr(c_i = k)$ is the assignment probability and the total number of the group is given by g $(k = 1, \ldots, g)$ and $q = 5$.

In a Bayesian analysis of finite mixture models, such as the heterogeneity model, parameter estimates by their posterior mean often lead to unreliable results due to the label-switching problem caused by the symmetry in the likelihood of the model parameter. This had been a significant setback to the heterogeneity model in the Bayesian framework and MCMC approach. One of the suggestions for this problem is to eliminate symmetry by introducing some constraints (Stephens 2000). Other authors have provided solutions to the label-switching problem (Jasra et al. 2005; Puolamäki and Kaski 2009; Sperrin et al. 2010; Stephens 2000). To overcome this problem, we adopt the following constraints; (i) $\pi_1 \leq \pi_2 \leq \ldots, \leq \pi_g$, (ii) $\sum_{k=1}^{g} \pi_k \mu_k = 0$, where $\mu_k = (\mu_{k1}, \ldots, \mu_{kq})^T$ (iii) $\sum_{k=1}^{g} \pi_k = 1$, (iv) $\mu_{gj} = \frac{-\sum_{k=1}^{g-1} \pi_k \mu_{kj}}{\pi_g}, j = 1, 2, \ldots, q$. As such, the expectation and the variance of $\mathbf{b}_i$ are $\sum_{k=1}^{g} \pi_k \mu_k$ and $\sum_{k=1}^{g} \mu_k \mu_k^T (1 - \pi_k) + \boldsymbol{D}$ respectively.

2.4. Joint model and bayesian estimation

Under the distribution assumptions and constraints discussed in previous sections, the marginal distribution of $(\mathbf{Y}_i, T_i, \delta_i)$ is given by

$$f(\mathbf{y}_i, t_i|\theta) = \sum_{k=1}^{g} \pi_k \left(\int (\prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \theta_y)) f(t_i, \delta_i|\mathbf{b}_i, \theta_t) \boldsymbol{\phi}(\mathbf{b}_i|\theta_{\mathbf{b}}) \, d\mathbf{b}_i \right),$$

where $\theta_y = (\beta, \sigma_\varepsilon^2)^T$, $\theta_t = (\varphi, \Phi, r)^T$, $\theta_b = (\pi_1, \ldots, \pi_{g-1}, \mu_1, \ldots, \mu_{g-1}, D)^T$, and $\theta = (\theta_y^T, \theta_t^T, \theta_b^T)$. Hence, the likelihood function is given as

$$L(\theta|\boldsymbol{y}, \mathbf{t}, \varphi) = \prod_{i=1}^{n} f(\boldsymbol{y}_i, t_i, \delta_i|\theta).$$

The numerical computation of the likelihood is not straightforward; therefore, a Bayesian approach using MCMC is employed. In Bayesian modelling, the prior distribution for unknown parameters should be defined. We assume independent priors for all components of θ and the prior distributions are given as:

$$\begin{array}{llll}
\beta & \sim N(\nu_\beta, \sum_\beta), & \varphi \sim N(\nu_\varphi, \sum_\varphi), \\
\Phi & \sim N_q(\nu_\Phi, \sum_\Phi), & \mu_j \sim (\nu_{\mu_j}, \sum_{\nu_j}), j = 1, \cdots, g-1, \\
r & \sim \Gamma(a_r, b_r), & \boldsymbol{D} \sim IW(\Phi, \psi), \\
(\pi_1, \ldots, \pi_g) & \sim \text{Con} - \text{Dir}(\alpha_1, \ldots, \alpha_g), & \pi_1 \leq, \ldots, \leq \pi_g, \\
\tau & \sim HN_q(\nu_\tau, s_\tau^2) & \sigma_\varepsilon^2 \sim I\Gamma(a_\sigma, b_\sigma),
\end{array}$$

where $j = 1, 2, \ldots, g-1$, $I\Gamma(\cdot, \cdot)$, $\Gamma(\cdot, \cdot)$, $\text{Con} - \text{Dir}(\cdot, \cdot)$, and $IW(\cdot, \cdot)$,

denote the inverse gamma distribution, the gamma distribution, Constrained-Dirichlet distribution (see Baghfalaki et al. 2017, for more details), and the inverse Wishart distribution, respectively. Assigning these priors is vital because they allow easy implementation in Bayesian software like

BUGS. The MCMC method was used for the posterior sampling via JAGS (Plummer et al. 2003). We present the code for one of the models in https://github.com/OludareAriyo/BentCable-Code.

3. Simulation studies

In the first simulation study, we generated a bent-cable model with a mixture of normal distribution for random effects. We consider the following longitudinal measurement model,

$$y_{ij} = \eta x_{ij} + \beta_{0i} + \beta_{1i} s_{ij} + \beta_{2i} q(s_{ij}|\tau_i, \gamma_i) + \varepsilon_{ij},$$

where $i = 1, 2, \ldots, n$, $j = 1, 2, 3, 4, 5$, and $\varepsilon_{ij} \sim N(0, 1)$. The bent-cable parameter $q_{ij}(s_{ij}, \tau_{ij}, \gamma_{ij}) = \frac{(s_{ij}-\tau_{ij}+\gamma_{ij})^2}{4\gamma_{ij}} I\{|s_{ij} - \tau_{ij}| \leq \gamma_{ij}\} + (s_{ij} - \gamma_{ij}) I\{t_i > \gamma_{ij}\}$. We generated x_{ij} from a uniform distribution $U(-1, 4)$ with $\beta_{0i} = \beta_0 + b_{1i}$, $\beta_{1i} = \beta_1 + b_{2i}$, $\beta_{2i} = \beta_2 + b_{3i}$, $\tau_i = \tau + b_{4i}$ and $\gamma_i = \gamma + b_{5i}$. The random effect $b_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i}, b_{5i}) \sim \sum_{k=1}^{g} \pi_k N_5(\mu_k, \boldsymbol{D})$ where $\boldsymbol{D} = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$. The true values of the parameters are considered as $\eta = 2$, $\beta_0 = 0.5$, $\beta_1 = -1$, $\beta_2 = 0.5$, $\tau = 1.5$, and $\gamma = 1$. For the Weibull proportional hazard model, we considered the following model,

$$h(t_i) = r \times t_i^{r-1} \exp\{\varphi_0 + \varphi_1 x_i + \boldsymbol{\phi}_1 b_{1i} + \boldsymbol{\phi}_2 b_{2i} + \boldsymbol{\phi}_3 b_{3i} + \boldsymbol{\phi}_4 b_{4i} + \boldsymbol{\phi}_5 b_{5i}\},$$

where $\varphi_0 = 1$, $\varphi_1 = -0.5$, $\boldsymbol{\phi}_1 = \boldsymbol{\phi}_2 = \ldots = \boldsymbol{\phi}_5 = 2$, and $r = 2$. We used noninformative prior distributions for the model parameters. Specifically, we assigned a normal distribution $N(0, 100)$ for each component of the population parameter vectors β, $\boldsymbol{\phi}$, and an inverse gamma prior distribution $IG(0.01, 0.01)$ for $\sigma\varepsilon$ and half-normal distributions with large variance for τ and γ. For the vector $(\pi_1, \pi_2, \ldots, \pi_g)$, we used a Constrained-Dirichlet distribution with parameters $(1, 1, \ldots, 1)$. We ran three parallel chains of 20,000 iterations each, discarding the first 10,000 to avoid correlation. We then thinned the chains to a spacing of 90. The chains showed rapid convergence, with all parameters displaying a $\hat{R}$ (Brooks and Gelman 1998) around 1.0 within 10,000 iterations. Furthermore, to evaluate the model's performance under different scenarios, we used two criteria for θ: the relative bias (Rel. Bias) and the root mean square error (RMSE) as:

$$\text{Rel.Bias}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{\theta}_i}{\theta} - 1\right), \quad \text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \theta\right)^2},$$

where $\hat{\theta}$ is the estimate of θ for the i^{th} samples and $N = 1000$.

Under the simulation described above, we considered the following scenarios;

- *Scenario 1*: We generated data from homogeneous normal distributions for random effects with a mean vector of $(0, 0, 0, 0, 0)^T$. The analysis assumes that the random effects follow a heterogeneous distribution with two components (i.e., $g = 2$), resulting in misspecification of the random effects distribution components.
- *Scenario 2*: The data is generated from heterogeneous random effects with two components, with ($\pi_1 = 0.4$ and $\pi_2 = 0.6$). The mean for the first component is $\mu_1 = (2, 2, 2, 2, 2)^T$ and the mean for the second component is $\mu_2 = (-2, -2, -2, -2, -2)^T$. We analyzed the simulated data, assuming (i) homogeneous normality for the random effects (specification) and (ii) heterogeneous normality with two components (correct specification) (i.e. $g = 2$).

- *Scenario 3*: Data generated from heterogeneous random effects with three components, such that $\pi_1 = \pi_2 = 0.3$, $\pi_3 = 0.4$, $\mu_1 = -\mu_2 = (2,2,2,2,2)^T$ and $\mu_3 = (0,0,0,0,0)^T$, and analysis with (i) homogeneous normal (ii) mixture of three components.

Table 1 presents the results of scenario 1, where the data is generated from a bent-cable with homogeneous random effects. The analysis assumes random effects as a mixture of two components. The results indicate that we lost nothing when we analyze the data using the heterogeneous joint model even though the data is coming from homogeneous random effects. This demonstrates that the performance of the heterogeneous joint model is reliable as the relative biases and the RMSE are small. Additionally, the table confirms that increasing the sample size reduces the relative biases and the RMSE.

The results of scenario 2 are presented in Table 2. In this scenario, data was generated using a bent cable with heterogeneous random effects having two components. The data was analyzed via homogeneous joint modelling and a heterogeneous mixture model for random effects. The results show that misspecification of the number of components significantly affects the parameters of the covariance matrix of random effects components as the relative biases and the RMSE are very high. However, the impact of the fixed parameters is negligible. On the other hand, the heterogeneous joint model is reliable when analyzed with the correct number of components. If the component of random effects is incorrectly specified, we lose the parameter's efficiency, especially the covariance parameters. This also confirmed the superior performance of the heterogeneous distribution for random effects.

Table 1. Simulation results for scenario 1 (est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and RMSE) for 1000 samples when the data generated from a bent-cable model with homogeneous random effects and analysis with a heterogeneous random distribution with two components.

		$n = 200$				$n = 500$			
Paratemer	Real values	Est	S.E	Rel. Bias	RSME	Est	S.E	Rel. Bias	RSME
η	2.0000	2.0198	0.0308	0.0010	0.0247	2.0058	0.0099	0.0001	0.0012
β_0	0.5000	0.4604	0.0926	−0.0920	0.1904	0.4999	0.0779	−0.1952	0.0898
β_1	−1.0000	−0.9103	0.0853	−0.1403	0.1483	−1.0765	0.0676	−0.0098	0.0777
β_2	0.5000	0.4620	0.0920	−0.0932	0.2835	0.4875	0.0457	−0.0932	0.2835
φ_0	1.0000	0.9712	0.1647	−0.1062	0.2265	1.0124	0.0899	−0.0985	0.0357
φ_1	−0.5000	−0.5575	0.0070	0.1111	0.5559	−0.5114	0.0057	0.0090	0.5559
γ	1.0000	1.4700	1.7531	0.8085	2.9896	1.2315	1.5679	0.6656	2.0330
τ	1.5000	1.8549	1.0644	0.4699	1.5424	1.6535	0.9878	0.4229	1.4455
r	2.0000	1.5110	0.0006	−0.0945	0.1890	1.7110	0.0004	−0.0945	0.1556
ϕ_1	2.0000	1.6119	0.0354	−0.1059	0.2121	1.7119	0.0211	−0.1321	0.1342
ϕ_2	2.0000	1.3379	0.0372	−0.1169	0.2352	1.6574	0.0246	−0.1679	0.1235
ϕ_3	2.0000	1.8910	0.0317	−0.1051	0.2110	1.9679	0.0213	−0.1347	0.1121
ϕ_4	2.0000	1.8150	0.0419	−0.0925	0.1887	2.0974	0.0219	0.0679	0.0776
ϕ_5	2.0000	2.1013	0.0413	0.0095	0.1927	2.0987	0.0075	0.0089	0.1889
$_{\sigma}2$	1.0000	1.0424	0.0007	0.0933	0.0303	1.0358	0.0006	0.0788	0.0223
λ_1	0.0000	0.4103	0.0068	*	0.0057	0.4065	0.0011	*	0.0022
λ_2	1.0000	0.5897	0.0048	0.5671	0.5935	0.4764	0.2011	0.5678	0.0045
$\mu 11$	1.0000	2.6857	0.3105	1.4359	1.7748	2.0009	0.3001	1.3390	1.1835
$\mu 22$	1.0000	1.9981	0.1197	1.2345	1.0479	1.6790	0.1155	1.2219	1.0123
$\mu 33$	1.0000	2.0009	0.4334	1.0794	1.2090	2.0010	0.4223	1.0668	1.2111
$\mu 44$	1.0000	2.6269	0.0861	1.1360	0.9690	2.0006	0.0768	1.0001	0.8977
$\mu 55$	1.0000	1.2222	0.1414	1.0359	1.1188	1.0123	0.0898	1.0099	1.0089
$\mu 12$	0.5000	0.9559	0.1813	0.4589	0.7685	0.7896	0.0987	0.3445	0.6897
$\mu 13$	0.5000	0.4331	0.3013	0.8769	0.9000	0.5111	0.3340	0.7657	0.7800
$\mu 14$	0.5000	0.4371	0.3213	0.8969	0.8990	0.5011	0.3234	0.7566	0.7580
$\mu 15$	0.5000	1.0924	0.0999	1.3593	0.9999	1.0766	0.0900	1.1136	0.9778
$\mu 23$	0.5000	0.5353	0.1253	0.0079	0.0346	0.5000	0.0999	0.0067	0.0035
$\mu 24$	0.5000	−0.9687	0.1969	0.0568	0.9877	−0.5569	0.1097	0.0437	0.7900
$\mu 25$	0.5000	0.5340	0.0720	0.0343	0.7654	0.5113	0.0665	0.0214	0.7347
$\mu 34$	0.5000	0.9430	0.0869	0.2360	0.8965	0.9212	0.0766	0.2111	0.7690
$\mu 35$	0.5000	0.5104	0.1137	0.0215	0.0433	0.5012	0.0768	0.0115	0.0323
$\mu 45$	0.5000	−0.4229	0.0828	−0.0137	0.1258	0.4789	0.0658	−0.0119	0.1232

Table 2. Simulation results for scenario 2 (est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and RMSE) for 1000 samples when the data generated from a bent-cable model with a mixture of normal distribution random effects and analysis with (i) standard normal effects (one component) and (ii) a mixture of normal distribution with two components when $n = 500.

		$g = 1$ component				$g = 2$ components			
Paratemer	Real values	Est	S.E	Rel. Bias	RSME	Est	S.E	Rel. Bias	RSME
η	2.0000	2.0080	0.6051	0.0004	0.1274	2.0014	0.0019	0.0003	0.0012
β_0	0.5000	0.4617	0.6502	−0.0962	2.1112	0.4792	0.0087	−0.0960	0.0197
β_1	−1.0000	2.2145	0.6186	−0.3365	1.8492	−0.9864	0.0088	−0.1864	0.0193
β_2	0.5000	0.3970	0.6513	−0.0908	8.5644	0.1277	0.0091	−0.0957	0.0291
φ_0	1.0000	−1.7555	0.6038	−0.1966	1.8492	−1.0290	0.0176	−0.1514	0.0318
φ_1	−0.5000	−0.5955	0.5946	0.1272	7.1605	−0.5275	0.0009	0.0215	0.0608
γ	1.0000	1.6121	2.1198	1.4357	1.8492	0.8173	0.2034	−0.1987	0.3082
τ	1.5000	1.8967	1.2764	0.7865	1.7890	1.4682	1.0789	0.5789	1.4567
r	2.0000	1.6214	0.5872	−0.0935	3.2900	1.7214	0.0001	−0.0893	0.0179
ϕ_1	2.0000	−0.1747	0.6242	−0.1138	1.8492	−0.1736	0.0037	−0.1087	0.0218
ϕ_2	2.0000	−0.4716	0.6296	−0.1294	6.4808	−0.4487	0.0043	−0.1224	0.0248
ϕ_3	2.0000	−0.1268	0.6189	−0.1113	1.8445	−0.1352	0.0036	−0.1068	0.0215
ϕ_4	2.0000	0.0809	0.6312	−0.1005	1.6429	0.0464	0.0045	−0.0977	0.0198
ϕ_5	2.0000	0.0531	0.6299	−0.1019	1.9524	0.0331	0.0044	−0.0983	0.0198
$_\sigma 2$	1.0000	0.1154	0.5878	−0.0926	1.9442	0.1168	0.0002	−0.0883	0.0078
λ_1	0.4000	–	–	–	–	0.4065	0.0011	0.0320	0.0022
λ_2	0.6000	–	–	–	–	0.5935	0.0011	0.0399	0.0023
$\mu 11$	1.0000	2.1391	0.7543	2.0323	183.0594	2.0770	0.0165	1.3390	1.1835
$\mu 22$	1.0000	1.0150	0.6449	1.2292	108.0825	1.0140	0.0059	1.2219	1.0123
$\mu 33$	1.0000	3.0704	0.7716	2.6789	124.6949	2.4371	0.0165	1.0668	1.2111
$\mu 44$	1.0000	0.7828	0.6648	−0.8930	99.9406	0.7634	0.0074	1.0001	0.8977
$\mu 55$	1.0000	0.8632	0.6765	−0.8447	115.3954	0.8918	0.0088	1.0099	1.0089
$\mu 12$	0.5000	−1.2165	0.6816	2.8045	79.2605	−1.1891	0.0095	0.3445	0.6897
$\mu 13$	0.5000	1.5256	0.7406	1.1802	92.8248	1.2873	0.0131	0.7657	0.7800
$\mu 14$	0.5000	−0.1290	0.6640	−2.2308	103.1309	0.1454	0.0089	0.0068	0.0035
$\mu 15$	0.5000	0.8751	0.6690	0.0822	3.5667	−0.1784	0.0077	1.1136	0.9778
$\mu 23$	0.5000	−1.1138	0.6775	0.5942	101.8692	0.1441	0.0084	0.0067	0.0035
$\mu 24$	0.5000	0.0918	0.6339	0.3587	78.9492	−0.9572	0.0081	0.0437	0.7900
$\mu 25$	0.5000	−0.0347	0.6367	2.4707	92.4628	0.1164	0.0049	0.0214	0.7347
$\mu 34$	0.5000	−0.1995	0.6680	0.2246	4.4650	−0.0668	0.0054	0.2111	0.7690
$\mu 35$	0.5000	0.2011	0.6749	−0.1431	12.9744	−0.1569	0.0071	0.0115	0.0323
$\mu 45$	0.5000	−0.3053	0.6519	−2.1802	92.8248	0.1773	0.0079	−0.0119	0.1232

Table 3 presents the results of the significant impact on estimating covariance parameters and fixed parameters when data is generated from a mixture of a normal distribution with three components, but analyzed with a bent-cable with standard normal distribution and with a mixture of normal distributional assumption for the random effects with the correct number of components. Overall, the diverse joint model is reliable even when the random components are incorrectly specified.

4. Application: Tehran Lipid and Glucose Study

4.1. Data description

The data analyzed in this paper was obtained from the Tehran Lipid and Glucose Study (TLGS), which is a population-based cohort study that was initiated in 1999–2001. It consisted of 15,010 residents (aged 3 years and older) in 13 districts of Tehran, Iran. The participants were selected through a multi-stage stratified cluster random sampling (Tohidi et al. 2009). The present study received approval from the Ethics Committee of the Research Institute for Endocrine Sciences in Iran, and informed consent was obtained from all participants prior to their involvement in the study. For other analysis of this data, refer to Baghfalaki et al. (2020); Pahlavanzade et al. (2019)

For the purpose of this study, 4,440 participants aged 41 and above were examined at baseline between 1999 and 2001. For this group, lipid markers and several covariates (including BMI, gender, age, cholesterol, and fasting blood sugar, FBS) were measured every 3 years

Table 3. Simulation results for scenario 3 (est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and RMSE) for 1000 samples when the data generated from a bent-cable model with a mixture of normal distribution random effects with three components and analysis with (i) standard normal effects (one component) and (ii) a mixture of a normal distribution with three components when $n = 500$.

		$g = 1$ component				$g = 3$ components			
Paratemer	Real values	Est	S.E	Rel. Bias	RSME	Est	S.E	Rel. Bias	RSME
η	2.0000	2.0080	0.0101	0.0042	0.0038	1.9989	0.0017	0.0000	0.0013
β_0	0.5000	0.4617	0.0642	−0.9728	−0.8665	0.4983	0.0091	−0.0095	0.0193
β_1	−1.0000	2.2145	0.0611	−3.4022	−3.0302	−0.9636	0.0084	−0.0082	0.0169
β_2	0.5000	0.3970	0.0643	−0.9183	−0.8179	0.5406	0.0091	−0.0099	0.0298
φ_0	1.0000	−1.7555	0.0596	−1.9874	−1.7701	0.8772	0.0200	−0.0088	0.0189
φ_1	−0.5000	−0.5955	0.0587	1.2861	1.1454	−0.4631	0.0009	−0.0126	0.0632
γ	1.0000	1.6121	0.2094	14.5147	12.9276	1.1093	0.2900	0.2934	0.4575
τ	1.5000	2.4995	1.8717	9.9855	8.8937	1.4995	−0.1726	0.0852	0.2118
r	2.0000	1.6214	0.0580	−0.9450	−0.8417	1.7863	0.0001	−0.0087	0.0174
ϕ_1	2.0000	−0.1747	0.0617	−1.1509	−1.0250	1.8719	0.0037	−0.0094	0.0156
ϕ_2	2.0000	−0.4716	0.0622	−1.3080	−1.1650	1.6192	0.0043	−0.0132	0.0134
ϕ_3	2.0000	−0.1268	0.0611	−1.1255	−1.0024	1.7892	0.0039	−0.0168	0.0124
ϕ_4	2.0000	0.0809	0.0623	−1.0156	−0.9045	2.0037	0.0042	0.0135	0.0112
ϕ_5	2.0000	0.0531	0.0622	−1.0303	−0.9176	2.0105	0.0039	0.0068	0.0078
$_\sigma 2$	1.0000	0.1154	0.0581	−0.9363	−0.8339	1.1510	0.0003	0.0009	0.0189
λ_1	0.3000	–	–	–	–	0.2132	0.0009	0.0079	0.0022
λ_2	0.3000	–	–	–	–	0.2116	0.0013	0.0097	0.0005
λ_3	0.4000	-	-	-	-	0.6989	0.0020	0.2032	1.8125
$\mu 11$	1.0000	2.1391	0.7469	31.6548	28.1936	1.2066	0.0394	0.1229	1.0701
$\mu 22$	1.0000	2.5556	0.6386	23.5361	20.9627	0.7152	0.0041	0.2679	1.2346
$\mu 33$	1.0000	3.0704	0.7640	38.1925	34.0165	1.1383	0.0085	−0.0893	0.9895
$\mu 44$	1.0000	1.7828	0.6583	2.0798	1.8524	0.6218	0.0064	−0.0845	1.1425
$\mu 55$	1.0000	1.8632	0.6698	2.5685	2.2877	0.7101	0.0065	0.2805	0.7848
$\mu 12$	0.5000	−1.2165	0.6748	39.4623	35.1475	0.7156	0.0059	0.1180	0.9191
$\mu 13$	0.5000	1.5256	0.7333	23.0408	20.5215	0.6042	0.0068	−0.2231	1.0211
$\mu 14$	0.5000	1.1290	0.6574	−11.4453	−10.1938	−0.5497	0.0048	0.0082	0.0353
$\mu 15$	0.5000	1.8751	0.6624	11.9401	10.6345	1.0678	0.5670	0.0594	1.0086
$\mu 23$	0.5000	2.1138	0.6708	17.1154	15.2440	−0.5347	0.0047	0.0359	0.7817
$\mu 24$	0.5000	0.9178	0.6277	14.7355	13.1243	0.0344	0.0033	0.2471	0.9155
$\mu 25$	0.5000	0.6780	0.6304	36.0869	32.1411	−0.5003	0.0036	0.0225	0.0442
$\mu 34$	0.5000	0.8999	0.6614	13.3791	11.9163	−0.1466	0.0049	−0.0143	0.1285
$\mu 35$	0.5000	0.9201	0.6682	9.6617	8.6053	0.4436	0.0051	−0.2180	0.9191
$\mu 45$	0.5000	0.3053	0.6455	−10.9338	−9.7383	−0.4568	0.0051	0.2032	1.8125

until 2015, resulting in a median follow-up of 12.4 years. Therefore, the number of visits is at most five. The first phase is from 1999 to 2001, the second phase is from 2001 to 2005, the third phase is from 2005 to 2008, the fourth phase is from 2008 to 2011, and the fifth phase is from 2011 to 2015.

The purpose of our study is to identify significant factors that contribute to the risk of death from cardiovascular disease (CVD). We consider systolic blood pressure (SBP) as a longitudinal response. Figure 1 shows the evolution of SBP for randomly selected subjects, separated by gender.

Out of 4440 patients, 229 patients died of CVD, while the remaining patients were considered as censoring data. Figure 2 presents the Kaplan-Meier survival curve estimates separated by gender. Table 4 presents a summary of the longitudinal response and the explanatory variables including body mass index (BMI), total cholesterol (TC), fasting blood sugar (FBS), gender, and age.

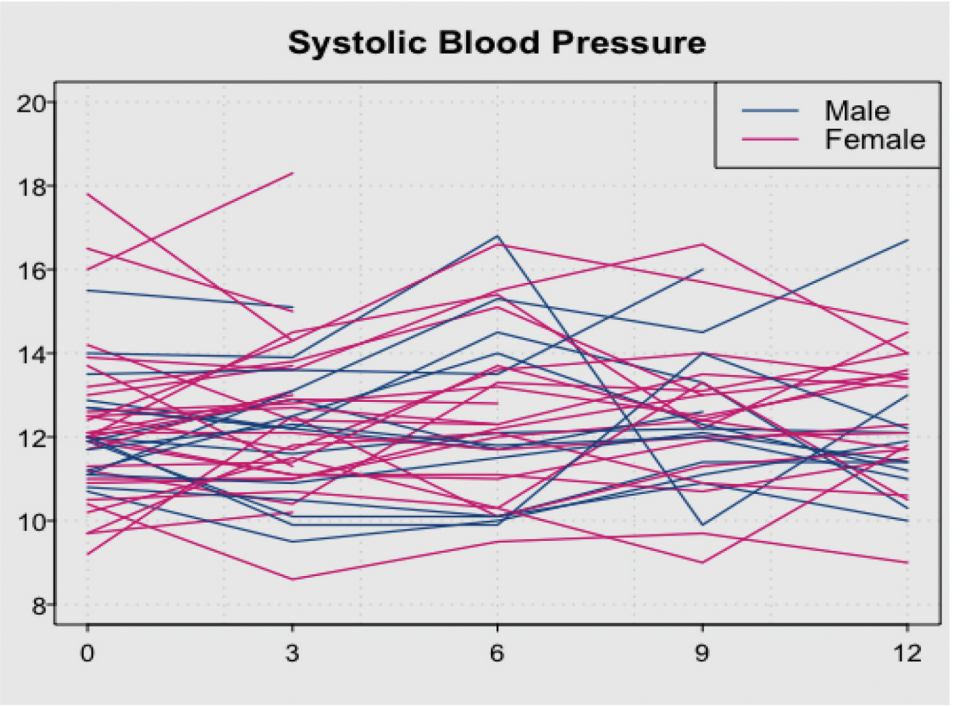

Figure 1. The profile plot for SBP over time for a randomly selected subjects, separated by gender, in the TLGS data.

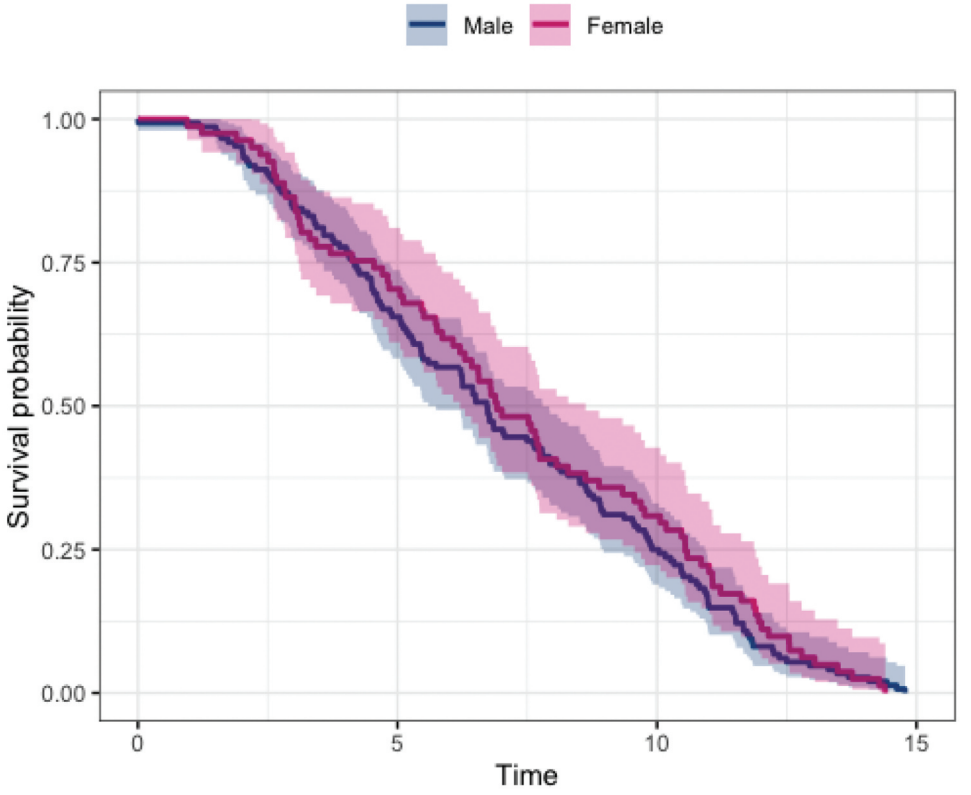

Figure 2. Kaplan-Meier plot of the survival data for all patients considering the gender, in the TLGS data.

Table 4. Baseline and follow-up characteristics of participants at five phases (baseline +4 follow-ups). Summarizations for TLGS data are shown as mean, median, and SD for continuous variables or frequency (percentage) for categorical variables.

		1st phase (1999–2001)	2nd phase (2002–2005)	3rd phase (2005—2008)	4th phase (2008–2011)	5th phase (2011–2015)
SBP	Mean	12.8423	12.7573	12.7860	12.9951	13.1169
	Median	12.5000	12.4000	12.5000	12.8000	12.9000
	SD	2.1224	2.1234	2.1543	2.1830	2.2133
BMI	Mean	27.9361	28.3055	28.4689	28.7699	28.7503
	Median	27.6332	27.9431	28.0102	28.3039	28.3268
	SD	4.5074	4.6562	4.7612	5.0040	5.1084
TC	Mean	226.6007	208.8331	205.4705	201.2286	197.5017
	Median	223.0000	206.0000	203.0000	199.0000	196.0000
	SD	47.3688	42.0178	40.1508	43.5756	41.7350
FBS	Mean	107.4236	109.2737	108.4634	113.9975	112.7628
	Median	94.0000	95.0000	95.0000	100.0000	101.0000
	SD	42.2859	42.1472	40.9945	41.3297	38.6007
Gender	Male	1974(44.45)				
	Female	2466(55.54)				
Age	Mean	55.379				
	Median	54.068				
	SD	9.609				

4.2. Specifying and fitting the model

In this study, we are interested in estimating longitudinal and survival data using a joint model. This model takes into consideration the gradual changes that occur over time, which may have been overlooked but can be captured by the bent-cable model.

$$y_{ij} = \eta_1 \text{Gender}_i + \eta_2 \text{SAge}_i + \eta_3 \text{BMI}_{ij} + \eta_4 \text{TC}_{ij} + \eta_5 \text{FBS}_{ij}$$

$$+ \beta_{0i} + \beta_{1i} s_{ij} + \beta_{2i} q(s_{ij} | \tau_i, \gamma_i)$$

where "SAge" represents the standardized values of age at baseline, $\beta_{0i}, \beta_{1i}, \beta_{2i}$, τ_i, and γ_i are defined in Equation (2). The random effects $\mathbf{b}_i = (b_{1i}, \ldots, b_{5i})$ represent individual variations in the bent-cable model and are shared parameters between the response and survival models.

For the time event model, we used a Weibull proportional hazard model $T_i \sim \text{Weibull}(\mu_i^*, r)$, where $\mu_i^* = \varphi_0 + \varphi_1 \text{Gender}_i + \varphi_2 \text{SAge}_i + \boldsymbol{\phi}_1 b_{1i} + \boldsymbol{\phi}_2 b_{2i} + \boldsymbol{\phi}_3 b_{3i} + \boldsymbol{\phi}_4 b_{4i} + \boldsymbol{\phi}_5 b_{5i}$.

For the Bayesian implementation of the model, we need to evaluate the hyperparameter values in the prior distribution. As such, the prior distributions for a fixed parameter are taken as $N(0, 1000)$ for each component of $(\beta_{0i}, \beta_{1i}, \beta_{2i})$ and for the center parameter τ and width parameter γ of the bent-cable, half-normal distributions $HN(0, 1000)$ and $HN(0, 1000)$ are used, respectively. For the scale parameter σ_ε^2, we used a non-informative inverse gamma distribution $IG(0.01, 0.01)$. Additionally, $\varphi_k, k = 0, 1, 2$ and $\boldsymbol{\phi}_k, k = 1, \cdots, 5$ are drawn from a normal distribution with mean 0 and variance 100. The matrix $\boldsymbol{D}$ follows an inverse Wishart distribution $IW(100 I_5, 5)$, and $(\pi_1, \pi_2, \ldots, \pi_g) \sim \text{constrained} - \text{Dirichlet}(1, 1, 1, \ldots)$ with $\pi_1 < \ldots < \pi_g$.

We conducted two parallel MCMC chains with various starting points, each running for 150,000 iterations, to account for the complex nature of the bent-cable model. After that, we retrained half of the iterations for the posterior analysis and discarded the first half as pre-convergence burn-in. A longer chain is necessary for convergence as the number of components increases. After thinning by 90, the autocorrelation was minimal, indicating effective mixing. The posterior standard deviation values for the parameters were less than 5%, and the Gelman-Rubin potential-scale reduction factors are close to 1.1, indicating good precision and convergence. The computational burden of running MCMC for a model with three components (g = 3), for example, was close to 11 h on an 11th Gen Intel(R) Core(TM) i5-1145G7 @ 2.60 GHz.

Table 5. The estimated parameters from the joint model on TLGS data. Est: posterior mean, SD: standard deviation, 2.5% CI: lower bound of credible interval and 97.5% CI: upper bound of credible interval.

	Est	SD	2.5%	97.5%	Est	SD	2.5%	97.5%	Est	SD	2.5%	97.5%
η_1	−0.007	0.033	−0.069	0.060	−0.109	0.087	−0.270	0.063	−0.138	0.088	−0.302	0.029
η_2	0.074	0.029	0.020	0.127	0.489	0.064	0.359	0.621	0.460	0.065	0.328	0.571
η_3	0.285	0.010	0.266	0.303	0.281	0.009	0.263	0.296	0.293	0.009	0.275	0.310
η_4	0.016	0.001	0.014	0.019	0.012	0.001	0.011	0.014	0.014	0.001	0.012	0.016
η_5	0.006	0.001	0.003	0.008	0.005	0.001	0.002	0.007	0.005	0.001	0.002	0.007
β_1	0.035	0.031	−0.022	0.095	0.355	0.096	0.193	0.557	0.399	0.105	0.184	0.594
β_2	0.046	0.019	0.014	0.089	0.019	0.041	−0.058	0.098	0.050	0.019	0.011	0.083
β_3	0.060	0.033	0.006	0.129	0.243	0.071	0.118	0.380	−0.011	0.091	−0.175	0.154
φ_1	−0.048	0.030	−0.099	0.018	−0.312	0.090	−0.465	−0.110	−0.317	0.094	−0.524	−0.134
φ_2	−0.014	0.035	−0.079	0.053	−0.101	0.097	−0.287	0.086	−0.109	0.081	−0.262	0.078
φ_3	−0.025	0.031	−0.086	0.033	−0.131	0.076	−0.264	0.021	−0.126	0.088	−0.311	0.044
τ	9.988	0.031	9.930	10.046	10.038	0.103	9.832	10.213	5.036	0.097	4.858	5.229
γ	5.052	0.029	5.008	5.111	5.032	0.088	4.865	5.179	10.014	0.113	9.786	10.202
r	0.407	0.043	0.319	0.486	0.557	0.057	0.443	0.667	0.555	0.067	0.419	0.676
ϕ_1	−0.019	0.029	−0.073	0.028	−0.104	0.060	−0.215	0.002	−0.095	0.057	−0.206	0.009
ϕ_2	0.001	0.030	−0.060	0.057	−0.031	0.107	−0.227	0.200	−0.002	0.094	−0.181	0.165
ϕ_3	−0.001	0.031	−0.059	0.065	−0.098	0.098	−0.291	0.085	0.011	0.092	−0.160	0.184
ϕ_4	−0.002	0.029	−0.047	0.054	−0.046	0.080	−0.210	0.140	−0.007	0.087	−0.184	0.156
ϕ_5	0.004	0.029	−0.052	0.058	0.003	0.088	−0.152	0.187	−0.003	0.089	−0.153	0.182
d_{11}	4.439	0.454	3.634	5.216	4.340	0.540	3.299	5.505	2.581	0.293	2.079	3.197
d_{21}	−0.229	0.070	−0.363	−0.100	−0.232	0.051	−0.338	−0.139	−0.089	0.024	−0.132	−0.048
d_{31}	−1.053	0.258	−1.637	−0.628	−0.873	0.257	−1.500	−0.464	−0.226	0.686	−1.473	1.086
d_{41}	−0.938	0.493	−1.912	0.009	−1.517	0.656	−3.126	−0.415	−0.320	0.879	−2.098	1.040
d_{51}	0.455	0.304	−0.101	1.062	0.340	0.546	−0.705	1.331	0.434	0.883	−0.831	2.583
d_{22}	0.307	0.021	0.270	0.347	0.070	0.007	0.058	0.083	0.058	0.005	0.050	0.067
d_{32}	0.099	0.035	0.038	0.185	0.069	0.021	0.033	0.116	0.006	0.032	−0.054	0.070
d_{42}	0.058	0.070	−0.075	0.190	0.119	0.052	0.037	0.244	0.012	0.039	−0.057	0.104
d_{52}	−0.047	0.042	−0.121	0.025	−0.041	0.030	−0.096	0.014	−0.011	0.038	−0.103	0.049
d_{33}	1.287	0.180	0.993	1.742	0.508	0.122	0.307	0.756	1.774	0.689	0.855	3.457
d_{43}	0.497	0.191	0.194	0.856	0.430	0.165	0.153	0.775	0.404	0.697	−0.540	2.031
d_{53}	−0.186	0.122	−0.437	0.018	−0.036	0.115	−0.231	0.188	−0.495	0.725	−2.240	0.458
d_{44}	4.163	0.877	2.898	5.973	2.189	0.807	1.039	3.941	2.135	1.185	0.697	5.342
d_{54}	0.747	0.367	0.022	1.428	−0.128	0.258	−0.678	0.339	−1.272	1.219	−4.412	−0.003
d_{55}	1.980	0.284	1.545	2.620	0.929	0.254	0.513	1.422	2.124	1.396	0.760	5.450
σ^2_ε	1.364	0.059	1.260	1.474	1.471	0.074	1.337	1.637	1.656	0.064	1.534	1.754
π_1	–	–	–	–	0.354	3.000	−4.657	4.479	0.334	0.037	0.287	0.427
π_2	–	–	–	–	0.646	3.000	−3.479	5.657	0.618	0.030	0.535	0.652
π_3	–	–	–	–	–	–	–	–	0.048	0.019	0.021	0.089
μ_{11}	–	–	–	–	0.396	0.098	0.223	0.584	0.134	0.094	−0.042	0.301
μ_{12}	–	–	–	–	0.093	0.039	0.023	0.163	−0.015	0.035	−0.086	0.053
μ_{13}	–	–	–	–	0.292	0.076	0.160	0.449	0.001	0.099	−0.162	0.187
μ_{14}	–	–	–	–	0.055	0.101	−0.122	0.235	−0.002	0.108	−0.216	0.186
μ_{15}	–	–	–	–	0.033	0.104	−0.162	0.243	0.002	0.099	−0.160	0.179
μ_{21}	–	–	–	–	−0.377	3.098	−1.818	1.741	0.376	0.092	0.207	0.527
μ_{22}	–	–	–	–	−0.091	0.700	−0.425	0.371	−0.022	0.007	−0.036	−0.009
μ_{23}	–	–	–	–	−0.233	1.813	−1.390	1.364	−0.013	0.117	−0.237	0.199
μ_{24}	–	–	–	–	−0.011	0.387	−0.471	0.276	0.053	0.078	−0.080	0.217
μ_{25}	–	–	–	–	0.047	0.443	−0.279	0.697	0.051	0.095	−0.129	0.221
μ_{31}	–	–	–	–	–	–	–	–	5.096	0.996	3.403	6.984
μ_{32}	–	–	–	–	–	–	–	–	−0.311	0.128	−0.612	−0.096
μ_{33}	–	–	–	–	–	–	–	–	−0.150	2.052	−4.734	3.926
μ_{34}	–	–	–	–	–	–	–	–	0.887	1.433	−1.076	4.589
μ_{35}	–	–	–	–	–	–	–	–	0.881	1.676	−2.120	4.488
DIC			10370.5				10071.4				9833.0	

The results for one, two, and three clusters are presented in Table 5. Based on the Deviance Information Criterion (DIC) values, the model with three clusters is the best fit. We should note that when running the joint model with more than three clusters, the values of DIC remain unchanged. Therefore, we have decided to keep the results with at most three clusters to save space. Based on these results, although none of the covariates are effective on time to event, age, BMI, total cholesterol, and fasting blood sugar are significant variables with a positive effect on SBP. To ensure that the data

"speak for itself," we conducted a sensitivity analysis to examine the impact of prior choice on the posterior distributions. For example, we utilized a uniform prior distribution $(unif(0, 100))$ for σ_ε^2 (Gelman et al. 2013) instead of $IG(0.01, 0.01)$, and for the precision matrix of the random effect $\mathbf{b}_i$ of dimension 5, we used $IW(100I_5, 7)$ with 7 degrees of freedom instead of 5. We observed that the change in prior distributions had an insignificant effect on the posterior mean of the parameters' distributions. As a result, the final findings are reasonable and robust, and the conclusion of our analysis remains unchanged despite changes in prior distributions and initial values.

5. Discussion

Longitudinal clinical studies collect biomarkers until a specific endpoint is reached. Joint modelling of response variables and time-to-event processes is preferable, and researchers have been actively studying the extensions of joint modelling (Alsefri et al. 2020; Ariyo and Adeleke 2022; Baghfalaki and Ganjali 2015; Baghfalaki et al. 2014a,c; 2017; Ibrahim et al. 2010; Papageorgiou et al. 2019; Rappl et al. 2023; Rizopoulos 2012; Wulfsohn and Tsiatis 1997; Zhudenkov et al. 2022). However, most studies have focused on survival data when multiphase trajectories were not identified. Only a few studies have used the bent-cable model when the longitudinal response exhibits multiphase trajectories (Dagne 2018a, 2018b). For those studies, a shared random assumption is made, assuming homogeneity and following a normal distribution. However, it is important to identify homogeneous subgroups when the underlying population is heterogeneous. This issue has received little attention in the literature, particularly for multi-phase longitudinal responses.

This paper proposes a bent-cable mixed model for the longitudinal measurement and a Weibull distribution for the survival part. A finite mixture of normal distribution assumptions accounts for the unobserved heterogeneity of the shared random effect model. The model uses a Bayesian approach with the Markov Chain Monte Carlo (MCMC) methodology, using the JAGS and R2jags package as an interface between the R platform and JAGS. An extensive simulation study is performed to evaluate the performance of the proposed models. In this study, we generated a bent-cable model with a mixture of normal distributions for random effects, varying the number of components. Finally, the model is applied to the Tehran Lipid and Glucose Study data sets.

The results show that the heterogeneous joint model is reliable whether the response exhibits multiphase trajectories or not. Also, misspecification of the number of components significantly affects the parameters of the covariance matrix of random effects components, and the impact of the fixed parameters is negligible. The heterogeneous joint model performs well when analyzed with the correct component, but if the component of random effects is incorrectly specified, the parameter's efficiency is lost especially for the covariance parameters. The study confirms the superior performance of the heterogeneous distribution for random effects in all scenarios considered in this study.

In Bayesian analysis, it is crucial to perform sensitivity analysis to determine if the posterior estimates change significantly when the priors vary. To this end, we conducted a sensitivity analysis using different sets of values for the hyper-parameters in the models and ran the MCMC sampling scheme. We found that the conclusion was similar to those presented in the article. However, the study has several limitations, even though the proposed model fits the Tehran Lipid and Glucose Study well. This article assumes that the shared parameter random effects follow a normal distribution. Nevertheless, the shared parameter model could assume other distributions. Another useful extension of this work is examining multivariate longitudinal time-to-event, which is currently under examination. In the simulation study, we evaluated the performance of the proposed method under the assumption of a fixed shape parameter $r = 2$, representing scenarios with increasing hazard rates. This approach facilitated consistency and computational feasibility but introduced a limitation by restricting the exploration of varying hazard dynamics. Future research could address this limitation by incorporating a broader range of shape parameters, enabling the assessment of the method's robustness under diverse hazard scenarios. Despite this limitation, our findings demonstrate the method's effectiveness and provide valuable insights into its applicability within the defined scope. These results lay a strong foundation for further exploration and refinement in subsequent studies.

Acknowledgements

The authors would like to express their gratitude to the Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran, for their contributions to providing a portion of the data from the Tehran Lipid and Glucose Study. The authors also appreciate the insightful comments from the anonymous referees and the Associate Editor, which significantly enhanced the quality of this work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research presented in this study was conducted on the High-Performance Computing Cluster supported by the University Essex, United Kingdom.

ORCID

Oludare Ariyo http://orcid.org/0000-0003-3375-1831
Kehinde Olobatuyi http://orcid.org/0000-0002-4635-7895
Taban Baghfalaki http://orcid.org/0000-0002-2100-4532

References

Alsefri, M., M. Sudell, M. Garca-Fiñana, and R. Kolamunnage-Dona. 2020. Bayesian joint modelling of longitudinal and time to event data: A methodological review. *BMC Medical Research Methodology* 20 (1):1–17.

Ariyo, O., and M. Adeleke. 2022. Simultaneous Bayesian modelling of skew-normal longitudinal measurements with non-ignorable dropout. *Computational Statistics* 37 (1):303–325. doi: 10.1007/s00180-021-01118-y.

Ariyo, O., E. Lesaffre, G. Verbeke, and A. Quintero. 2022. Bayesian model selection for longitudinal count data. *Sankhya B* 84 (2):516–547. doi: 10.1007/s13571-021-00268-9.

Ariyo, O., A. Quintero, J. Muñoz, G. Verbeke, and E. Lesaffre. 2020. Bayesian model selection in linear mixed models for longitudinal data. *Journal of Applied Statistics* 47 (5):890–913. doi: 10.1080/02664763.2019.1657814.

Baghfalaki, T., and M. Ganjali. 2015. A Bayesian approach for joint modeling of skew-normal longitudinal measurements and time to event data. *REVSTAT-Statistical Journal* 13 (2):169–191.

Baghfalaki, T., M. Ganjali, and D. Berridge. 2014a. Joint modeling of multivariate longitudinal mixed measurements and time to event data using a Bayesian approach. *Journal of Applied Statistics* 41 (9):1934–1955. doi: 10.1080/02664763.2014.898132.

Baghfalaki, T., M. Ganjali, and D. Berridge. 2014b. Joint modeling of multivariate longitudinal mixed measurements and time to event data using a Bayesian approach. *Journal of Applied Statistics* 41 (9):1934–1955. doi: 10.1080/02664763.2014.898132.

Baghfalaki, T., M. Ganjali, and R. Hashemi. 2014c. Bayesian joint modeling of longitudinal measurements and time-to-event data using robust distributions. *Journal of Biopharmaceutical Statistics* 24 (4):834–855. doi: 10.1080/10543406.2014.903657.

Baghfalaki, T., M. Ganjali, and G. Verbeke. 2017. A shared parameter model of longitudinal measurements and survival time with heterogeneous random-effects distribution. *Journal of Applied Statistics* 44 (15):2813–2836. doi: 10.1080/02664763.2016.1266309.

Baghfalaki, T., S. Kalantari, M. Ganjali, F. Hadaegh, and B. Pahlavanzadeh. 2020. Bayesian joint modeling of ordinal longitudinal measurements and competing risks survival data for analysing Tehran lipid and glucose study. *Journal of Biopharmaceutical Statistics* 30 (4):689–703. doi: 10.1080/10543406.2020.1730876.

Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7 (4):434–455. doi: 10.1080/10618600.1998.10474787.

Chiu, G., R. Lockhart, and R. Routledge. 2006. Bent-cable regression theory and applications. *Journal of the American Statistical Association* 101 (474):542–553. doi: 10.1198/016214505000001177.

Dagne, G. A. 2018a. Bayesian two-part bent-cable Tobit models with skew distributions: Application to AIDS studies. *Statistical Methods in Medical Research* 27 (12):3696–3708. doi: 10.1177/0962280217710679.

Dagne, G. A. 2018b. Heterogeneous growth bent-cable models for time-to-event and longitudinal data: Application to AIDS studies. *Journal of Biopharmaceutical Statistics* 28 (6):1216–1230. doi: 10.1080/10543406.2018.1489407.

Dobson, A. J., and A. G. Barnett. 2018. *An introduction to generalised linear models*. New York: Chapman and Hall/CRC.

Elashoff, R. M., X. Huang, and G. Li. 2010. A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Statistics and Its Interface* 3 (2):185–195. doi: 10.4310/SII.2010.v3.n2.a6.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian data analysis*. New York: CRC press.

Guo, X., and B. P. Carlin. 2004. Separate and joint modeling of longitudinal and event time data using standard computer packages. *American Statistician* 58 (1):16–24. doi: 10.1198/0003130042854.

Ibrahim, J. G., H. Chu, and L. M. Chen. 2010. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology* 28 (16):2796. doi: 10.1200/JCO.2009.25.0654.

Jasra, A., C. C. Holmes, and D. A. Stephens. 2005. Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science* 20 (1):50–67. doi: 10.1214/088342305000000016.

Kneib, T. 2013. Beyond mean regression. *Statistical Modelling* 13 (4):275–303. doi: 10.1177/1471082X13494159.

Lin, X., and R. J. Carroll. 2001. Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association* 96 (455):1045–1056. doi: 10.1198/016214501753208708.

Pahlavanzade, B., F. Zayeri, T. Baghfalaki, O. Mozafari, D. Khalili, F. Azizi, and A. Abadi. 2019. Association of lipid markers with coronary heart disease and stroke mortality: A 15-year follow-up study. *Iranian Journal of Basic Medical Sciences* 22 (11):1325. doi: 10.22038/ijbms.2019.35617.8775.

Papageorgiou, G., K. Mauff, A. Tomer, and D. Rizopoulos. 2019. An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual Review of Statistics and Its Application* 6 (1):223–240.

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

Plummer, M. 2012. JAGS version 3.3. 0 user manual [Computer software manual]. International agency for research on cancer. Lyon: *France*. https://sourceforge.net/projects/mcmc-jags/files/Manuals/3.x/jags_user_manual.pdf/download.

Plummer, M., et al. 2003b. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*, vol. 124, 1–10, Vienna, Austria.

Puolamäki, K., and S. Kaski. 2009. Bayesian solutions to the label switching problem. In International symposium on intelligent data analysis, 381–392. Springer.

Pusponegoro, N. H., K. A. Notodiputro, B. Sartono, and B. Sartono. 2017. Linear mixed model for analyzing longitudinal data: A simulation study of children growth differences. *Procedia Computer Science* 116:284–291. doi: 10.1016/j.procs.2017.10.071 .

Rappl, A., T. Kneib, S. Lang, and E. Bergherr. 2023. Spatial joint models through bayesian structured piece-wise additive joint modelling for longitudinal and time-to-event data. *Statistics and Computing* 33 (6). doi: 10.1007/s11222-023-10293-5.

Rizopoulos, D. 2012. *Joint models for longitudinal and time-to-event data: With applications in R*. New York: CRC press.

Rizopoulos, D., G. Verbeke, and G. Molenberghs. 2008. Shared parameter models under random effects misspecification. *Biometrika* 95 (1):63–74. doi: 10.1093/biomet/asm087.

Sousa, I. 2011. A review on joint modelling of longitudinal measurements and time-to-even. *REVSTAT-Statistical Journal* 9 (1):57–81.

Sperrin, M., T. Jaki, and E. Wit. 2010. Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models. *Statistics and Computing* 20 (3):357–366. doi: 10.1007/s11222-009-9129-8.

Stephens, M. 2000. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62 (4):795–809.

Su, Y.-S., and M. Yajima. 2015. R2jags: Using R to run 'JAGS'. *R package version 0.5-7*. 34.

Tohidi, M., F. Hadaegh, H. Harati, and F. Azizi. 2009. C-reactive protein in risk prediction of cardiovascular outcomes: Tehran lipid and glucose study. *International Journal of Cardiology* 132 (3):369–374. doi: 10.1016/j.ijcard.2007.11.085.

Verbeke, G., G. Molenberghs, and G. Verbeke. 1997. *Linear mixed models for longitudinal data* (pp.63–153). New York: Springer.

Viviani, S., M. Alfó, and D. Rizopoulos. 2014. Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing* 24 (3):417–427. doi: 10.1007/s11222-013-9378-4.

Wu, L., W. Liu, and X. Hu. 2010. Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. *Biometrics* 66 (2):327–335. doi: 10.1111/j.1541-0420.2009.01308.x.

Wulfsohn, M. S., and A. A. Tsiatis. 1997. A joint model for survival and longitudinal data measured with error. *Biometrics* 53 (1):330–339.

Xu, J., and S. L. Zeger. 2001. Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 50 (3):375–387.

Zhudenkov, K., S. Gavrilov, A. Sofronova, O. Stepanov, N. Kudryashova, G. Helmlinger, and K. Peskov. 2022. A workflow for the joint modeling of longitudinal and event data in the development of therapeutics: Tools, statistical methods, and diagnostics. *CPT: Pharmacometrics & Systems Pharmacology* 11 (4):425–437. doi: 10.1002/psp4.12763.