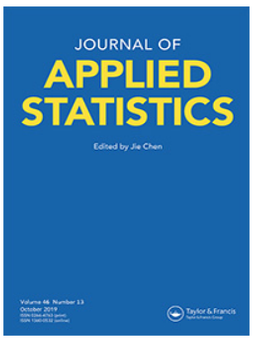

Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Joint modelling of longitudinal binary data and survival data

Yi-Ting Hwang, Chia-Hui Huang, Chun-Chao Wang, Tzu-Yin Lin & Yi-Kuan Tseng

To cite this article: Yi-Ting Hwang, Chia-Hui Huang, Chun-Chao Wang, Tzu-Yin Lin & Yi-Kuan Tseng (2019) Joint modelling of longitudinal binary data and survival data, Journal of Applied Statistics, 46:13, 2357-2371, DOI: 10.1080/02664763.2019.1590540

To link to this article: https://doi.org/10.1080/02664763.2019.1590540

JOURNAL OF APPLIED STATISTICS
2019, VOL. 46, NO. 13, 2357–2371
https://doi.org/10.1080/02664763.2019.1590540

Taylor & Francis
Taylor & Francis Group

Joint modelling of longitudinal binary data and survival data

Yi-Ting Hwang[a], Chia-Hui Huang[a], Chun-Chao Wang[a], Tzu-Yin Lin[a] and Yi-Kuan Tseng[b]

[a]National Taipei University, New Taipei City, Taiwan; [b]Graduate Institute of Statistics, National Central University, Taoyuan, Taiwan

ABSTRACT

The medical costs in an ageing society substantially increase when the incidences of chronic diseases, disabilities and inability to live independently are high. Healthy lifestyles not only affect elderly individuals but also influence the entire community. When assessing treatment efficacy, survival and quality of life should be considered simultaneously. This paper proposes the joint likelihood approach for modelling survival and longitudinal binary covariates simultaneously. Because some unobservable information is present in the model, the Monte Carlo EM algorithm and Metropolis-Hastings algorithm are used to find the estimators. Monte Carlo simulations are performed to evaluate the performance of the proposed model based on the accuracy and precision of the estimates. Real data are used to demonstrate the feasibility of the proposed model.

ARTICLE HISTORY
Received 2 May 2018
Accepted 26 February 2019

KEYWORDS
Cox model; generalized linear model; metropolis-hastings algorithm; Monte Carlo EM algorithm; quality of life

1. Introduction

The medical costs in an ageing society dramatically increase when the incidences of chronic diseases, disabilities and inability to live independently are high. Healthy lifestyles not only affect elderly individuals but also influence the entire community. When assessing the efficacy of a treatment, two important primary endpoints are the survival time and the quality of life (QOL) of an individual. Thus, when studying potential influential factors that are associated with a healthy lifestyle, the survival time and the QOL should be addressed simultaneously.

Jointly studying two outcomes is quite common in medical studies. Many phase III clinical trials consider both survival time and QOL as the primary endpoints. Thus, a composite measure of quality and quantity of life is needed to be able to assess the efficacy of treatments. Many different techniques have been proposed to derive such a composite measure [3]. Although the composite measure accounts for the missing QOL due to death, it does not handle the missing data resulting from drop-out from the study prior to death.

CONTACT Yi-Ting Hwang hwangyt@gm.ntpu.edu.tw

Supplemental data for this article can be accessed here. https://doi.org/10.1080/02664763.2019.1590540

By contrast, modelling-based approaches can explicitly address drop-out. Modelling-based approaches for jointly considering continuously repeated measures and the survival outcome have been discussed. The difference between various models is how to incorporate the predicted values derived from the general linear model into the survival model. Three major parameter estimations include the two-stage approach [26], the likelihood-based approach [14,28] and the Bayesian approach [4]. A detailed overview is presented in [25].

A limited number of studies that jointly assess the association between longitudinal QOL data and time-to-event data are discussed. First, the quality of life might be measured by a well-developed questionnaire, such as the 36 item short-form health survey (SF36; [27]) and preference-based health utilities index instrument [7,9], which are often summarized in a continuous measure for the QOL. Ribaudo, Thompson and Allen-Mersh [20] formed trivariate normal models for analysing the QOL and the log transformation of the survival time. Xu and Zeger [29] jointly modelled the general linear model for longitudinal QOL and the Cox proportional hazards model conditionally on the latent quantity of QOL that is obtained from the general linear model. Zeng and Cai [30] assumed common random effects in both models to account for a possible dependence between the QOL and the survival time.

The majority of the recent simultaneous modellings are focused on modelling the relationship between the continuous longitudinal measures and the time-to-event data. However, in some instances, it might only be possible to obtain a binary outcome that indicates the respondent's health condition or the QOL because of limited resources. Furthermore, Rizopoulos *et al.* [22] presented a study that has an aim of finding the association between longitudinal proteinuria measurements and the time to renal graft failure for patients with a renal allograft, where proteinuria is treated as a binary outcome. Also, Choi *et al.* [5] investigated the ethnic difference in binary longitudinal satisfaction outcomes and the survival time. Since the binary repeated measurement is related to the individual survival time, both outcomes need to be modelled simultaneously. Rizopoulos *et al.* [22] suggested using the parametric survival model with a random effect to capture unobserved heterogeneity and a mixed effects logistic regression to model the binary longitudinal effect and assumed that the random effects from these two models have a multiplicative relation. Furthermore, Choi *et al.* [5] suggested a joint analysis of survival time and longitudinal categorical outcomes, where the survival time was modelled by the stratified multiplicative hazards model and longitudinal categorical outcomes was analyzed by the generalized linear model. In particular, the joint analysis between these two models was linked by the common random effects.

Rather than sharing the random effect in the two models and using the parametric survival model, we adapt the joint model formulation proposed in [26,28], where the true mean trajectory for each subject is obtained from the mixed effect logistic regression and is used as a time-dependent covariate in the survival model. The joint likelihood function approach is used to find the estimator of parameters.

The remainder of this paper is organized as follows. A general framework and model assumptions for modelling the longitudinal binary outcome and the survival time are presented in Section 2. The joint likelihood function is established and the corresponding numerical algorithm is described in Section 3. Simulations are performed to evaluate the performance of the estimators in Section 4. In Section 5, data collected by the Department

of Health (DOH) in Taiwan are used to illustrate the usage of the proposed model. Further discussion and generalizations are provided in Section 6.

2. Model and notation

Let T_i and C_i denote the survival time and censoring time for the ith patient, $i = 1, \ldots, n$. Assume that T_i and C_i are independent. In practice, either T_i or C_i is observed. Let the ith observable data be denoted as $X_i = \min(T_i, C_i)$ and $\Delta_i = I[T_i \leq C_i]$, where $I[A]$ is an indicator function of the event A. For the ith patient, let $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{ip})'$ be the vector of p time-independent covariates and $Z_i^*(t)$ be a time-dependent covariate. Additionally, let the full time trajectory of $Z_i^*(t)$ prior to t be denoted as $\bar{\boldsymbol{Z}}_i^* = \{Z_i^*(s), s \leq t\}$. Moreover, let $\lambda(t \mid \boldsymbol{\eta}_i, \bar{\boldsymbol{Z}}_i^*)$ denote the hazard function at time t for the ith patient given $\boldsymbol{\eta}_i$ and $\bar{\boldsymbol{Z}}_i^*$. If it is possible to observe the complete history of $Z_i^*(t)$, then the association between the survival and the corresponding risk factors can be established by the Cox proportional hazards model:

$$\begin{aligned}\lambda(t \mid \boldsymbol{\eta}_i, \bar{\boldsymbol{Z}}_i^*) &= \lambda_0(t) \exp\{\boldsymbol{\eta}_i' \boldsymbol{\gamma} + \beta Z_i^*(t)\} \\ &= \lambda_0(t) \exp\left\{\sum_{k=1}^{p} \gamma_k \eta_{ik} + \beta Z_i^*(t)\right\}, \end{aligned} \tag{1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)'$ and β are parameters associated with risk factors and a time-dependent covariate. The parameters in (1) can be estimated through the partial likelihood function proposed by [6].

The complete history of the time-dependent covariate is often unable to be obtained in practice since the respondent is only subject to follow-up on a regular basis. Furthermore, measurement errors may exist when obtaining the time-dependent covariate. When this covariate is measured continuously, a general linear model for the repeated measure can be used to derive the true complete history of the time-dependent covariate. Based on this model and (1), Wulfsohn and Tsiatis [28] constructed a joint likelihood function that simultaneously takes both the survival and the continuous repeated measure into account to obtain the estimators of model parameters.

However, when this time-dependent covariate is a binary variable, a generalized linear model for the repeated measure is needed to find the true subject-specific trajectory. Suppose that for the ith patient, a sequence of binary measures observed at time $\boldsymbol{t}_i = (t_{i1}, t_{i2}, \ldots, t_{im_i})'$ is obtained and is denoted as Z_{ij}, $j = 1, \ldots, m_i$. The time-dependent covariate can be assumed to be a function of $\boldsymbol{\theta}' \boldsymbol{f}(u)$, where $\boldsymbol{f}(u)$ is a $(q \times 1)$ vector of a function of u and $\boldsymbol{\theta}$ is a $(q \times 1)$ random effect vector. In practice, it is often assumed to be a known polynomial function, e.g.

$$\boldsymbol{\theta}' \boldsymbol{f}(u) = \sum_{l=0}^{q-1} \theta_{il} u^l. \tag{2}$$

The generalized linear model for the repeated binary measure can then be formulated as follows:

(1) Suppose that $\boldsymbol{\theta}_i$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_{q-1})'$ and $q \times q$ covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0,q-1} \\ \sigma_{10} & \sigma_{11} & \cdots & \sigma_{1,q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{q-1,0} & \sigma_{q-1,1} & \cdots & \sigma_{q-1,q-1} \end{pmatrix}.$$

(2) Given (2), the conditional distribution of Z_{ij} is a Bernoulli distribution with probability of success π defined as

$$\pi(\boldsymbol{\theta}_i; t_{ij}) = \frac{\exp(\sum_{l=0}^{q-1} \theta_{il} t_{ij}^l)}{1 + \exp(\sum_{l=0}^{q-1} \theta_{il} t_{ij}^l)}, \tag{3}$$

where $\pi(\boldsymbol{\theta}_i;\ t_{ij}) = E[Z_{ij} \mid \boldsymbol{\theta}_i]$.

(3) Given $\boldsymbol{\theta}_i$, Z_{ij}, $j = 1, \ldots, m_i$, $i = 1, \ldots, n$, are independent.

3. Proposed models

The proposed models include the generalized linear mixed model and the Cox model. The following discusses the estimation approach. To ease the notation, set $q = 2$. To be able to incorporate the binary repeated measure into the Cox model, we define the following models:

$$\pi(\boldsymbol{\theta}_i;\ t_{ij}) = \frac{\exp(\theta_{i0} + \theta_{i1} t_{ij})}{1 + \exp(\theta_{i0} + \theta_{i1} t_{ij})}, \tag{4}$$

$$\lambda(t \mid \boldsymbol{\eta}_i, \bar{\boldsymbol{\pi}}_i) = \lambda_0(t) \exp\{\boldsymbol{\eta}_i' \boldsymbol{\gamma} + \beta \pi(\boldsymbol{\theta}_i; t)\}, \tag{5}$$

where $\bar{\boldsymbol{\pi}}_i = \{\pi(\boldsymbol{\theta}_i; s),\ s \le t\}$.

The following constructs the joint likelihood function for the observable data and obtains the model estimators based on the observed likelihood function. Let $\boldsymbol{\mathcal{D}}_i = (\boldsymbol{Z}_i, \boldsymbol{t}_i, X_i, \Delta_i, \boldsymbol{\eta}_i)$ denote the observable data for the ith patient and $\boldsymbol{\Omega} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda_0, \beta, \boldsymbol{\gamma}\}$ denote the parameter space, where $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{im_i})'$. Then, the observed joint likelihood function for $\boldsymbol{\mathcal{D}}_i$, $i = 1, \ldots, n$, is

$$L(\boldsymbol{\Omega}) = \prod_{i=1}^{n} \left\{ \int \left[\prod_{j=1}^{m_i} f(Z_{ij} \mid \boldsymbol{\theta}_i) \right] f(\boldsymbol{\theta}_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) f(X_i, \Delta_i \mid \boldsymbol{\theta}_i, \lambda_0, \beta, \boldsymbol{\gamma})\, \mathrm{d}\boldsymbol{\theta}_i \right\}, \tag{6}$$

where

$$\begin{aligned} f(Z_{ij} \mid \boldsymbol{\theta}_i) &= \pi(\boldsymbol{\theta}_i;\ t_{ij})^{Z_{ij}} (1 - \pi(\boldsymbol{\theta}_i;\ t_{ij}))^{1 - Z_{ij}}, \\ f(\boldsymbol{\theta}_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{\exp\{-(\boldsymbol{\theta}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\mu})/2\}}{(2\pi)^{q/2} |\boldsymbol{\Sigma}|^{1/2}}, \\ f(X_i, \Delta_i \mid \boldsymbol{\theta}_i, \lambda_0, \beta, \boldsymbol{\gamma}) &= \left[\lambda_0(X_i) \exp\{\boldsymbol{\eta}_i' \boldsymbol{\gamma} + \beta \pi(\boldsymbol{\theta}_i; X_i)\}\right]^{\Delta_i} \\ &\quad \times \exp\left[- \int_0^{X_i} \lambda_0(u) \exp\{\boldsymbol{\eta}_i' \boldsymbol{\gamma} + \beta \pi(\boldsymbol{\theta}_i; u)\}\, \mathrm{d}u \right]. \end{aligned} \tag{7}$$

Since random effects are unobservable, the estimates cannot be found directly. The EM algorithm is needed. Let $\hat{\boldsymbol{\Omega}} = \{\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \hat{\lambda}_0, \hat{\beta}, \hat{\boldsymbol{\gamma}}\}$ denote the estimates for the current step. In the E step, given $\hat{\boldsymbol{\Omega}}$, the expected log-likelihood value can be computed. Given the observable data and the parameter estimates, we can derive the conditional probability density function as

$$f(\boldsymbol{\theta}_i \mid X_i, \Delta_i, \boldsymbol{Z}_i, \hat{\boldsymbol{\Omega}}) = \frac{f(X_i, \Delta_i \mid \boldsymbol{\theta}_i, \hat{\lambda}_0, \hat{\beta}, \hat{\boldsymbol{\gamma}}) f(\boldsymbol{\theta}_i \mid \boldsymbol{Z}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})}{\int_{-\infty}^{\infty} f(X_i, \Delta_i \mid \boldsymbol{\theta}_i, \hat{\lambda}_0, \hat{\beta}, \hat{\boldsymbol{\gamma}}) f(\boldsymbol{\theta}_i \mid \boldsymbol{Z}_i, \boldsymbol{t}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})\, \mathrm{d}\boldsymbol{\theta}_i}. \tag{8}$$

Let $h(\cdot)$ be any arbitrary function. Given the observable data and the parameter estimates, the conditional expectation of $h(\cdot)$ has the following representation:

$$\begin{aligned} E_i[h(\pi(\boldsymbol{\theta}_i; t) \mid \boldsymbol{Z}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}] &= E_i[h(\pi(\boldsymbol{\theta}_i; t)] \\ &= \frac{\int_{-\infty}^{\infty} h(\pi(\boldsymbol{\theta}_i; t)) f(X_i, \Delta_i \mid \boldsymbol{\theta}_i, \hat{\lambda}_0, \hat{\beta}, \hat{\boldsymbol{\gamma}}) f(\boldsymbol{\theta}_i \mid \boldsymbol{Z}_i, \boldsymbol{t}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})\, \mathrm{d}\boldsymbol{\theta}_i}{\int_{-\infty}^{\infty} f(X_i, \Delta_i \mid \boldsymbol{\theta}_i, \hat{\lambda}_0, \hat{\beta}, \hat{\boldsymbol{\gamma}}) f(\boldsymbol{\theta}_i \mid \boldsymbol{Z}_i, \boldsymbol{t}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})\, \mathrm{d}\boldsymbol{\theta}_i}. \end{aligned} \tag{9}$$

In contrast to the joint likelihood approach for the general linear model and the survival model as given in [28], $f(\boldsymbol{\theta}_i \mid \boldsymbol{Z}_i, \boldsymbol{t}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ defined in (8) is a mixture distribution. The random sample cannot be generated directly. The Metropolis-Hastings algorithm proposed by [19] and [12] is used to generate the random sample. The detailed step is listed in Appendix 1. Let the M MH samples be denoted as $\pi(\boldsymbol{\theta}_i^{(k)}; t)$, $k = 1, \ldots, M$, where only the last $M-B$ samples are used in the computation. The conditional expectation in (9) can be approximated by

$$E_i(h(\pi(\boldsymbol{\theta}_i; t)) \approx \frac{\sum_{k=B+1}^{M} h(\pi(\boldsymbol{\theta}_i^{(k)}; t)) f(X_i, \Delta_i \mid \boldsymbol{\theta}_i^{(k)}, \hat{\lambda}_0, \hat{\beta}, \hat{\boldsymbol{\gamma}})/(M-B)}{\sum_{k=B+1}^{M} f(X_i, \Delta_i \mid \boldsymbol{\theta}_i^{(k)}, \hat{\lambda}_0, \hat{\beta}, \hat{\boldsymbol{\gamma}})/(M-B)}. \tag{10}$$

From the E step, we obtain the complete data $\mathcal{D}_i^* = (\boldsymbol{Z}_i, \pi(\boldsymbol{\theta}_i; t), X_i, \Delta_i, \boldsymbol{\eta}_i)$. Based on these data, we can then obtain the MLE of $\hat{\boldsymbol{\Omega}}$. The detailed derivation of the M step is given in Appendix 2, and the estimates are denoted as follows:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} E_i(\boldsymbol{\theta}_i), \tag{11}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} E_i[(\boldsymbol{\theta}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{\theta}_i - \hat{\boldsymbol{\mu}})'], \tag{12}$$

$$\hat{\lambda}_0(u) = \sum_{i=1}^{n} \frac{\Delta_i I(X_i = u)}{\sum_{j=1}^{n} E_j \left[\exp\left\{ \beta \pi(\boldsymbol{\theta}_j; u) + \boldsymbol{\eta}_j' \boldsymbol{\gamma} \right\} \right] I(X_j \geq u)}. \tag{13}$$

Since $\boldsymbol{\gamma}$ and β do not have closed forms, FMINSEARCH in MATLAB is used to find the numerical solution for $\boldsymbol{\gamma}$ and β [18].

The standard error of estimates cannot be derived directly. Adapting the estimation procedure suggested by [15,23], the standard error of estimates is obtained by the bootstrap samples. The procedure to obtain the bootstrap estimates is as follows:

Step 1. Use $\mathcal{D} = \{\mathcal{D}_i,\ i = 1, \ldots, n\}$ to generate a bootstrap sample $\mathcal{D}_l^b$ and use EM to find MLE $\hat{\boldsymbol{\Omega}}_l^b$.

Step 2. Repeat Step 1 L times to obtain $\hat{\boldsymbol{\Omega}}_l^b,\ l = 1, \ldots L$.

Step 3. Compute $\hat{\text{Cov}}(\hat{\boldsymbol{\Omega}}^b)$ as

$$\hat{\text{Cov}}(\hat{\boldsymbol{\Omega}}^b) = \frac{1}{L-1}\sum_{l=1}^{L}(\hat{\boldsymbol{\Omega}}_l^b - \bar{\boldsymbol{\Omega}}^b)(\hat{\boldsymbol{\Omega}}_l^b - \bar{\boldsymbol{\Omega}}^b)',$$

where $\bar{\boldsymbol{\Omega}}^b = \sum_{l=1}^{L}\hat{\boldsymbol{\Omega}}_l^b/L$.

4. Monte Carlo simulations

Monte Carlo simulations are conducted to evaluate the performance of the estimates of the proposed model. Eleven repeated measures that are measured equally spaced are considered. Let the binary repeated measures Z_{ij} be generated from a generalized linear random effects model ($q = 2$) as

$$\log\left(\frac{\pi(\boldsymbol{\theta}_i;\ t_{ij})}{1-\pi(\boldsymbol{\theta}_i;\ t_{ij})}\right) = \theta_{0i} + \theta_{1i}t_{ij}, \tag{14}$$

where $t_{ij} = j, j = 1, 2, \ldots, 11$, and $\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i})'$ has a bivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix}\sigma_{00} & \sigma_{01}\\ \sigma_{10} & \sigma_{11}\end{pmatrix}.$$

In particular, we set $\boldsymbol{\mu} = (2, -1)'$ and $\sigma_{00} = 1$, $\sigma_{01} = 0.4$ and $\sigma_{11} = 0.5$.

In addition to the time-dependent covariate, the survival time is assumed to be associated with a binary fixed effect (η_i), which is generated from the Bernoulli random generator with mean 0.5. The survival time T_i is generated from the Cox model defined as

$$\lambda(t \mid \eta_i, \bar{\boldsymbol{\pi}}_i) = \lambda_0(t)\exp\{\beta\pi(\boldsymbol{\theta}_i; t) + \gamma\eta_i\}, \tag{15}$$

where the baseline survival function is assumed to be linear in time, i.e. $\lambda_0(t) = \alpha t$ and $\alpha = 0.05$, $\beta = -0.1$ and $\gamma = 0.3$. Moreover, the censoring time is assumed to have an exponential distribution with mean μ_c, where μ_c is determined by the probability of censoring. In this paper, two different probabilities of censoring, 10% and 30%, are chosen. The corresponding settings for μ_c are 70 and 15. The following results are based on 500 random samples. Due to censoring, the average number of repeated measures for 15% and 30% censoring rates are 4.53 and 3.80, respectively.

The performance of the estimates is evaluated by 5 indices: bias, percent of bias, $\sqrt{\text{SSE}}$, SE and CP. The bias is defined as the average of the estimate minus the true value computed from 500 samples. The relative bias is defined as the average of the bias relative to the true value. When the estimate is unbiased, the percent of bias is close to zero. The $\sqrt{\text{SSE}}$ is the sample standard error of estimates. The SE for the joint likelihood approach is the average of the standard error derived from the bootstrap samples, where the number of bootstrap

samples equals $L = 40$, which is suggested by [16]. The CP is the 95% coverage probability, which is computed from the average of 500 samples of the status of the confidence interval of the parameter covering the true value.

The proposed model settings include two models: the generalized linear model and the Cox model. Similar to the two-stage approach proposed in [26], the estimates can be derived from two stages. The first stage is to find the predicted individual mean trajectories derived from the generalized linear model. The second stage uses the predicted individual mean value $\hat{\pi}(\boldsymbol{\theta}_i; t)$, which is treated as the time-dependent covariate in (1). The Cox model becomes

$$\lambda(t \mid \boldsymbol{\eta}_i, \bar{\boldsymbol{\pi}}_i) = \lambda_0(t)\exp\{\boldsymbol{\eta}_i'\boldsymbol{\gamma} + \beta\hat{\pi}(\boldsymbol{\theta}_i; t)\}.$$

The following uses estimates derived from the two-stage model as a comparison. The advantage of this method is the estimates can be derived directly from the well-known statistical packages. In this paper, the GLIMMIX procedure in SAS is used to find $\hat{\pi}(\boldsymbol{\theta}_i; t)$, which are obtained by the marginal log likelihood function, and the adaptive Gauss-Hermite quadrature rule is chosen to solve the estimates iteratively. The PHREG procedure in SAS is then used to find the estimates of $\boldsymbol{\gamma}$ and β. The SE for the two-stage approach is the average of the standard error derived from the GLIMMIX procedure.

Table 1 provides the performance of the simulation when the probability of censoring is 10%. When the two-stage approach is used, two estimates related to the Cox model have larger biases compared to the estimates obtained by the joint likelihood approach. In particular, the variable of interest evaluated by β is incorrectly estimated by the two-stage estimation. Both approaches are able to yield small biases of the estimates related to the generalized linear model. Most of relative bias of the estimates are slightly larger than 0. Specifically, only the estimates of the covariance of the intercept and the slope (σ_{01}) and the fixed effect (γ) have negative relative biases for the two-stage estimation, whereas only one estimate of the covariance of the intercept and the slope (σ_{01}) has negative relative biases

Table 1. Performance of estimates for two estimation procedures under a 10% censoring rate.

		Two-stage model					Joint model				
Parameter	True	Bias	RB[#]	$\sqrt{\text{SSE}}$	SE[$]	CP	Bias	RB	$\sqrt{\text{SSE}}$	SE*	CP
$n = 200$											
μ_0	2	0.0229	1.14%	0.2084	0.2100	95.20%	0.0402	2.01%	0.1537	0.1572	93.00%
μ_1	−1	−0.0074	0.74%	0.1134	0.1117	95.00%	−0.0258	2.58%	0.0955	0.0933	92.60%
σ_{00}	1	0.0873	8.73%	0.6298	0.6171	86.20%	−0.0573	−5.73%	0.3097	0.3426	87.60%
σ_{01}	0.4	−0.0170	−4.24%	0.2056	0.5126	90.80%	0.0347	8.66%	0.1358	0.1479	96.00%
σ_{11}	0.5	0.0247	4.94%	0.1902	0.9342	91.00%	0.0255	5.11%	0.1365	0.1458	95.20%
β	−0.1	−1.2219	1221.88%	0.1870	0.1073	0.00%	−0.0056	5.57%	0.2674	0.2731	95.00%
γ	0.3	−0.0750	−25.00%	0.1441	0.1044	79.60%	0.0032	1.05%	0.1524	0.1560	94.80%
$n = 400$											
μ_0	2	0.0017	0.66 %	0.1516	0.1485	93.20%	0.0438	2.19%	0.0963	0.0981	90.60%
μ_1	−1	0.0032	0.03%	0.0775	0.0789	96.00%	−0.0275	2.75%	0.0591	0.0622	94.00%
σ_{00}	1	0.0315	6.94%	0.4317	0.4450	91.60%	−0.0834	−8.34%	0.1444	0.1678	89.20%
σ_{01}	0.4	0.0071	−7.17%	0.1502	0.3615	98.60%	0.0522	13.05%	0.0846	0.0930	94.80%
σ_{11}	0.5	0.0188	7.87%	0.1318	0.6480	98.60%	0.0374	7.48%	0.0846	0.0913	95.00%
β	−0.1	−1.2135	1124.98%	0.1272	0.0750	0.00%	−0.0183	18.34%	0.1830	0.1837	94.20%
γ	0.3	−0.0697	−22.23%	0.1077	0.0732	71.80%	0.0086	2.87%	0.1072	0.1071	92.94%

[#] RB stands for the relative bias.
[$] Average SE computed from the GLIMMIX procedure.
* Average SE derived from bootstrap samples.

for the two-stage estimation. The relative bias for β is extremely large under the two-stage estimation. On the contrary, the relative bias for β under the joint model estimation is less than 10%. The performance in terms of $\sqrt{\text{SSE}}$ and SE is similar when the joint likelihood estimation is used. Nevertheless, $\sqrt{\text{SSE}}$ and SE derived from the two-stage approach are not similar. In particular, the SEs of the estimates for the Cox model are smaller than $\sqrt{\text{SSE}}$, whereas those of the estimates for the covariance of the generalized linear model are larger than $\sqrt{\text{SSE}}$. The CPs for the estimates derived from the joint likelihood approach are closer to the nominal confidence level than those obtained from the two-stage approach. In particular, the CP for the estimate of β derived from the two-stage approach is 0 due to a larger bias.

The performance in term of the sample size states as follows. The bias increases slightly as n increases. In term, the relative bias also increases. Nevertheless, as expected, the sample size has a great impact on SE. Increasing the sample size reduces the $\sqrt{\text{SSE}}$ and SE. Finally, the performance in term of CP varies. Most of the estimators derived from the generalized linear model become closer to the nominal confidence level, but the estimators derived from the survival model are away from the nominal confidence level slightly.

Table 2 lists the results of the simulation when the probability of censoring is 30%. The bias of the estimates derived from the joint likelihood approach increases slightly. The relative bias also increases. In contrast, the bias of the estimates derived from the two-stage approach decreases slightly when $n = 200$, whereas when $n = 400$, only some bias of the estimates decreases.

When using the joint likelihood approach, $\sqrt{\text{SSE}}$ and SE of the estimate increase slightly. In particular, $\sqrt{\text{SSE}}$ remains slightly larger than SE. However, the magnitudes of $\sqrt{\text{SSE}}$ and SE derived from the two-stage approach change when the censoring rate changes. Since the bias of the estimates related to the binary repeated measures derived from the joint likelihood approach is slightly larger, the CP for these estimates is slightly away from the

Table 2. Performance of estimates for two estimation procedures under a 30% censoring rate.

		Two-stage model					Joint model				
Parameter	True	Bias	RB[#]	$\sqrt{\text{SSE}}$	SE[$]	CP	Bias	RB	$\sqrt{\text{SSE}}$	SE*	CP
$n = 200$											
μ_0	2.0	0.0177	0.89%	0.2156	0.2214	95.00%	0.0520	2.60%	0.1525	0.1674	95.80%
μ_1	−1.0	−0.0020	0.20%	0.1216	0.1214	95.20%	−0.0304	3.04%	0.0970	0.1019	94.80%
σ_{00}	1.0	0.1288	12.88%	0.7181	0.6627	88.20%	−0.0242	−2.42%	0.3433	0.3961	95.40%
σ_{01}	0.4	−0.0328	−8.20%	0.2252	0.5629	91.20%	0.0430	10.76%	0.1558	0.1723	95.60%
σ_{11}	0.5	0.0398	7.96%	0.1993	1.0014	91.40%	0.0445	8.91%	0.1351	0.1679	97.20%
β	−0.1	−1.1243	1124.26%	0.1817	0.1238	0.00%	0.0211	−21.07%	0.3113	0.3191	94.80%
γ	0.3	−0.0638	−21.26%	0.1768	0.1204	78.60%	0.0013	0.42%	0.1776	0.1761	93.00%
$n = 400$											
μ_0	2.0	0.0132	0.66%	0.1557	0.1590	95.00%	0.0687	3.43%	0.1120	0.1065	87.40%
μ_1	−1.0	−0.0003	0.03%	0.0848	0.0867	94.40%	−0.0410	4.10%	0.0694	0.0683	88.40%
σ_{00}	1.0	0.0694	6.94%	0.4712	0.4891	91.60%	−0.0764	−7.64%	0.1550	0.1868	90.60%
σ_{01}	0.4	−0.0287	−7.17%	0.1724	0.3992	96.80%	0.0540	13.51%	0.0985	0.1007	93.60%
σ_{11}	0.5	0.0393	7.87%	0.1651	0.7438	97.20%	0.0500	10.00%	0.0995	0.1003	94.40%
β	−0.1	−1.1250	1124.98%	0.1269	0.0864	0.00%	−0.0268	26.77%	0.2242	0.2144	92.20%
γ	0.3	−0.0667	−22.23%	0.1286	0.0842	74.40%	0.0014	0.47%	0.1257	0.1214	94.60%

[#]RB stands for the relative bias.
[$]Average SE computed from the GLIMMIX procedure.
*Average SE derived from bootstrap samples.

nominal confidence level. The performance in terms of the CP for the two-stage model is similar to that for the lower censoring rate.

A compute with an Intel(R) core (TM) i7-7700 CPU @ 3.60 GHz processor and 64 GB RAM is used to perform all the simulations. The average convergent time including finding the estimates and the corresponding standard error based on 40 bootstrap samples for a data set is 30 minutes.

5. A case study

Due to ageing, the Department of Health (DOH) in Taiwan launched a longitudinal study 30 years ago to investigate the physical and mental functions in the elderly in Taiwan. This study aimed to formulate the needs of medical insurance and social support in the future. This study began in 1987. The respondents were followed every 3-4 years for approximately 20 years. Excluding data that were answered by surrogates and had missing important study variables, there were 739 eligible individuals that resided in northern Taiwan. Each respondent was contacted in 1989, 1993, 1996, 1999, 2003 and 2007. The number of repeated measures varies from 1 to 6 times, and the average number of repeated measures was 3.12 times.

Although the study was a longitudinal follow-up study, the questionnaire for each run was not consistent. We were able to identify four binary questions that were related to satisfaction and were consistent in 6 follow-up runs. This first question was 'compared with others, your life is better than others'. The second question was 'You feel satisfied?'. The third question was 'You expect some good things would happen?'. The fourth question was 'You have the best time currently'. When the respondent chooses more than 2 items, the respondent is defined as having high satisfaction; otherwise, they are defined as having low satisfaction. Figure 1 displays the percent and log odds of having high satisfaction over time, where the solid line stands for the percent of satisfaction (%) and the dashed line represents the log odds. The percent and log odds of high satisfaction decrease at the beginning and then increase after 1993.

Three explanatory variables were selected, including age, gender and hypertension. The respondent's age in 1989 was used and was classified into 2 groups: less than 75 and equal to or over 75 years of age. Additionally, the status of hypertension used the information collected in 1989. Table 3 lists the median survival time stratified by each variable. The log rank test is used to assess the association between the survival time and risk factors. Female respondents who were less than 75 years of age and did not have hypertension had significantly larger median survival times.

Based on the log odds over time for the high satisfaction in Figure 1, the individual time trajectory for the longitudinal binary response is assumed to have a piecewise linear form and is defined as

$$\log\left(\frac{\pi(\boldsymbol{\theta}_i; t_{ij})}{1-\pi(\boldsymbol{\theta}_i;\ t_{ij})}\right) = \theta_{0i} + \theta_{1i}t_{ij} + \theta_{2i}(t_{ij})_+,$$

where $(t)_+$ equals t if $t > 4$ and equals 0 otherwise. Controlling for age, gender and hypertension, the Cox model is defined as

$$\lambda(t \mid \boldsymbol{\eta}_i, \bar{\boldsymbol{\pi}}_i) = \lambda_0(t)\exp(\beta\pi(\boldsymbol{\theta}_i; t) + \gamma_1\eta_{i1} + \gamma_2\eta_{i2} + \gamma_3\eta_{i3}), \tag{16}$$

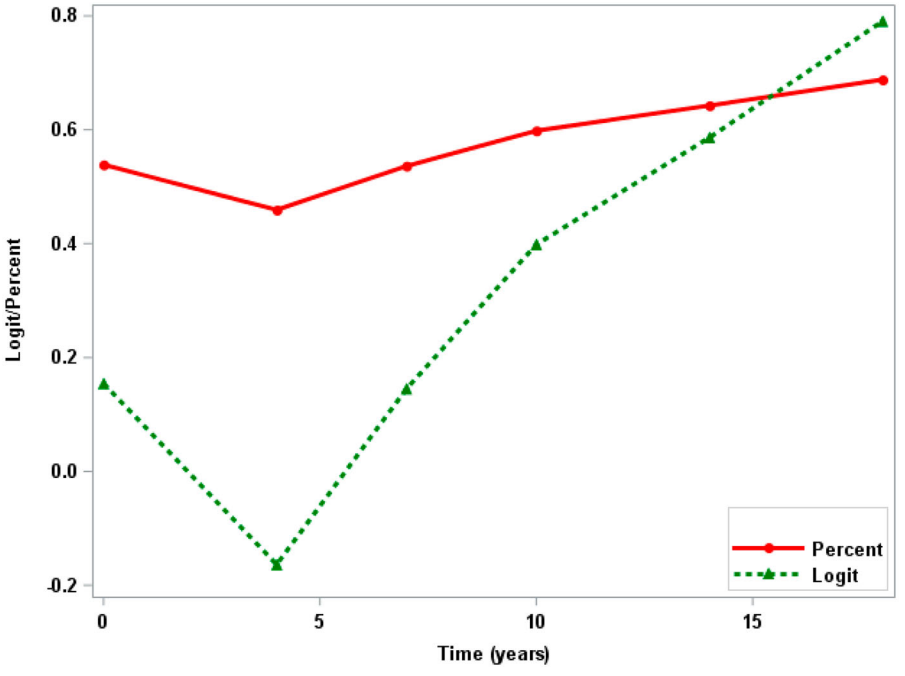

Figure 1. Percent and log odds of having high satisfaction for the case study.

Table 3. Median survival time stratified by gender, age and hypertension for the case study.

Variable	Category	Median survival	*p*-value*
Gender	Female (reference)	3.95	0.0004
	Male	3.06	
Hypertension	No (reference)	3.51	0.0111
	Yes	3.01	
Age	$<$ 75 (reference)	3.88	$<$ 0.0001
	$\geq$ 75	2.17	

*computed by log rank test.

where η_{i1} stands for male, η_{i2} stands for the status of hypertension, and η_{i3} represents the status of respondents aged 75 years old and older for subject i.

Table 4 provides the estimates derived from the two-stage approach and the joint likelihood approach. The parameter of interest, quality of life, is not significant when the two-stage approach is used, whereas it becomes very significant when the joint likelihood estimation is used. The estimate ($\hat{\beta} = -0.6067$) means that the higher the score of quality of life is, the lower is the hazard rate. When the score of the quality of life measured in the propensity ($\pi(\boldsymbol{\theta}_i; t)$) increases 0.1 unit, the hazard rate reduces 0.94 times. Estimates for other controlling variables derived from the two approaches are similar. Males who were 75 years old and over and had hypertension have higher hazard rates. Both approaches demonstrate that age has the greatest impact on survival. Furthermore, since the SE of estimates for the random effects derived from the two-stage approach is very large,

Table 4. Estimators derived from the two-stage and joint likelihood approaches for the case study.

	Two-stage approach					Joint likelihood approach				
Parameter	Estimate	SE*	*p*-value	95% CI		Estimate	SE*	*p*-value	95% CI	
β	−0.0705	0.3560	0.8431	0.464	1.873	−0.6067	0.1871	0.0012	−0.9734	−0.2400
γ_1	0.3132	0.1150	0.0065	1.092	1.714	0.4141	0.0810	< 0.0001	0.2553	0.5729
γ_2	0.3806	0.1303	0.0035	1.133	1.889	0.3294	0.1071	0.0021	0.1195	0.5393
γ_3	0.7728	0.1374	< .0001	1.655	2.835	0.9532	0.0796	< 0.0001	0.7972	1.1092
μ_0	0.1916	0.1132	0.0453	−0.030	0.414	−0.1846	0.0383	< 0.0001	−0.2597	−0.1095
μ_1	−0.3990	0.1557	0.9948	−0.704	−0.094	0.3038	0.0157	< 0.0001	0.2730	0.3346
μ_2	0.8457	0.2138	< .0001	0.427	1.265	−0.4622	0.0043	< 0.0001	−0.4706	−0.4538
σ_{00}	1.1383	2.4022	0.3178	−3.570	5.847	0.2072	0.0341	< 0.0001	0.1404	0.2740
σ_{01}	0.1664	3.3677	0.4803	−6.434	6.767	0.0893	0.0151	< 0.0001	0.0597	0.1189
σ_{02}	0.1702	1.4519	0.4533	−2.676	3.016	0.0034	0.0014	< 0.0001	0.0007	0.0061
σ_{11}	1.4816	7.2205	0.4187	−12.670	15.634	0.0451	0.0068	< 0.0001	0.0318	0.0584
σ_{12}	−2.0117	1.8652	0.8696	−5.668	1.644	0.0020	0.0006	0.0009	0.0008	0.0032
σ_{22}	2.9683	2.5142	0.1189	−1.960	7.896	0.0012	0.0002	< 0.0001	0.0008	0.0016

*SE based on bootstrap samples.

all the estimators except those for μ_0 and μ_2 are not significant. In particular, the two-stage approach results in an increasing mean trajectory in time, while the joint likelihood approach yields estimates with smaller SEs. In turn, all estimates are very significant from 0. Furthermore, when taking the individual survivals into account, the mean trajectory of the QOL increases slightly and then declines.

6. Discussion

This paper constructs models for analysing longitudinal binary QOL data and survival time simultaneously. Although the two-stage approach is a straightforward approach and is easily implemented with existing software, we again show the efficiency of estimates is not good. The joint likelihood approach requires some intensive computations, but this approach provides more accurate estimates. For the case study, the estimates derived from the two approaches are very different. The individual mean trajectory derived from the two-stage model is similar to the mean response curve, whereas that from the joint likelihood approach reveals an increasing trend and then decreasing trend over time. However, since the number of eligible participants decreases as time increases, the mean response curve presented in Figure 1 might be misleading.

The point mass construction for estimating the baseline hazard function is suggested by [28]. A similar construction can be also referred to [15]. Moreover, under the point mass construction, the consistency and asymptotic normality of the infinite dimensional cumulative baseline hazard and finite dimensional parameters have been theoretically verified in [5,30].

The proposed joint likelihood function is much complex than that discussed in [26] owing to the mixture distribution. Besides the usual EM algorithm, the MH algorithm is required. The MH algorithm is used to generate the random sample to find empirical estimates of the conditional expectation defined in (9). The candidate-generating distribution $g(\boldsymbol{\theta})$ is assumed to have a multivariate normal distribution with a mean vector $\boldsymbol{\theta}^{(B-1)}$ and a variance-covariance matrix $\boldsymbol{I}$. Normally, the MH algorithm uses data obtained from the previous step to update data for the current step. In a preliminary inspection, when the

variance-covariance matrix is replaced by the sample values from the previous step, the EM algorithm does not converge. Nevertheless, when $\boldsymbol{I}$ is chosen, the EM algorithm converges. The reason for this result might be that the estimates of the variance and covariance have slightly larger biases and have more conservative CP.

The predictive power of the logistic regression can be assessed by the ROC curve and the area under the ROC curve. Heagerty and Zheng [13] illustrated how to obtain the ROC curves for the standard Cox regression. Risopoulos [21] derived accuracy measures based on the ROC curve and the area under the ROC curve under the joint modelling framework for continuous longitudinal and time-to-event data. Based on these results, the predictive power of the proposed model might be derived.

The data used to illustrate the feasibility of the model actually have an ordinal repeated measure. To analyse the ordinal repeated measure, the proportional odds model might be considered. A similar formulation may be constructed. The observed joint likelihood function consists of the multinomial distribution, the multivariate normal distribution and the setting from the Cox model. The computation part will be more complex. In addition, the appropriateness of the proportional odds assumption has to be verified.

The Cox model is used to model the survival time. An extended hazard model proposed by [24] is a more general survival model and can be considered as an alternative model. Furthermore, Agresti [1] and Griffith, Hill and Pope [11] showed that the maximum likelihood estimator (MLE) of $\boldsymbol{\mu}$ can have significant bias for small samples. Also, the maximum likelihood estimates of logistic regression parameters are biased when the event of interest is rare [17]. Albert and Chib [2] used the exact Bayesian methods to perform exact inference for the small samples. Firth [8] proposed a penalized likelihood function to correct the bias. Based on the preliminary simulation, When $\pi(\boldsymbol{\theta}_i; \boldsymbol{t}) = 0.5$, the bias of of the estimate of β is small as given in Table 5. However, when $\pi(\boldsymbol{\theta}_i; \boldsymbol{t}) = 0.1$, the bias of the estimate of β is indeed 3 times larger than that when $\pi(\boldsymbol{\theta}_i; \boldsymbol{t}) = 0.3$ as given in Table 6. Thus, when

Table 5. Performance of estimate when $\pi(\theta; t_{ij}) = 0.5$ under a 30% censoring rate and $n = 400$.

Parameter	True	Average	bias	$\sqrt{SSE}$	CP*
μ_0	2	1.952	−0.048	0.1138	0.9486
μ_1	−0.5	−0.5024	−0.0024	0.0628	0.9571
σ_{00}	1	1.0273	0.0273	0.2637	0.9486
σ_{01}	−0.4	−0.4413	−0.0413	0.1445	0.9414
σ_{11}	0.5	0.5143	0.0143	0.1057	0.9543
β	−0.1	−0.1085	−0.0085	0.2159	0.9443
γ	0.3	0.3072	0.0072	0.1197	0.9529

*Standard error of CP is based on $\sqrt{\text{SSE}}$.

Table 6. Performance of estimate when $\pi(\boldsymbol{\theta}_i; \boldsymbol{t}) = 0.1$ under a 30% censoring rate and $n = 400$.

Parameter	True	Average	Bias	$\sqrt{SSE}$	CP*
μ_0	−1.0	−0.9948	0.0052	0.1180	0.9470
μ_1	−1.0	−0.9767	0.0233	0.1046	0.9382
σ_{00}	1.0	1.0089	0.0089	0.3854	0.9470
σ_{01}	−0.4	−0.4676	−0.0676	0.1931	0.9435
σ_{11}	0.5	0.5151	0.0151	0.1350	0.9417
β	−0.1	−0.0254	0.0746	0.6627	0.9435
γ	0.3	0.3040	0.0040	0.1208	0.9523

*Standard error of CP is based on $\sqrt{\text{SSE}}$.

the sample size is small or the event of interest is rare, the proposed estimation needs to be modified.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] A. Agresti, *Categorical Data Analysis*, 3rd ed., John Wiley and Sons Inc., Hoboken, 2013.
[2] J.H. Albert and S. Chib, *Bayesian analysis of binary and polychotomous response data*, J. Am. Stat. Assoc. 88 (1993), pp. 669–679.
[3] L.J. Billingham and K.R. Abrams, *Simultaneous analysis of quality of life and survival data*, Stat. Methods. Med. Res. 11 (2002), pp. 25–48.
[4] E.R. Brown and J.G. Ibrahim, *A bayesian semiparametric joint hierarchical model for longitudinal and survival data*, Biometrics 59 (2003), pp. 221–228.
[5] J. Choi, J. Cai, D. Zeng, and A.F. Olshan, *Joint analysis of survival time and longitudinal categorical outcomes*, Stat. Biosci. 7 (2015), pp. 19–47.
[6] D.R. Cox, *Partial likelihood*, Biometrika 62 (1975), pp. 269–276.
[7] D. Feeny, *The health utilities index: A tool for assessing health benefits*, PRO Newsletter 34 (2005), pp. 2–6.
[8] D. Firth, *Bias reduction of maximum likelihood estimates*, Biometrika 80 (1993), pp. 27–38.
[9] W.J. Furlong, D.H. Feeny, G.W. Torrence, and R.D. Barr, *The health utilities index (hui) system for assessing health-related quality of life in clinical studies*, Ann. Med. 33 (2001), pp. 375–384.
[10] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Behtari, and D.B. Rubin, *The Health Utilities Index (HUI) System for Assessing Health-related Quality of Life in Clinical Studies*, Chapman and Hall/CRC, USA, 2013.
[11] W.E. Griffith, R.C. Hill, and P.J. Pope, *Small sample properties of probit model estimators*, J. Am. Stat. Assoc. 339 (1987), pp. 929–937.
[12] W.K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika 57 (1970), pp. 97–109.
[13] P.J. Heagerty and Y. Zheng, *Survival model predictive accuracy and roc curves*, Biometrics 61 (2005), pp. 92–105.
[14] R. Henderson, P. Diggle, and A. Dobson, *Joint modelling of longitudinal measurements and event time data*, Biostatistics 4 (2000), pp. 465–480.
[15] F. Hsieh, Y.K. Tseng, and J.L. Wang, *Joint model of survival time and longitudinal data: Likelihood approach revisit*, Biometrics 62 (2006), pp. 1037–1043.
[16] Y.T. Hwang, C.C. Wang, C.H. Wang, Y.K. Tseng, and Y.J. Chang, *Joint model of multiple longitudinal measures and a binary outcome: An application to predict orthostatic hypertension for subacute stroke patients*, Biom. J. 57 (2015), pp. 661–675.
[17] G. King and L. Zeng, *Logistic regression in rare events data*, Polit. Anal. 9 (2001), pp. 137–163.
[18] J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright, *Convergence properties of the nelder-mead simplex method in low dimensions*, SIAM J. Optim. 9 (1998), pp. 112–147.
[19] N. Metropolis and S. Ulam, *The monte carlo method*, J. Am. Stat. Assoc. 44 (1949), pp. 335–341.
[20] H.J. Ribaudo, S.G. Thompson, and T.G. Allen-Mersh, *A joint analysis of quality of life and survival using a random effect selection model*, Stat. Med. 19 (2000), pp. 3237–3250.
[21] D. Rizopoulos, *Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data*, Biometrics 67 (2011), pp. 819–829.
[22] D. Rizopoulos, G. Verbeke, E. Lesaffre, and Y. Vanrenterghem, *A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros*, Biometrics 64 (2008), pp. 611–619.
[23] Y.K. Tseng, F. Hsieh, and J.L. Wang, *Joint modelling of accelerated failure time and longitudinal data*, Biometrika 92 (2005), pp. 587–603.

[24] Y.K. Tseng, Y.R. Su, M. Mao, and J.L. Wang, *An extended hazard model with longitudinal covariates*, Biometrika 102 (2015), pp. 135–150.
[25] A. Tsiatis and M. Davidian, *Joint modeling of longitudinal and time-to-event data: An overview*, Stat. Sin. 14 (2004), pp. 809–834.
[26] A. Tsiatis, V. DeGruttola, and M.S. Wulfsohn, *Modeling the relationship of survival to longitudinal data measured with error - applications to survival and cd4 counts in patients with aids*, J. Am. Stat. Assoc. 90 (1995), pp. 27–37.
[27] J.E. Ware and C.D. Sherbourne, *The mos 36-item short-form health survey (sf-36): I. Conceptual framework and item selection*, Med. Care 30 (1992), pp. 473–483.
[28] M.S. Wulfsohn and A.A. Tsiatis, *A joint model for survival and longitudinal data measured with error*, Biometrics 53 (1997), pp. 330–339.
[29] J. Xu and S.L. Zeger, *Joint analysis of longitudinal data comprising repeated measures and times to events*, Appl. Stat. 50 (2001), pp. 375–387.
[30] D. Zeng and J. Cai, *Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time*, Ann. Statist. 33 (2005), pp. 2132–2163.

Appendix 1. MH algorithm

Since the data include both categorical data and continuous data, the joint distribution function is a rather complex function. Due to the mixture of data, we cannot generate the random sample directly. The Metropolis-Hastings algorithm proposed by [10,12,19] is chosen to obtain the desired sample.

Let g_θ denote a candidate-generating distribution. A sample is randomly selected from g_θ and is accepted when it falls in the acceptance region. The process then continues until the desired sample is determined. The following provides the selection procedures:

Step 1. Let $B=0$ and given the current estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, generate a random sample $\boldsymbol{\theta}_i$ of size n from MVN($\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}$), where $\boldsymbol{\theta}_i^{(B)} = [\theta_{0i}\ \ \theta_{1i}]$, $i = 1, \ldots, n$. Let this random sample be denoted as $\vec{\boldsymbol{\theta}}^{(B)} = [\boldsymbol{\theta}_1^{(B)}\ \ \boldsymbol{\theta}_2^{(B)}\ \ \cdots\ \ \boldsymbol{\theta}_n^{(B)}]'$.

Step 2. Let $B=B+1$, $k=0$ and $\vec{\boldsymbol{\theta}}^{(B)} = \vec{\boldsymbol{\theta}}^{(B-1)}$.

Step 3. Let $k=k+1$ and generate $\boldsymbol{\theta}_k^* = (\theta_{0k}^*, \theta_{1k}^*)$ from $g(\boldsymbol{\theta}) = \text{MVN}(\vec{\boldsymbol{\theta}}^{(B-1)}, \boldsymbol{I})$, where $\boldsymbol{I}$ is an identity matrix.

Step 4. Replace the kth row in $\vec{\boldsymbol{\theta}}^{(B)}$ by $\boldsymbol{\theta}_k^*$ and denote the new sample

$$\vec{\boldsymbol{\theta}}_k^* = [\boldsymbol{\theta}_1^{(B)}\quad \boldsymbol{\theta}_2^{(B)}\quad \cdots\quad \boldsymbol{\theta}_{k-1}^{(B)}\quad \boldsymbol{\theta}_k^*\quad \boldsymbol{\theta}_{k+1}^{(B)}\quad \cdots\quad \boldsymbol{\theta}_n^{(B)}]'.$$

Step 5. Generate a random number p^* from $U(0,1)$ and replace $\vec{\boldsymbol{\theta}}^{(B)}$ by $\vec{\boldsymbol{\theta}}_k^*$ if $p^* \le p$, where

$$\begin{aligned} p = A_k(\vec{\boldsymbol{\theta}}^{(B-1)}, \vec{\boldsymbol{\theta}}_k^*) &= \min\left\{1, \frac{f(\vec{\boldsymbol{\theta}}_k^* \mid \boldsymbol{Z}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) g(\vec{\boldsymbol{\theta}}^{(B-1)})}{f(\vec{\boldsymbol{\theta}}^{(B-1)} \mid \boldsymbol{Z}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) g(\vec{\boldsymbol{\theta}}_k^*)}\right\} \\ &= \min\left\{1, \prod_{j=1}^{m_k} \exp\{Z_{kj}[\theta_{0k}^* - \theta_{0k}^{(B-1)} + (\theta_{1k}^* - \theta_{1k}^{(B-1)})t_{kj}]\}\right. \\ &\quad \times \left[\frac{1+\exp(\theta_{0k}^{(B-1)} + \theta_{1k}^{(B-1)} t_{kj})}{1+\exp(\theta_{0k}^* + \theta_{1k}^* t_{kj})}\right] \\ &\quad \left. \times \exp\left[-\frac{1}{2}(\boldsymbol{\theta}_k^* - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\theta}_k^* - \hat{\boldsymbol{\mu}}) + \frac{1}{2}(\boldsymbol{\theta}_k^{(B-1)} - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\theta}_k^{(B-1)} - \hat{\boldsymbol{\mu}})\right]\right\}, \end{aligned} \tag{A1}$$

and $f(\boldsymbol{Z}|\vec{\boldsymbol{\theta}})$ is defined in (7).

Step 6. Repeat Steps 3 and 5 until $k = n$ times to obtain the first Monte Carlo sample $\boldsymbol{\theta}^{(1)}$.
Step 7. Repeat Steps 2–5 M times to obtain $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}$.

Appendix 2. M step for the joint model

From (6), the conditional expectation can be partitioned into three parts according to the type of coefficients. Only the second and third parts include the unknown parameters of interest. The conditional expectation given the observable data for the second part is

$$E\left[\ln\prod_{i=1}^{n} f(\boldsymbol{\theta}_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \mathcal{D}\right] = -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n} E_i[(\boldsymbol{\theta}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu})]. \tag{A2}$$

Differentiating (A2) with respect to $\boldsymbol{\mu}$ yields

$$\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n} E_i(\boldsymbol{\theta}_i - \boldsymbol{\mu}). \tag{A3}$$

The estimates in (11) are obtained by setting (A3) equal to 0. Let $\boldsymbol{U} = \boldsymbol{\Sigma}^{-1}$. Taking the derivative of (A2) with respect to $\boldsymbol{\Sigma}^{-1}$, we obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{U}}\left[\frac{n}{2}\ln|\boldsymbol{U}| - \frac{1}{2}\mathrm{tr}\left\{\sum_{i=1}^{n} E_i\left(\boldsymbol{\theta}_i - \frac{\sum_{j=1}^{n}\boldsymbol{\theta}_j}{n}\right)\left(\boldsymbol{\theta}_i - \frac{\sum_{j=1}^{n}\boldsymbol{\theta}_j}{n}\right)'\right\}\right] \\
&= \frac{n}{2}\{2\boldsymbol{U}^{-1} - \mathrm{diag}(\boldsymbol{U}^{-1})\} - \frac{1}{2}\left[2\sum_{i=1}^{n} E_i\left(\boldsymbol{\theta}_i - \frac{\sum_{j=1}^{n}\boldsymbol{\theta}_j}{n}\right)\left(\boldsymbol{\theta}_i - \frac{\sum_{j=1}^{n}\boldsymbol{\theta}_j}{n}\right)'\right. \\
&\quad \left.-\mathrm{diag}\left\{\sum_{i=1}^{n} E_i\left(\boldsymbol{\theta}_i - \frac{\sum_{j=1}^{n}\boldsymbol{\theta}_j}{n}\right)\left(\boldsymbol{\theta}_i - \frac{\sum_{j=1}^{n}\boldsymbol{\theta}_j}{n}\right)'\right\}\right],
\end{aligned} \tag{A4}$$

where diag is a diagonal matrix and tr is the sum of diagonal elements. The estimates in (12) are obtained by setting (A4) equal to 0.

The conditional expectation given the observable data for the second part is focused on the estimation for $\boldsymbol{\gamma}$, β and $\lambda_0(u)$. The conditional expectation is

$$\begin{aligned}
E\left[\ln\prod_{i=1}^{n} f(X_i, \Delta_i \mid \boldsymbol{\theta}_i, \lambda_0, \beta, \boldsymbol{\gamma}) \mid \mathcal{D}\right] &= \sum_{i=1}^{n}\Delta_i \ln\lambda_0(X_i) \\
&+ \sum_{i=1}^{n}\{\Delta_i(\beta E_i[\pi(\boldsymbol{\theta}_i; X_i)] + \boldsymbol{\eta}_i'\boldsymbol{\gamma})\} - \sum_{i=1}^{n}\int_0^{X_i}\lambda_0(u)E_i[\mathrm{e}^{\beta\pi(\boldsymbol{\theta}_i;u)+\boldsymbol{\eta}_i'\boldsymbol{\gamma}}]\,\mathrm{d}u
\end{aligned}$$

Differentiating (A5) with respect to $\lambda_0(u)$ yields

$$\sum_{i=1}^{n}\left[\frac{\Delta_i I(X_i = u)}{\lambda_0(u)} - E_i\left[\exp\{\beta\pi(\boldsymbol{\theta}_i; u) + \boldsymbol{\eta}_i'\boldsymbol{\gamma}\}\right]Y_i(u)\right], \tag{A5}$$

where $Y_i(u) = I(X_i \geq u)$ is an indication function for X_i. The estimates in (13) are obtained by setting (A5) equal to 0. Differentiating (A5) with respect to β, we obtain

$$\sum_{i=1}^{n}\left\{\Delta_i E_i[\pi(\boldsymbol{\theta}_i; X_i)] - \sum_{j=1}^{n}\lambda_0(X_j)E_i[\pi(\boldsymbol{\theta}_i; X_j)\,\mathrm{e}^{\beta\pi(\boldsymbol{\theta}_i;X_j)+\boldsymbol{\eta}_i'\boldsymbol{\gamma}}]Y_i(X_j)\right\}, \tag{A6}$$

where $\lambda_0(u)$ is a function of β and $\boldsymbol{\gamma}$. Consequently, the estimates of β and $\boldsymbol{\gamma}$ do not have a closed form. β and $\boldsymbol{\gamma}$ are solved by FMINSEARCH in MATLAB.