Sankhyā : The Indian Journal of Statistics
2017, Volume 79-B, Part 2, pp. 190-216

Penalized Likelihood Approach for Simultaneous Analysis of Survival Time and Binary Longitudinal Outcome

Jaeun Choi
Albert Einstein College of Medicine, New York, USA
Jianwen Cai and Donglin Zeng
University of North Carolina at Chapel Hill, North Carolina, USA

Abstract

In this paper we consider simultaneous analysis of survival time and binary longitudinal outcome where random effects are introduced to account for the dependence between the two different types of outcomes due to unobserved factors and assumed to follow a Gaussian distribution with mean zero. The estimator based on maximum likelihood approach using an Expectation-Maximization algorithm is consistent and asymptotically normally distributed. However, the EM algorithm may be intensive on numerical integrations with large sample sizes and large numbers of longitudinal observations per subject. We develop a more computationally efficient estimation procedure based on a penalized likelihood obtained by Laplace approximation. Through simulation studies, we compare numerical performances on the computing time, bias, and mean squared error from the proposed penalized likelihood estimation procedure and the EM algorithm of maximum likelihood estimation. We also illustrate the proposed approach with a liver transplantation data set.

AMS (2000) subject classification. Primary 62; Secondary H99.
Keywords and phrases. Generalized linear mixed model, Laplace approximation, Penalized likelihood estimator, Random effect, Simultaneous modeling, Stratified Cox proportional hazards model

1 Introduction

In biomedical or public health research, it is common that both longitudinal outcomes over time and survival endpoint are collected for the same subject along with the subject's characteristics or risk factors. Investigators are interested in finding important variables which predict both longitudinal outcomes and survival time. For this purpose, simultaneous modeling is needed since the two different types of outcomes are correlated within the same subject. Among the existing approaches for the joint analysis of longitudinal data and survival time, modeling survival time conditional on longitudinal data or vice versa was more widely considered, compared to simultaneous modeling. Estimating the distribution of survival time given longitudinal data was studied by numerous authors, for example, Tsiatis et al. (1995), Wulfsohn and Tsiatis (1997), Henderson et al. (2000), Tsiatis and Davidian (2001), Xu and Zeger (2001a, 2001b), Song et al. (2002), Larsen (2004), Tseng et al. (2005), Song and Wang (2007), Ye et al. (2008a) and Chakraborty and Das (2010) among others. The trend of longitudinal outcomes conditional on survival time was studied by Wu and Carroll (1988), Hogan and Laird (1997), Albert and Follmann (2000, 2007) and Ding and Wang (2008) among others. On the other hand, simultaneous modeling of the longitudinal and survival data was proposed by Xu and Zeger (2001b) and Zeng and Cai (2005a), and further studied by Elashoff et al. (2007, 2008), Liu et al. (2008), Rizopoulos et al. (2008), Rizopoulos et al. (2008) and most recently Choi et al. (2013). Wang and Taylor (2001), Brown and Ibrahim (2003), Dunson and Herring (2005), Chen et al. (2009), Hu et al. (2009) and Huang et al. (2011) studied simultaneous modeling in the Bayesian perspective.

In the joint models, random effects are often incorporated to accommodate the latent dependence between survival time and longitudinal outcomes, and often assumed to be normally distributed so that we can integrate a complete data likelihood over random effects to obtain a full likelihood. The maximum likelihood approach using an Expectation-Maximization algorithm provides the estimators which are asymptotically consistent and follows an asymptotic Gaussian process (Zeng and Cai 2005a, 2005b; Choi et al. 2013). However, the EM algorithm may be intensive on computation with large sample sizes and large numbers of longitudinal observations per subject (Wulfsohn and Tsiatis 1997; Ye et al. 2008b). The numerical integrations required for a full likelihood approach can be cumbersome and intractable. One possible alternative is to use penalized likelihood approach. In this approach, the likelihood can be obtained by Laplace approximation and a penalty is used for regarding random effects as fixed effects.

The penalized quasi-likelihood (PQL) approach is the most common estimation procedure in generalized linear mixed models (GLMM). The PQL was proposed as an approximate Bayes procedure for some commonly occurring GLMM's by Laird (1978) and the PQL method exploited by Green (1987) for semiparametric regression analysis is available for inference in hierarchical models where the focus is on shrinkage estimation of the random effects (Robinson, 1991). Breslow and Clayton (1993) proposed to use the PQL with some modifications to a Laplace expansion for a GLMM in order to motivate standard estimating equations that may be solved by iterative application of normal theory procedures. Breslow and Lin (1995) and Lin and Breslow (1996) derived the general expressions for the asymptotic biases in approximate estimators of regression coefficients and variance component in the GLMMs with a single source of extraneous variation and multiple components of dispersion, respectively. The PQL also has been studied in a wide variety of GLMMs by Bartlett and Sutradhar (1999), Raudenbush et al. (2000), Dean et al. (2004), Huber et al. (2004), Localio et al. (2006), Diaz (2007), Lin (2007), Nelson and Leroux (2008), Dang et al. (2008), Qiu et al. (2008), Jang and Lim (2009), Masaoud and Stryhn (2010), Fong et al. (2010), Wood (2011) and Krivobokova et al. (2012). Furthermore, the PQL is already built in SAS GLIMMIX procedure and used for the analysis of the GLMM. On the other hand, in survival analysis, a penalized partial likelihood was proposed for multivariate frailty models by Ripatti and Palmgren (2000) and further studied for the mixed and time-varying effects survival model by Kauermann et al. (2008), for the time-varying effects recurrent event model by Yu et al. (2013) and for a joint model of recurrent events and a terminal event by Yu and Liu (2011) and Mazroui et al. (2012). In joint modeling framework of longitudinal and survival data, Ye et al. (2008b) proposed a penalized joint likelihood for a selection model and considered a continuous longitudinal process to be included as a covariate for survival time. Their penalized joint likelihood is obtained by replacing the full survival likelihood with a partial likelihood in the Laplace approximation to the full joint likelihood function, which is not equal to the actual form derived from the full joint likelihood function. However, there is no work done on the penalized likelihood approach for the simultaneous modeling of longitudinal outcomes and survival time. Furthermore, the previous study using the penalized likelihood in the selection model (Ye et al., 2008b) only considered continuous longitudinal data from a normal distribution and cannot be applied to analyzing binary or categorical longitudinal outcomes.

In this paper, we propose to use a penalized likelihood to develop a more efficient estimation procedure on computation for simultaneous modeling than the EM algorithm of the maximum likelihood approach. We consider a generalized linear mixed model for longitudinal outcome to incorporate both categorical and continuous data, although we particularly focus on binary data, and study a stratified Cox proportional hazards model for survival time. In this estimation procedure, all the parameters are estimated together at the same time. The organization of this paper is as follows. We present a simultaneous modeling for longitudinal outcomes and survival time with random effects in Section 2 and describe the proposed estimation procedure in Section 3. Numerical results from simulation studies are given in Section 4, and our proposed method is illustrated with a liver transplantation data set in Section 5. In Section 6, we discuss some further consideration.

2 Model Formulation and Notation

We use $Y(t)$ to denote the value of a longitudinal marker process at time t. Suppose $Y(t)$ is from a distribution belonging to an exponential family in order to incorporate either for continuous or categorical measurements. Let T denote survival time, and suppose that the survival time T is possibly subject to right censoring. Suppose a set of n subjects are followed over an interval $[0, \tau]$, where τ is the study end time. Denote $\boldsymbol{b}_i$, $i = 1, \ldots, n$, as a vector of subject-specific random effects of dimension d_b and $\boldsymbol{b}_i$'s are mutually independent and identically distributed from a multivariate normal with mean zero and covariance matrix $\boldsymbol{\Sigma}_b$.

Given the random effects $\boldsymbol{b}_i$, the observed covariates, and the observed outcome history till time t, we assume that the longitudinal outcome $Y_i(t)$ at time t for subject i follows a distribution from the exponential family with density,

$$\exp\left\{\frac{y_i\eta_i(t|\boldsymbol{b}_i) - B(\eta_i(t|\boldsymbol{b}_i))}{A(D_i(t;\phi))} + C(y_i, D_i(t;\phi))\right\}, \tag{2.1}$$

satisfying

$$\eta_i(t|\boldsymbol{b}_i) = g(\mu_i(t|\boldsymbol{b}_i)) = \boldsymbol{X}_i(t)\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_i(t)\boldsymbol{b}_i,$$

where $\mu_i(t|\boldsymbol{b}_i) = \mathrm{E}(Y_i(t)|\boldsymbol{b}_i) = B'(\eta_i(t|\boldsymbol{b}_i))$ and $v_i(t|\boldsymbol{b}_i) = \mathrm{Var}(Y_i(t)|\boldsymbol{b}_i) = B''(\eta_i(t|\boldsymbol{b}_i))A(D_i(t;\phi)) = v(\mu_i(t|\boldsymbol{b}_i)\ )A(D_i(t;\phi))$. Here, $g(\cdot)$ and $v(\cdot)$ are known link and variance functions, respectively, ϕ is the dispersion parameter, the functions $A(\cdot)$, $B(\cdot)$, $C(\cdot)$, and $D(\cdot)$ are known, $\boldsymbol{X}_i(t)$ and $\widetilde{\boldsymbol{X}}_i(t)$ are the row vectors of the observed covariates for subject i, and $\boldsymbol{\beta}$ is a column vector of coefficients for $\boldsymbol{X}_i(t)$. The random effect $\boldsymbol{b}_i$ is allowed to differ for different individuals. Additionally, $\boldsymbol{X}_i(t)$ and $\widetilde{\boldsymbol{X}}_i(t)$ can be completely different or share some components, and may include dummy variables for different strata. The missingness of $Y_i(t)$ is assumed to be non-informative. Note that the logistic distribution for binary $Y_i(t)$ has $A(D_i(t_j;\phi)) = 1$, $B_{ij}(\boldsymbol{\beta};\boldsymbol{b}_i) = \log\left(1 + \exp\left\{\boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\right\}\right)$, and $C(Y_{ij}; D_i(t_j;\phi)) = 0$.

Given the random effects $\boldsymbol{b}_i$, the observed covariates, and the observed survival history before time t, the conditional hazard rate function for the survival time T_i of subject i is assumed to follow a stratified multiplicative hazards model,

$$\lambda_s(t)\exp\{\widetilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\psi} \circ \boldsymbol{b}_i) + \boldsymbol{Z}_i(t)\boldsymbol{\gamma}\}, \tag{2.2}$$

where $\boldsymbol{Z}_i(t)$ and $\widetilde{\boldsymbol{Z}}_i(t)$ are the row vectors of the observed covariates and may share some components, $\boldsymbol{\psi}$ is a vector of parameters of the coefficients for random effects, $\boldsymbol{\gamma}$ is a column vector of coefficients for $\boldsymbol{Z}_i(t)$, and $\lambda_s(t)$ is the s-th stratum baseline hazard rate function so that the baseline hazard rate is allowed to vary across levels of the stratification variable. Note that $\boldsymbol{Z}_i(t)$ does not include dummy variables for strata since baseline hazard rate is stratum-specific. We assume common fixed effects and random effects across strata in both hazard and longitudinal models. However, the model may allow for possibly different covariate effects for different strata, which can be achieved by including interaction terms of the covariates with the indicator variables for the stratification variable. Subjects in different strata are assumed to be independent. In equation (2.2), for any vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ of the same dimension, $\boldsymbol{a}_1 \circ \boldsymbol{a}_2$ denotes the component-wise product. In addition, $\widetilde{\boldsymbol{X}}_i(t)$ and $\widetilde{\boldsymbol{Z}}_i(t)$ have the same dimensions as $\boldsymbol{b}_i$'s.

Under models (2.1) and (2.2), the two outcomes $Y(t)$ and T are independent conditional on the covariates and random effect. The parameter $\boldsymbol{\psi}$ in model (2.2) characterizes the dependence between the longitudinal outcomes and the survival time due to latent random effect: When the k-th component of $\boldsymbol{\psi}$ is 0 (i.e. $\psi_k = 0$), it implies that the dependence between the survival time and longitudinal responses is not due to the corresponding latent variable b_{ik}, the k-th component of $\boldsymbol{b}_i$; $\psi_k \neq 0$ implies that such dependence may be due to the corresponding latent variable b_{ik}.

Let n_i be the number of the observed longitudinal measurements for subject i, and assume that the distributions of n_i and the observation times for longitudinal measurements are independent of the parameters of interest conditional on $\boldsymbol{b}_i$ in this joint model. We also assume n_i is bounded, which is a reasonable assumption in many biomedical studies. The observed data from n subjects are $(n_i, Y_{ij}, \boldsymbol{X}_{ij}, \widetilde{\boldsymbol{X}}_{ij})$, $j{=}1,\ldots,n_i$, $i{=}1,\ldots,n$, and $(V_i, \Delta_i, S_i, \{(\boldsymbol{Z}_i(t), \widetilde{\boldsymbol{Z}}_i(t)) : t \leq V_i\})$, $i=1,\ldots,n$, where for subject i, $(Y_{ij}, \boldsymbol{X}_{ij}, \widetilde{\boldsymbol{X}}_{ij})$ is the j-th observation of $(Y_i(t), \boldsymbol{X}_i(t), \widetilde{\boldsymbol{X}}_i(t))$, C_i is the right-censoring time and independent of T_i and $Y_i(t)$ given the covariates and the random effects, $V_i = \min(T_i, C_i)$, S_i denotes the stratum, and $\Delta_i = I(T_i \leq C_i)$. We also assume that missing longitudinal obervations are not informative.

The goal of this simultaneous modeling is to estimate and make inferences on the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T, \text{Vech}(\boldsymbol{\Sigma}_b)^T, \boldsymbol{\psi}^T, \boldsymbol{\gamma}^T)^T$ and the baseline cumulative hazard functions with S strata, $\boldsymbol{\Lambda}(t) = (\Lambda_1(t), \ldots, \Lambda_S(t))^T$, where $\Lambda_s(t) = \int_0^t \lambda_s(u)du$, $s = 1, \ldots, S$. The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ are from the longitudinal model, $\boldsymbol{\psi}$ and $\boldsymbol{\gamma}$ are from the hazard model, and $\boldsymbol{\Sigma}_b$ is associated with the random effects. Vech$(\cdot)$ operator creates a column vector from a matrix by stacking the diagonal and upper-triangle elements of the matrix.

3 Maximum Penalized Likelihood Estimation

For all n subjects, we write $\boldsymbol{Y} = (\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_n^T)^T$, $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$, $\boldsymbol{V} = (V_1, \ldots, V_n)^T$, and $\boldsymbol{b} = (\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_n^T)^T$. Then, using $\eta_i(t_j|\boldsymbol{b}_i) = \boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i$ and denoting $B_{ij}(\boldsymbol{\beta}; \boldsymbol{b}_i) = B(\eta_i(t_j|\boldsymbol{b}_i))$, the likelihood function of the complete data $(\boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})$ has the form,

$$
\begin{aligned}
& L_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b}) \\
= \ & \prod_{s=1}^{S}\prod_{i=1}^{n}\left[f(\boldsymbol{Y}_i, V_i|\boldsymbol{b}_i)f(\boldsymbol{b}_i)\right]^{I(S_i=s)} = \prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{b}_i)\left(\prod_{s=1}^{S}\left[f(V_i|\boldsymbol{b}_i)\right]^{I(S_i=s)}\right) f(\boldsymbol{b}_i) \\
= \ & \prod_{i=1}^{n} \exp\left\{\sum_{j=1}^{n_i}\left[\frac{Y_{ij}(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i) - B_{ij}(\boldsymbol{\beta}; \boldsymbol{b}_i)}{A(D_i(t_j;\phi))} + C(Y_{ij}; D_i(t_j;\phi))\right]\right\} \\
& \times \left(\prod_{s=1}^{S}\left[\lambda_s(V_i)^{\Delta_i}\exp\left\{\Delta_i\left[\widetilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi}\circ\boldsymbol{b}_i) + \boldsymbol{Z}_i(V_i)\boldsymbol{\gamma}\right]\right.\right.\right. \\
& \left.\left.\left. - \int_0^{V_i}\exp\left\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi}\circ\boldsymbol{b}_i) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma}\right\}d\Lambda_s(u)\right\}\right]^{I(S_i=s)}\right) \\
& \times (2\pi)^{-d_b/2}|\boldsymbol{\Sigma}_b|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{b}_i^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i\right\},
\end{aligned}
$$

and the full likelihood function of the observed data $(\boldsymbol{Y}, \boldsymbol{V})$ for the parameter $(\boldsymbol{\theta}, \boldsymbol{\Lambda})$ is expressed as

$$
L_f(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V}) = \int_{\boldsymbol{b}} L_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})d\boldsymbol{b}. \tag{3.1}
$$

The primary difficulty in implementing this full likelihood inference lies in the integrations needed to evaluate the complete data likelihood $L_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})$.

In the EM algorithm of maximum likelihood approach, the random effect $\boldsymbol{b}_i$ is considered as missing data for $i = 1, \ldots, n$. Thus, the M-step solves the conditional score equations from complete data log-likelihood given observations, where the conditional expectation is evaluated in the E-step. The procedure involves iterating between the two steps until convergence is achieved. In the E-step calculating the conditional expectations of some known functions of $\boldsymbol{b}_i$ needed in the next M-step, a numerical approximation method such as the Gauss-Hermite Quadrature is required for the integration with the posterior probability of random effects. When sample size (n), the number of observations per subject (n_i), and the number of parameters to be estimated are large, the task involving the integration in the E-step is intensive with the iterations in this algorithm and potentially slow down the convergence. Therefore, to make the simultaneous modeling more practical, we aim to develop an algorithm which relieves the computational burden.

Our proposed estimation method is to calculate the maximum penalized likelihood estimates for $(\boldsymbol{\theta}, \boldsymbol{\Lambda}(t))$ over a set in which $\boldsymbol{\theta}$ is in a bounded set and $\Lambda_s(t)$ of $\boldsymbol{\Lambda}(t)$ belongs to a space consisting of all the increasing functions with $\Lambda_s(0) = 0$, $s = 1, \ldots, S$. We let each $\Lambda_s(t)$ of $\boldsymbol{\Lambda}(t)$, $s = 1, \ldots, S$, be an increasing and right-continuous step function with jumps only at the observed failure times belonging to stratum s. The penalized likelihood is obtained by Laplace approximation, and the proposed approach is expected to be less intensive in computation in the sense that it imposes the penalty for considering the random effect as the fixed effect in the likelihood and therefore no calculation for integrating the likelihood over random effects is needed.

3.1. Laplace Approximation. The full likelihood (3.1) can be written as

$$L_f(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V}) = (2\pi)^{-nd_b/2}|\boldsymbol{\Sigma}_b|^{-n/2} \int_{\boldsymbol{b}} \exp\left\{\sum_{i=1}^{n}\left[l_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) - \frac{1}{2}\boldsymbol{b}_i^T \boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i\right]\right\} d\boldsymbol{b}, \tag{3.2}$$

where the logarithm of the conditional joint density given an unobserved random effect $\boldsymbol{b}_i$ is

$$\begin{aligned} l_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) = & \sum_{j=1}^{n_i}\left[\frac{Y_{ij}(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i) - B_{ij}(\boldsymbol{\beta}; \boldsymbol{b}_i)}{A(D_i(t_j; \phi))} + C(Y_{ij}; D_i(t_j; \phi))\right] \\ & + \sum_{s=1}^{S} I(S_i = s)\Big[\Delta_i \log(\lambda_s(V_i)) + \Delta_i[\widetilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi} \circ \boldsymbol{b}_i) + \boldsymbol{Z}_i(V_i)\boldsymbol{\gamma}] \\ & - \int_0^{V_i} \exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi} \circ \boldsymbol{b}_i) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma}\} d\Lambda_s(u)\Big]. \end{aligned} \tag{3.3}$$

In Eq. 3.2, define

$$-\boldsymbol{\kappa}(\boldsymbol{b}) = \sum_{i=1}^{n}\left[l_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) - \frac{1}{2}\boldsymbol{b}_i^T \boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i\right] = \sum_{i=1}^{n}\left[-\boldsymbol{\kappa}_i(\boldsymbol{b}_i)\right] \tag{3.4}$$

and apply Laplace's approximation as following,

$$-\boldsymbol{\kappa}_i(\boldsymbol{b}_i) \approx -\boldsymbol{\kappa}_i(\widetilde{\boldsymbol{b}}_i) - \frac{1}{2}(\boldsymbol{b} - \widetilde{\boldsymbol{b}}_i)^T \boldsymbol{\kappa}_i''(\widetilde{\boldsymbol{b}}_i)(\boldsymbol{b}_i - \widetilde{\boldsymbol{b}}_i),$$

where $\boldsymbol{\kappa}'$ and $\boldsymbol{\kappa}''$ denote the d_b vector and $d_b \times d_b$ dimensional matrix of first- and second-order partial derivatives of $\boldsymbol{\kappa}$ with respect to $\boldsymbol{b}$ and $\widetilde{\boldsymbol{b}}$ denotes the solution to $\boldsymbol{\kappa}'(\boldsymbol{b}) = 0$ that minimizes $\boldsymbol{\kappa}(\boldsymbol{b})$. Then, the full likelihood function (3.2) can be approximated as followings,

$$\begin{aligned}
& L_f(\boldsymbol{\theta}, \boldsymbol{\Lambda}; \boldsymbol{Y}, \boldsymbol{V}) \\
= \; & (2\pi)^{-nd_b/2} |\boldsymbol{\Sigma}_b|^{-n/2} \int_{\boldsymbol{b}} \exp\{-\boldsymbol{\kappa}(\boldsymbol{b})\} d\boldsymbol{b} = \prod_{i=1}^{n} \left[(2\pi)^{-d_b/2} |\boldsymbol{\Sigma}_b|^{-1/2} \int_{\boldsymbol{b}} \exp\{-\boldsymbol{\kappa}_i(\boldsymbol{b}_i)\} d\boldsymbol{b} \right] \\
\approx \; & \prod_{i=1}^{n} \left[(2\pi)^{-d_b/2} |\boldsymbol{\Sigma}_b|^{-1/2} \int_{\boldsymbol{b}} \exp\left\{ -\boldsymbol{\kappa}_i(\widetilde{\boldsymbol{b}}_i) - \frac{1}{2} (\boldsymbol{b}_i - \widetilde{\boldsymbol{b}}_i)^T \boldsymbol{\kappa}_i''(\widetilde{\boldsymbol{b}}_i) (\boldsymbol{b}_i - \widetilde{\boldsymbol{b}}_i) \right\} d\boldsymbol{b} \right] \\
= \; & \prod_{i=1}^{n} \left[|\boldsymbol{\Sigma}_b|^{-1/2} \exp\{-\boldsymbol{\kappa}_i(\widetilde{\boldsymbol{b}}_i)\} |\boldsymbol{\kappa}_i''(\widetilde{\boldsymbol{b}}_i)|^{-1/2} (2\pi)^{-d_b/2} |\boldsymbol{\kappa}_i''(\widetilde{\boldsymbol{b}}_i)|^{1/2} \right. \\
& \left. \times \int_{\boldsymbol{b}} \exp\left\{ -\frac{1}{2} (\boldsymbol{b}_i - \widetilde{\boldsymbol{b}}_i)^T \boldsymbol{\kappa}_i''(\widetilde{\boldsymbol{b}}_i) (\boldsymbol{b}_i - \widetilde{\boldsymbol{b}}_i) \right\} d\boldsymbol{b} \right] \\
= \; & \prod_{i=1}^{n} \left[|\boldsymbol{\Sigma}_b|^{-1/2} \exp\{-\boldsymbol{\kappa}_i(\widetilde{\boldsymbol{b}}_i)\} |\boldsymbol{\kappa}_i''(\widetilde{\boldsymbol{b}}_i)|^{-1/2} \right] \\
= \; & |\boldsymbol{\Sigma}_b|^{-n/2} \exp\left\{ \sum_{i=1}^{n} \left[-\boldsymbol{\kappa}_i(\widetilde{\boldsymbol{b}}_i) - \frac{1}{2} \log |\boldsymbol{\kappa}_i''(\widetilde{\boldsymbol{b}}_i)| \right] \right\}. \qquad (3.5)
\end{aligned}$$

Note that, from Eq. 3.4,

$$\begin{aligned}
\boldsymbol{\kappa}_i(\widetilde{\boldsymbol{b}}_i) &= -\widetilde{l}_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) + \frac{1}{2} \widetilde{\boldsymbol{b}}_i^T \boldsymbol{\Sigma}_b^{-1} \widetilde{\boldsymbol{b}}_i, \\
\boldsymbol{\kappa}_i'(\widetilde{\boldsymbol{b}}_i) &= -\widetilde{l}'_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) + \boldsymbol{\Sigma}_b^{-1} \widetilde{\boldsymbol{b}}_i, \\
\boldsymbol{\kappa}_i''(\widetilde{\boldsymbol{b}}_i) &= -\widetilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) + \boldsymbol{\Sigma}_b^{-1}, \qquad (3.6)
\end{aligned}$$

where $\widetilde{l}_{i|b_i}(\boldsymbol{\theta}, \Lambda_s)$ is Eq. 3.3 evaluated at $\widetilde{\boldsymbol{b}}_i$, and $\widetilde{l}'_{i|b_i}(\boldsymbol{\theta}, \Lambda_s)$ and $\widetilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s)$ are the first and second derivatives of Eq. 3.3 with respect to $\boldsymbol{b}_i$ evaluated at $\widetilde{\boldsymbol{b}}_i$. Then, the first order Laplace approximation (3.5) to the full likelihood becomes

$$\begin{aligned}
& \exp\left\{ \sum_{i=1}^{n} \left[-\frac{1}{2} \log |\boldsymbol{\Sigma}_b| - \left(-\widetilde{l}_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) + \frac{1}{2} \widetilde{\boldsymbol{b}}_i^T \boldsymbol{\Sigma}_b^{-1} \widetilde{\boldsymbol{b}}_i \right) - \frac{1}{2} \log \left| -\widetilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) + \boldsymbol{\Sigma}_b^{-1} \right| \right] \right\} \\
&= \exp\left\{ \sum_{i=1}^{n} \left[-\frac{1}{2} \log \left| \boldsymbol{I}_{d_b} - \boldsymbol{\Sigma}_b \widetilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) \right| + \widetilde{l}_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) - \frac{1}{2} \widetilde{\boldsymbol{b}}_i^T \boldsymbol{\Sigma}_b^{-1} \widetilde{\boldsymbol{b}}_i \right] \right\} \\
&= \exp\left\{ \sum_{i=1}^{n} \widetilde{l}_{fi}(\boldsymbol{\theta}, \Lambda_s) \right\} = \exp\left\{ \widetilde{l}_f(\boldsymbol{\theta}, \Lambda_s) \right\},
\end{aligned}$$

where $\widetilde{l}_f(\boldsymbol{\theta}, \Lambda_s)$ is the first order Laplace approximation to the full log-likelihood function $l_f(\boldsymbol{\theta}, \Lambda_s) = \log L_f(\boldsymbol{\theta}, \Lambda_s)$. That is,

$$\widetilde{l}_f(\boldsymbol{\theta}, \boldsymbol{\Lambda}) = \sum_{i=1}^{n} \widetilde{l}_{f_i}(\boldsymbol{\theta}, \Lambda_s) = \sum_{i=1}^{n} \left[-\frac{1}{2} \log \left| \boldsymbol{I}_{d_b} - \boldsymbol{\Sigma}_b \widetilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) \right| + \left(\widetilde{l}_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) - \frac{1}{2} \widetilde{\boldsymbol{b}}_i^T \boldsymbol{\Sigma}_b^{-1} \widetilde{\boldsymbol{b}}_i \right) \right], \qquad (3.7)$$

where $\tilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) = l''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s; \widetilde{\boldsymbol{b}}_i)$ and $\tilde{l}_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) = l_{i|b_i}(\boldsymbol{\theta}, \Lambda_s; \widetilde{\boldsymbol{b}}_i)$, and the first and second derivatives of $l_{i|b_i}(\boldsymbol{\theta}, \Lambda_s; \widetilde{\boldsymbol{b}}_i)$ in Eq. 3.3 with respect to $\boldsymbol{b}_i$ are

$$\begin{aligned} l'_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) &= \sum_{j=1}^{n_i} \left[\frac{Y_{ij} \widetilde{\boldsymbol{X}}_{ij}}{A(D_i(t_j;\phi))} - \frac{B'_{ij}(\boldsymbol{\beta}; \boldsymbol{b}_i)}{A(D_i(t_j;\phi))} \right] \\ &+ \sum_{s=1}^{S} I(S_i = s) \Big[\Delta_i(\widetilde{\boldsymbol{Z}}_i(V_i) \circ \boldsymbol{\psi}^T) - \int_0^{V_i} \exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi} \circ \boldsymbol{b}_i) \\ &+ \boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}(\widetilde{\boldsymbol{Z}}_i(u) \circ \boldsymbol{\psi}^T) d\Lambda_s(u) \Big] \end{aligned}$$

and

$$\begin{aligned} l''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) &= -\sum_{j=1}^{n_i} \frac{B''_{ij}(\boldsymbol{\beta}; \boldsymbol{b}_i)}{A(D_i(t_j;\phi))} - \sum_{s=1}^{S} I(S_i = s) \int_0^{V_i} \exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi} \circ \boldsymbol{b}_i) \\ &+ \boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}(\widetilde{\boldsymbol{Z}}_i^T(u) \circ \boldsymbol{\psi})^{\otimes 2} d\Lambda_s(u), \end{aligned} \tag{3.8}$$

where $B'_{ij}(\boldsymbol{\beta}; \boldsymbol{b}_i)$ and $B''_{ij}(\boldsymbol{\beta}; \boldsymbol{b}_i)$ are the first and second derivatives of $B_{ij}(\boldsymbol{\beta}; \boldsymbol{b}_i)$ with respect to $\boldsymbol{b}_i$. $-\frac{1}{2}\widetilde{\boldsymbol{b}}_i^T \boldsymbol{\Sigma}_b^{-1} \widetilde{\boldsymbol{b}}_i$ in Eq. 3.7 is the penalty term for regarding the random effects as fixed effects in the likelihood. The first order Laplace approximated log-likelihood Eq. 3.7 can be used for estimation and one may call it a penalized log-likelihood.

However, the estimation using Eq. 3.7 is still computationally intensive due to the calculation of Eq. 3.8, and therefore we further approximate (3.8) by obtaining its expected value. By some algebra, the first term of Eq. 3.8, $-\sum_{j=1}^{n_i} B''_{ij}(\boldsymbol{\beta}; \boldsymbol{b}_i)/A(D_i(t_j;\phi))$, can be expressed as $\widetilde{\boldsymbol{X}}_i^T \boldsymbol{W}_i \widetilde{\boldsymbol{X}}_i$, where $\boldsymbol{W}_i$ is the $n_i \times n_i$ diagonal matrix with $w_{ij} = \left[A(D_i(t_j;\phi)) g'(\mu_{ij}^b)\right]^{-1}$, $g(\cdot)$ is a canonical link function, $\mu_{ij}^b = \mathrm{E}\left(Y_{ij}|\boldsymbol{b}_i\right)$, $g'(\mu_{ij}^b)$ is the derivative of $g(\mu_{ij}^b)$ with respect to μ_{ij}^b, and $\widetilde{\boldsymbol{X}}_i = (\widetilde{\boldsymbol{X}}_{i1}^T, \ldots, \widetilde{\boldsymbol{X}}_{in_i}^T)^T$. The generalized linear model (GLM) iterative weights $\boldsymbol{W}_i$ (i.e. w_{ij}) vary slowly or not at all as the function of the mean, and hence, by taking an expectation over $\boldsymbol{W}_i$, $\mathrm{E}\left[\widetilde{\boldsymbol{X}}_i^T \boldsymbol{W}_i \widetilde{\boldsymbol{X}}_i\right]$ becomes a constant. Also, in the second term of Eq. 3.8,

$$\begin{aligned} &\int_0^{V_i} \exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi} \circ \boldsymbol{b}_i) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}(\widetilde{\boldsymbol{Z}}_i^T(u) \circ \boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(u) \circ \boldsymbol{\psi}^T) d\Lambda_s(u) \\ &= \int (\widetilde{\boldsymbol{Z}}_i^T(u) \circ \boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(u) \circ \boldsymbol{\psi}^T) I(V_i \geq u) \exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi} \circ \boldsymbol{b}_i) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma}\} \lambda_s(u) du \\ &= \int (\widetilde{\boldsymbol{Z}}_i^T(u) \circ \boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(u) \circ \boldsymbol{\psi}^T) I(T_i \geq u) I(C_i \geq u) \frac{1}{S_{sT}(u)} dF_{sT}(u) \end{aligned}$$

$$= \left(\widetilde{\boldsymbol{Z}}_i^T(V_i) \circ \boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(V_i) \circ \boldsymbol{\psi}^T\right) \Delta_i,$$

where $S_{sT}(\cdot)$ and $F_{sT}(\cdot)$ are the survival function and cumulative density function, respectively, for survival time of the s-stratum. By taking an expectation over Δ_i, $\mathrm{E}\left[(\widetilde{\boldsymbol{Z}}_i^T(V_i) \circ \boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(V_i) \circ \boldsymbol{\psi}^T)\Delta_i\right]$ becomes a constant, including only $\boldsymbol{\psi}$. Therefore, we obtain the expected value of Eq. 3.8,

$$\mathrm{E}\left[l''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s)\right] = \mathrm{E}\left[\widetilde{\boldsymbol{X}}_i^T \boldsymbol{W}_i \widetilde{\boldsymbol{X}}_i\right] - \sum_{s=1}^{S} I(S_i = s)\, \mathrm{E}\left[(\widetilde{\boldsymbol{Z}}_i^T(V_i) \circ \boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(V_i) \circ \boldsymbol{\psi}^T)\Delta_i\right], \tag{3.9}$$

which becomes a constant, not involving $\boldsymbol{b}_i$. By using Eq. 3.9 instead of $\widetilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s)$, Eq. 3.7 can be approximated to the following, defined as $l_P(\boldsymbol{\theta}, \boldsymbol{\Lambda})$,

$$\begin{aligned} l_P(\boldsymbol{\theta}, \boldsymbol{\Lambda}) = \sum_{i=1}^{n} \Bigg[& -\frac{1}{2} \log \Bigg| \boldsymbol{I}_{d_b} - \boldsymbol{\Sigma}_b \Bigg(\mathrm{E}\left[\widetilde{\boldsymbol{X}}_i^T \boldsymbol{W}_i \widetilde{\boldsymbol{X}}_i\right] - \sum_{s=1}^{S} I(S_i = s) \\ & \times \mathrm{E}\left[(\widetilde{\boldsymbol{Z}}_i^T(V_i) \circ \boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(V_i) \circ \boldsymbol{\psi}^T)\Delta_i\right]\Bigg)\Bigg| \\ & + \left(\widetilde{l}_{i|b_i}(\boldsymbol{\theta}, \Lambda_s) - \frac{1}{2}\widetilde{\boldsymbol{b}}_i^T \boldsymbol{\Sigma}_b^{-1} \widetilde{\boldsymbol{b}}_i\right) \Bigg]. \end{aligned} \tag{3.10}$$

Since the first term in Eq. 3.10 which corresponds to $\left|\boldsymbol{I}_{d_b} - \boldsymbol{\Sigma}_b \mathrm{E}\left[l''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s)\right]\right|$ includes only $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\psi}$, it contributes to the estimating equations of $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\psi}$ and is ignored to obtain the estimating equations of $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$. We choose $\boldsymbol{\theta}$ to maximize $l_p(\boldsymbol{\theta}, \boldsymbol{\Lambda})$ in Eq. 3.10. That is, $(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{b}})$ jointly maximize Eq. 3.10.

3.2. Implementation. We conduct the Newton-Rapshon method for estimating equations to obtain $\widetilde{\boldsymbol{b}}$ and $\widehat{\boldsymbol{\theta}}$. The procedure involves iterating between the following two steps until convergence is achieved: at the k-th iteration,

Step1 : Conduct one-step Newton-Rapshon iteration to obtain the solution $\widetilde{\boldsymbol{b}}$ of $\boldsymbol{\kappa}'(\boldsymbol{b}) = 0$. The $(k+1)$-th estimate is $\widetilde{\boldsymbol{b}}^{(k+1)} = \widetilde{\boldsymbol{b}}^{(k)} - \left[\boldsymbol{\kappa}''(\widetilde{\boldsymbol{b}}^{(k)})\right]^{-1}\left[\boldsymbol{\kappa}'(\widetilde{\boldsymbol{b}}^{(k)})\right]^T$, where $\widetilde{\boldsymbol{b}}^{(k)} = \widetilde{\boldsymbol{b}}^{(k)}(\widehat{\boldsymbol{\theta}}^{(k-1)})$, $\boldsymbol{\kappa}'(\boldsymbol{b}) = (\boldsymbol{\kappa}'_1(\boldsymbol{b}_1)^T, \ldots, \boldsymbol{\kappa}'_n(\boldsymbol{b}_n)^T)^T$ and $\boldsymbol{\kappa}''(\boldsymbol{b}) = (\boldsymbol{\kappa}''_1(\boldsymbol{b}_1)^T, \ldots, \boldsymbol{\kappa}''_n(\boldsymbol{b}_n)^T)^T$, and the functions $\boldsymbol{\kappa}'_i(\boldsymbol{b}_i)$ and $\boldsymbol{\kappa}''_i(\boldsymbol{b}_i)$, $i = 1, \ldots, n$, are given in Eq. 3.6.

Step2 : By one-step Newton-Rapshon iteration, the $(k+1)$-th estimate is calculated as $\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} - \left[S'_P(\widehat{\boldsymbol{\theta}}^{(k)T})\right]^{-1}\left[S_P(\widehat{\boldsymbol{\theta}}^{(k)T})\right]^T$, where $S_P(\boldsymbol{\theta})$ is the score equation for $\boldsymbol{\theta}$ from Eq. 3.10 and $S'_P(\boldsymbol{\theta})$ is the first derivative of

$S_P(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. With $(\widehat{\boldsymbol{\theta}}^{(k+1)}, \widetilde{\boldsymbol{b}}^{(k+1)})$, the $(k+1)$-th Breslow-type estimate of the baseline cumulative hazard for the s-th stratum is obtained as an empirical function which has jumps only at the observed failure time,

$$\begin{aligned}\Lambda_s^{(k+1)}(t) &= \Lambda_s^{(k+1)}(t; \widehat{\boldsymbol{\theta}}^{(k+1)}, \widetilde{\boldsymbol{b}}^{(k+1)}) \\ &= \sum_{i:V_i \le t} \frac{\Delta_i I(S_i = s)}{\sum_{l:V_l \ge V_i} \exp\left\{\widetilde{\boldsymbol{Z}}_l(V_i)(\widehat{\boldsymbol{\psi}}^{(k+1)} \circ \widetilde{\boldsymbol{b}}_l^{(k+1)}) + \boldsymbol{Z}_l(V_i)\widehat{\boldsymbol{\gamma}}^{(k+1)}\right\} I(S_l = s)}. \quad (3.11)\end{aligned}$$

For variance estimation of $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Lambda}}(t))$, we adopt the observed information matrix by Louis (1982) and conduct the Expectation step used in the maximum likelihood approach with the estimates by the penalized likelihood method. For the numerical calculation of the observed information matrix, we consider $\Lambda_s\{V_i\}$, the jump size of $\Lambda_s(t)$ at V_i belonging to stratum s for which $\Delta_i = 1$, instead of $\lambda_s(V_i)$. That is, $\boldsymbol{\Lambda}\{\cdot\} = (\boldsymbol{\Lambda}_1^T\{\cdot\}, \ldots, \boldsymbol{\Lambda}_S^T\{\cdot\})^T$ with $\boldsymbol{\Lambda}_s\{\cdot\} = (\Lambda\{T_{s1}\}, \ldots, \Lambda\{T_{sm_s}\})^T$ for m_s failure times among n_s subjects $(0 \le m_s \le n_s)$ of the s-th stratum, $s = 1, \ldots, S$. Then, by Louis (1982),

$$\begin{aligned}\mathrm{I}(\boldsymbol{\theta}, \boldsymbol{\Lambda}\{\cdot\}; \boldsymbol{Y}, \boldsymbol{V}) &= \mathrm{E}_{b|Y,V}[B_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}\{\cdot\}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})|\boldsymbol{Y}, \boldsymbol{V}] \\ &- \mathrm{E}_{b|Y,V}[U_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}\{\cdot\}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})U_c^T(\boldsymbol{\theta}, \boldsymbol{\Lambda}\{\cdot\}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})|\boldsymbol{Y}, \boldsymbol{V}] \\ &+ \mathrm{E}_{b|Y,V}[U_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}\{\cdot\}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})]\, \mathrm{E}_{b|Y,V}[U_c^T(\boldsymbol{\theta}, \boldsymbol{\Lambda}\{\cdot\}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})],\end{aligned}$$

where $U_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}\{\cdot\}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})$ and $B_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}\{\cdot\}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})$ are the first derivative vector and the negative of the second derivative matrix for the complete data log-likelihood $l_c(\boldsymbol{\theta}, \boldsymbol{\Lambda}\{\cdot\}; \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})$, respectively. For subject i with $S_i = s$, given observations and the penalized likelihood estimate $(\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s)$, we calculate the following conditional expectation of a known function $q(\boldsymbol{b}_i)$ needed in the observed information matrix,

$$\begin{aligned}E\left[q(\boldsymbol{b}_i)|\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s\right] &= \frac{\int_{\boldsymbol{b}_i} q(\boldsymbol{b}_i) f\left(\boldsymbol{Y}_i, V_i|\boldsymbol{b}_i, \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s\right) f\left(\boldsymbol{b}_i|\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s\right) d\boldsymbol{b}_i}{\int_{\boldsymbol{b}_i} f\left(\boldsymbol{Y}_i, V_i|\boldsymbol{b}_i, \widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s\right) f\left(\boldsymbol{b}_i|\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s\right) d\boldsymbol{b}_i} \\ &= \frac{\int_{\boldsymbol{z}_G} q\left(R(\boldsymbol{z}_G)\right) K(\boldsymbol{z}_G) \exp\{-\boldsymbol{z}_G^T \boldsymbol{z}_G\} d\boldsymbol{z}_G}{\int_{\boldsymbol{z}_G} K(\boldsymbol{z}_G) \exp\{-\boldsymbol{z}_G^T \boldsymbol{z}_G\} d\boldsymbol{z}_G}, \quad (3.12)\end{aligned}$$

where $\boldsymbol{z}_G$ follows a multivariate Gaussian distribution with mean zero, $\boldsymbol{z}_G = R^{-1}(\boldsymbol{b}_i)$, $K(\boldsymbol{z}_G) = \exp\{\boldsymbol{z}_G^T \boldsymbol{z}_G\} f\left(\boldsymbol{Y}_i, V_i|R(\boldsymbol{z}_G), \boldsymbol{\theta}^{(k)}, \Lambda_s^{(k)}\right) f\left(R(\boldsymbol{z}_G)|\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}_s\right)$, and Gauss-Hermite Quadrature numerical approximation is used for the calculation of integration. Note that, in Eq. 3.12, the functions of $R(\cdot)$ and $K(\cdot)$ have different expressions for different longitudinal distributions.

The proposed penalized likelihood approach for simultaneous modeling can be applied to all generalized linear mixed models of longitudinal outcomes. Since we focus on binary longitudinal data, next we provide the expressions of the penalized log-likelihood and relevant equations for binary longitudinal outcomes with survival time. On the other hand, the corresponding expressions for Normal longitudinal distribution as an example of continuous longitudinal outcome are given in Appendix.

3.2.1. Binary longitudinal data and survival time. Logistic distribution has $A(D_i(t_j;\phi)) = 1$, $B_{ij}(\boldsymbol{\beta};\boldsymbol{b}_i) = \log\left(1+\exp\left\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\right\}\right)$, and $C(Y_{ij};D_i(t_j;\phi)) = 0$ in Eq. 2.1. Thus, the $\boldsymbol{\kappa}_i'(\boldsymbol{b}_i)$ and $\boldsymbol{\kappa}_i''(\boldsymbol{b}_i)$, $i = 1,\ldots,n$, in Step 1 are

$$\begin{aligned}
&\boldsymbol{\kappa}_i'(\boldsymbol{b}_i)\\
&= -\Bigg[\sum_{j=1}^{n_i}\left(Y_{ij} - \frac{\exp\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\}}{1+\exp\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\}}\right)\widetilde{\boldsymbol{X}}_{ij}\\
&\quad + \sum_{s=1}^{S} I(S_i=s)\left(\Delta_i(\widetilde{\boldsymbol{Z}}_i(V_i)\circ\boldsymbol{\psi}^T) - \int_0^{V_i}\exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi}\circ\boldsymbol{b}_i)+\boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}(\widetilde{\boldsymbol{Z}}_i(u)\circ\boldsymbol{\psi}^T)d\Lambda_s(u)\right)\Bigg]\\
&\quad + \boldsymbol{b}_i^T\boldsymbol{\Sigma}_b^{-1}
\end{aligned}$$

and

$$\begin{aligned}
&\boldsymbol{\kappa}_i''(\boldsymbol{b}_i)\\
&= -\Bigg[\sum_{j=1}^{n_i}\left(-\frac{\exp\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\}}{(1+\exp\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\})^2}\right)\widetilde{\boldsymbol{X}}_{ij}^T\widetilde{\boldsymbol{X}}_{ij}\\
&\quad + \sum_{s=1}^{S} I(S_i=s)\left(-\int_0^{V_i}\exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi}\circ\boldsymbol{b}_i)+\boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}(\widetilde{\boldsymbol{Z}}_i^T(u)\circ\boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(u)\circ\boldsymbol{\psi}^T)d\Lambda_s(u)\right)\Bigg]\\
&\quad + \boldsymbol{\Sigma}_b^{-1}.
\end{aligned}$$

In Step 2, the penalized log-likelihood (3.10) has the following form for binary longitudinal outcomes and survival time,

$$\begin{aligned}
&l_P(\boldsymbol{\theta},\boldsymbol{\Lambda})\\
&= \sum_{i=1}^{n}\Bigg[-\frac{1}{2}\log\left|\boldsymbol{I}_{d_b} - \boldsymbol{\Sigma}_b\left(\mathrm{E}\left[\widetilde{\boldsymbol{X}}_i^T\boldsymbol{W}_i\widetilde{\boldsymbol{X}}_i\right] - \sum_{s=1}^{S} I(S_i=s)\,\mathrm{E}\left[(\widetilde{\boldsymbol{Z}}_i^T(V_i)\circ\boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(V_i)\circ\boldsymbol{\psi}^T)\Delta_i\right]\right)\right|\\
&\quad + \sum_{j=1}^{n_i}\left[Y_{ij}(\boldsymbol{X}_{ij}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}_{ij}\widetilde{\boldsymbol{b}}_i) - \log\left(1+\exp\left\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}_{ij}\widetilde{\boldsymbol{b}}_i\right\}\right)\right]\\
&\quad + \sum_{s=1}^{S} I(S_i=s)\Bigg[\Delta_i\log(\lambda_s(V_i)) + \Delta_i[\widetilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi}\circ\widetilde{\boldsymbol{b}}_i)+\boldsymbol{Z}_i(V_i)\boldsymbol{\gamma}]\\
&\qquad\qquad - \int_0^{V_i}\exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi}\circ\widetilde{\boldsymbol{b}}_i)+\boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}d\Lambda_s(u)\Bigg]\\
&\quad - \frac{1}{2}\widetilde{\boldsymbol{b}}_i^T\boldsymbol{\Sigma}_b^{-1}\widetilde{\boldsymbol{b}}_i\Bigg)\Bigg].
\end{aligned} \tag{3.13}$$

For the two expected values, $\mathrm{E}\left[\widetilde{\boldsymbol{X}_i}^T\boldsymbol{W}_i\widetilde{\boldsymbol{X}_i}\right]$ and $\mathrm{E}\left[(\widetilde{\boldsymbol{Z}}_i^T(V_i)\circ\boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(V_i)\circ\boldsymbol{\psi}^T)\Delta_i\right]$ in Eq. 3.13, we evaluate $\boldsymbol{W}_i = \mu_{ij}^b(1-\mu_{ij}^b)\boldsymbol{I}_i$, with $\mu_{ij}^b = \exp\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+$

$\widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\}/\big(1+\exp\{\boldsymbol{X}_{ij}\boldsymbol{\beta}+\widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\}\big)$ and a n_i-dimensional identity matrix $\boldsymbol{I}_i$, and $\Delta_i = \int_0^{V_i} \exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi}\circ\boldsymbol{b}_i)+\boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}d\Lambda_s(u)$ at the estimates of parameters and cumulative hazards at the previous iteration and the estimates of random effects from Step 1 at the current iteration. That is, we use

$$\widehat{\boldsymbol{W}}_i = \frac{\exp\{\boldsymbol{X}_{ij}\widehat{\boldsymbol{\beta}}+\widetilde{\boldsymbol{X}}_{ij}\widetilde{\boldsymbol{b}}_i\}}{(1+\exp\{\boldsymbol{X}_{ij}\widehat{\boldsymbol{\beta}}+\widetilde{\boldsymbol{X}}_{ij}\widetilde{\boldsymbol{b}}_i\})^2}\boldsymbol{I}_i \quad \text{and} \quad \widehat{\Delta}_i = \exp\{\widetilde{\boldsymbol{Z}}_i(V_i)(\widehat{\boldsymbol{\psi}}\circ\widetilde{\boldsymbol{b}}_i)+\boldsymbol{Z}_i(V_i)\widehat{\boldsymbol{\gamma}}\}\widehat{\Lambda}_s(V_i), \quad (3.14)$$

respectively. $S_P(\boldsymbol{\theta})$ and $S'_P(\boldsymbol{\theta})$ are the first and second derivatives of Eq. 3.13 with respect to $\boldsymbol{\theta}$. The Breslow-type estimator of the baseline cumulative hazard for the s-th stratum has the same expression given in Eq. 3.11 for all different longitudinal distributions.

4. Simulation Studies

In this section, through simulation studies, we compare numerical performances on the computing time, bias, and mean squared error (MSE) of the penalized likelihood method and the EM algorithm used in maximum likelihood estimation for the simultaneous modeling of binary longitudinal outcomes and survival time with a random intercept.

We assume that Y_{ij} is a binary outcome following

$$P(Y_{ij}=y_{ij}|b_i) = \exp\big\{y_{ij}\eta_{ij} - \log(1+\exp\{\eta_{ij}\})\big\}, \qquad y_{ij}=0,\ 1, \quad (4.1)$$

with $\eta_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3ij} + b_i$ for $j = 1, \ldots, n_i$, and

$$h(t|b_i) = \lambda(t)\exp\{\psi b_i + \boldsymbol{Z}_i(t)\boldsymbol{\gamma}\} = \lambda(t)\exp\{\psi b_i + \gamma_1 Z_{1i} + \gamma_2 Z_{2i}\}, \quad (4.2)$$

where $b_i \sim N(0, \sigma_b^2)$, $X_{1i} \equiv Z_{1i}$ are generated from a Bernoulli distribution with success probability being 0.5, and $X_{2i} \equiv Z_{2i}$ are simulated from the uniform distribution between 0 and 1. They are included in both hazard and longitudinal models. There is one additional covariate denoted as X_{3ij}, the time at measurement, included in the longitudinal model. For this time at which longitudinal data are observed, we suppose the longitudinal data are collected for every unit of time over the follow-up ranging 0 through 2.4 and simulate 4 different units which are 0.3, 0.1, 0.05 and 0.03 producing the average numbers of longitudinal observations (n_i) per subject to be 4, 8, 15 and 25, respectively. Then, the longitudinal data are generated from the Bernoulli distribution with the success probability $P(Y_{ij}=1|b_i)$ given in Eq. 4.1 at X_{3ij}. To generate the survival time, we first generate u_i from uniform (0,1) distribution. For a given hazard function λ, the survival time is then generated by $t_i = -\log(u_i) \times \exp\{-(\psi b_i + \gamma_1 Z_{1i} + \gamma_2 Z_{2i})\}/\lambda$. Censoring time is generated from the uniform distribution between 0.4 and 2.4 so that the censoring proportion is around 25∼35%. The observed survival time is obtained by the minimum of the generated survival and censoring times. For the comparison of the estimated baseline cumulative hazards over simulations, we consider three time points: 0.9, 1.4, and 1.9, which correspond to the quartiles of the true survival distribution. The three time points are not the only distinct survival times but are selected to report the estimated cumulative hazard function at these points.

We consider $\psi = -0.1$ indicating negative dependency between longitudinal process and survival time model. The parameters in the longitudinal and hazard models are chosen as $\beta_0 = -1$, $\beta_1 = 1$, $\beta_2 = -0.5$, $\beta_3 = -0.2$, $\psi = -0.1$, $\gamma_1 = -0.1$, $\gamma_2 = 0.1$, and $\lambda(t) = 1$, and the variance of random effects, σ_b^2, is chosen as 0.5. Different sample sizes (n=200, 400) are simulated with 1000 replications.

Tables 1 and 2 report the simulation results of maximum likelihood estimation (MLE) using the logarithm of Eq. 3.1 and maximum penalized likelihood estimation (MPLE) using Eq. 3.13 for $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_b^2, \psi, \boldsymbol{\gamma}^T)^T$ and baseline cumulative hazards at the given three time points in the simultaneous modeling of binary longitudinal outcomes and survival time with sample sizes of 200 and 400, respectively.

In Tables 1 and 2, "True" gives the true values of parameters; the middle 6 columns under "MLE" and the right 6 columns under "MPLE" are the results of the maximum likelihood estimates from the EM algorithm and the proposed maximum penalized likelihood estimates, respectively; the averages of the estimates are in "Est."; the averages of the bias estimates of the parameter estimates subtracted from true values are in "Bias"; the sample standard deviations from 1000 simulations are reported in "SSD"; "ESE" is the average of 1000 standard error estimates based on the observed information matrix; "MSE" gives the mean squared error calculated by adding the squared bias and the squared sample standard deviations; "CP" is the coverage proportion of 95% nominal confidence intervals based on the estimated standard error "ESE". Note that "ESE" under "MPLE" is based on the observed information matrix obtained by maximum likelihood approach using the maximum penalized likelihood estimates. Satterthwaite method is used for the coverage proportion of σ_b^2.

From Tables 1 and 2, we can see that the bias of the proposed MPLE is small for most cases although it is bigger than the MLE's, but overall the bias of the MPLE decreases for large n_i and large n like the MLE's does. On

Table 1: Summary of simulation results of maximum likelihood estimation (MLE) and maximum penalized likelihood estimation (MPLE) in the simultaneous modeling of binary longitudinal outcomes and survival time (n=200)

			MLE						MPLE					
n_i	Par.	True	Est.	Bias	SSD	ESE	MSE	CP	Est.	Bias	SSD	ESE	MSE	CP
4	β_0	−1.0	−1.013	−.013	.247	.243	.061	.948	−.932	.068	.226	.228	.056	.948
	β_1	1.0	1.004	.004	.203	.205	.041	.953	.928	−.072	.186	.190	.040	.933
	β_2	−.5	−.479	.021	.358	.349	.128	.944	−.444	.056	.330	.327	.112	.940
	β_3	−.2	−.201	−.001	.207	.209	.043	.960	−.191	.009	.193	.203	.037	.969
	γ_1	−.1	−.096	.004	.169	.174	.029	.964	−.098	.002	.170	.177	.029	.965
	γ_2	.1	.098	−.002	.301	.302	.091	.948	.100	.000	.301	.314	.091	.954
	ψ	−.1	−.111	−.011	.316	.316	.100	.980	−.131	−.031	.404	.472	.164	.989
	σ_b^2	.5	.516	.016	.204	.217	.042	.949	.360	−.140	.138	.172	.038	.997
	$\Lambda(\ .9)$	.9	.917	.017	.191	.184	.061	.945	.932	.032	.196	.200	.052	.952
	$\Lambda(1.4)$	1.4	1.446	.046	.302	.297	.043	.945	1.471	.071	.314	.328	.040	.959
	$\Lambda(1.9)$	1.9	1.977	.077	.440	.442	.134	.958	2.013	.113	.460	.494	.122	.966
8	β_0	−1.0	−.994	.006	.201	.198	.040	.946	−.927	.073	.187	.188	.040	.928
	β_1	1.0	.990	−.010	.165	.168	.027	.954	.927	−.073	.154	.158	.029	.933
	β_2	−.5	−.504	−.004	.296	.288	.087	.948	−.471	.029	.277	.273	.078	.945
	β_3	−.2	−.206	−.006	.156	.156	.024	.953	−.199	.001	.148	.152	.022	.956
	γ_1	−.1	−.102	−.002	.179	.173	.032	.939	−.103	−.003	.179	.173	.032	.939
	γ_2	.1	.114	.014	.288	.300	.083	.962	.116	.016	.288	.305	.083	.965
	ψ	−.1	−.112	−.012	.230	.232	.053	.976	−.114	−.014	.266	.264	.071	.974
	σ_b^2	.5	.502	.002	.138	.142	.019	.961	.402	−.098	.106	.115	.021	.998
	$\Lambda(\ .9)$	.9	.906	.006	.176	.181	.040	.959	.913	.013	.178	.188	.035	.964
	$\Lambda(1.4)$	1.4	1.421	.021	.282	.289	.028	.958	1.432	.032	.287	.299	.025	.961
	$\Lambda(1.9)$	1.9	1.949	.049	.413	.428	.090	.965	1.964	.064	.419	.441	.081	.968
15	β_0	−1.0	−.989	.011	.168	.166	.028	.946	−.938	.062	.159	.159	.029	.931
	β_1	1.0	.997	−.003	.145	.142	.021	.943	.948	−.052	.137	.136	.021	.933
	β_2	−.5	−.508	−.008	.248	.245	.062	.950	−.481	.019	.235	.235	.055	.952
	β_3	−.2	−.201	−.001	.109	.113	.012	.958	−.198	.002	.105	.112	.011	.961
	γ_1	−.1	−.083	.017	.167	.172	.028	.955	−.084	.016	.168	.172	.028	.954
	γ_2	.1	.114	.014	.301	.299	.091	.951	.118	.018	.301	.301	.091	.953
	ψ	−.1	−.098	.002	.185	.190	.034	.956	−.090	.010	.200	.200	.040	.950
	σ_b^2	.5	.487	−.013	.098	.100	.010	.966	.424	−.076	.082	.085	.012	.944
	$\Lambda(\ .9)$	.9	.900	.000	.182	.179	.028	.955	.902	.002	.183	.182	.025	.957
	$\Lambda(1.4)$	1.4	1.400	.000	.297	.283	.021	.948	1.403	.003	.297	.287	.019	.953
	$\Lambda(1.9)$	1.9	1.914	.014	.431	.416	.062	.949	1.919	.019	.434	.422	.055	.952
25	β_0	−1.0	−.992	.008	.149	.150	.022	.949	−.951	.049	.142	.145	.023	.938
	β_1	1.0	.997	−.003	.132	.130	.017	.947	.957	−.043	.125	.125	.017	.934
	β_2	−.5	−.501	−.001	.223	.225	.050	.950	−.481	.019	.209	.217	.044	.964
	β_3	−.2	−.203	−.003	.086	.090	.007	.954	−.200	.000	.084	.089	.007	.960
	γ_1	−.1	−.101	−.001	.177	.172	.031	.941	−.098	.002	.174	.172	.030	.941
	γ_2	.1	.100	.000	.310	.299	.096	.943	.112	.012	.305	.300	.093	.947
	ψ	−.1	−.091	.009	.177	.169	.031	.947	−.084	.016	.185	.173	.034	.938
	σ_b^2	.5	.490	−.010	.083	.084	.007	.956	.446	−.054	.073	.073	.008	.931
	$\Lambda(\ .9)$	.9	.913	.013	.188	.182	.022	.941	.910	.010	.186	.183	.020	.944
	$\Lambda(1.4)$	1.4	1.428	.028	.305	.288	.018	.937	1.421	.021	.305	.289	.016	.932
	$\Lambda(1.9)$	1.9	1.958	.058	.454	.426	.053	.947	1.946	.046	.450	.426	.046	.946

Table 2: Summary of simulation results of maximum likelihood estimation (MLE) and maximum penalized likelihood estimation (MPLE) in the simultaneous modeling of binary longitudinal outcomes and survival time (n=400)

			MLE						MPLE					
n_i	Par.	True	Est.	Bias	SSD	ESE	MSE	CP	Est.	Bias	SSD	ESE	MSE	CP
4	β_0	−1.0	−1.003	−.003	.172	.170	.030	.949	−.924	.076	.158	.159	.031	.932
	β_1	1.0	1.004	.004	.145	.143	.021	.946	.929	−.071	.132	.133	.023	.916
	β_2	−.5	− .503	−.003	.248	.244	.061	.941	−.465	.035	.229	.229	.053	.944
	β_3	−.2	−.196	.004	.147	.147	.022	.952	−.186	.014	.137	.142	.019	.959
	γ_1	−.1	−.100	.000	.124	.121	.015	.936	−.101	−.001	.124	.121	.015	.938
	γ_2	.1	.114	.014	.213	.210	.046	.950	.116	.016	.213	.213	.046	.957
	ψ	−.1	−.096	.004	.205	.203	.042	.963	−.112	−.012	.262	.270	.069	.978
	σ_b^2	.5	.496	−.004	.136	.150	.018	.961	.349	−.151	.095	.119	.032	1.000
	$\Lambda(.9)$	.9	.898	−.002	.129	.126	.030	.948	.905	.005	.131	.131	.025	.950
	$\Lambda(1.4)$	1.4	1.406	.006	.208	.201	.021	.949	1.418	.018	.213	.208	.018	.953
	$\Lambda(1.9)$	1.9	1.916	.016	.302	.295	.062	.944	1.934	.034	.311	.306	.053	.951
8	β_0	−1.0	−.998	.002	.147	.139	.022	.938	−.930	.070	.138	.132	.024	.903
	β_1	1.0	1.001	.001	.120	.118	.014	.946	.937	−.063	.112	.111	.017	.904
	β_2	−.5	−.509	−.009	.205	.203	.042	.951	−.476	.024	.192	.192	.038	.944
	β_3	−.2	−.195	.005	.112	.109	.013	.939	−.188	.012	.107	.107	.012	.942
	γ_1	−.1	−.098	.002	.125	.120	.016	.953	−.098	.002	.125	.120	.016	.953
	γ_2	.1	.106	.006	.207	.209	.043	.948	.107	.007	.206	.210	.043	.950
	ψ	−.1	−.104	−.004	.155	.156	.024	.967	−.103	−.003	.178	.176	.032	.965
	σ_b^2	.5	.498	−.002	.093	.099	.009	.963	.401	−.099	.072	.080	.015	.937
	$\Lambda(.9)$	.9	.902	.002	.127	.126	.022	.943	.905	.005	.127	.128	.019	.942
	$\Lambda(1.4)$	1.4	1.413	.013	.206	.200	.015	.946	1.417	.017	.206	.203	.013	.949
	$\Lambda(1.9)$	1.9	1.924	.024	.299	.292	.043	.946	1.930	.030	.300	.296	.038	.949
15	β_0	−1.0	−.996	.004	.117	.117	.014	.946	−.944	.056	.110	.112	.015	.923
	β_1	1.0	1.000	.000	.099	.101	.010	.954	.949	−.051	.095	.096	.012	.925
	β_2	−.5	−.504	−.004	.172	.173	.030	.955	−.477	.023	.163	.166	.027	.955
	β_3	−.2	−.199	.001	.081	.080	.006	.943	−.196	.004	.078	.079	.006	.947
	γ_1	−.1	−.100	.000	.118	.120	.014	.962	−.100	.000	.118	.120	.014	.964
	γ_2	.1	.108	.008	.208	.209	.043	.955	.110	.010	.208	.209	.043	.954
	ψ	−.1	−.099	.001	.128	.130	.016	.959	−.089	.011	.140	.136	.020	.953
	σ_b^2	.5	.495	−.005	.070	.071	.005	.959	.431	−.069	.058	.060	.008	.886
	$\Lambda(.9)$	.9	.899	−.001	.126	.126	.014	.955	.901	.001	.126	.127	.012	.955
	$\Lambda(1.4)$	1.4	1.405	.005	.198	.199	.010	.951	1.408	.008	.198	.200	.009	.953
	$\Lambda(1.9)$	1.9	1.917	.017	.289	.290	.030	.948	1.915	.015	.286	.292	.027	.951
25	β_0	−1.0	−1.004	−.004	.104	.106	.011	.949	−.961	.039	.099	.102	.011	.939
	β_1	1.0	.998	−.002	.093	.092	.009	.945	.954	−.046	.087	.088	.010	.921
	β_2	−.5	−.486	.014	.158	.159	.025	.943	−.467	.033	.150	.153	.024	.944
	β_3	−.2	−.198	.002	.063	.063	.004	.962	−.197	.003	.063	.063	.004	.964
	γ_1	−.1	−.101	−.001	.118	.120	.014	.961	−.100	.000	.118	.120	.014	.965
	γ_2	.1	.099	−.001	.217	.208	.047	.941	.103	.003	.213	.209	.046	.948
	ψ	−.1	−.096	.004	.115	.117	.013	.959	−.081	.019	.120	.120	.015	.957
	σ_b^2	.5	.493	−.007	.058	.059	.003	.964	.446	−.054	.051	.052	.006	.872
	$\Lambda(.9)$	.9	.911	.011	.135	.127	.011	.931	.910	.010	.132	.127	.010	.940
	$\Lambda(1.4)$	1.4	1.419	.019	.213	.200	.009	.933	1.414	.014	.207	.200	.008	.940
	$\Lambda(1.9)$	1.9	1.925	.025	.299	.291	.026	.947	1.925	.025	.294	.292	.023	.949

the other hand, the estimate of σ_b^2 by the MPLE is smaller than its true value showing the biggest bias, but it is improved soon being close to the true value as n_i increases. It is already known that the penalized quasi-likelihood (PQL) used for GLMMs tends to underestimate somewhat the variance components when applied to clustered binary data but the situation improves rapidly for binomial observations having denominators greater than one (Breslow and Clayton, 1993). The result from our simulation studies conforms this fact. For both MLE and MPLE, the ESE calculated from the observed information matrix is close to the SSD from the 1000 estimates. They decrease over n_i except for the baseline cumulative hazards estimates, and they also decrease as n increases. The MPLE has smaller SSD and ESE than the MLE for most cases. As for MSE representing both bias and SSD together, the MSE of the MPLE appears to be smaller than or close to the MLE's. The MSEs from both MLE and MPLE decrease as n_i and n increase. The 95% CPs are close to 0.95 except those for ψ of both MLE and MPLE with small n_i(=4 and 8) at small n(=200), for σ_b^2 of the MPLE with small n_i(=4 and 8) at small n(=200), and for σ_b^2 of the MPLE with the very small or large n_i(=4, 15 and 25) at large n(=400). For both MLE and MPLE, the CP of ψ is recovered for large n_i and large n. Thus, with small n_i and small n, the test for ψ is conservative, which strengthens the test results when rejecting the null ($\psi = 0$), and the type I error becomes closer to the nominal level as n_i and n increase. While the high CP of σ_b^2 by the MLE is improved for both large n_i and large n, the CP of σ_b^2 by the MPLE appears to be improved for large n_i at small n and small n_i at large n. With the small n(=200) of Table 1, the high CPs of σ_b^2 by the MPLE at small n_i(=4 and 8) are recovered at large n_i(=15 and 25). On the other hand, with the relatively small n_i(=8), the high CP of σ_b^2 by the MPLE shown at small n(=200) in Table 1 is improved for large n(=400) in Table 1. In additional simulation studies conducted with larger n(=800) whose results are not provided in this paper, the high CPs of σ_b^2 by the MPLE shown at the smallest n_i(=4) of n=200 and 400 in both Tables 1 and 2 actually reached 95% nominal level for n=800. However, the CP of σ_b^2 by the MPLE at large n_i(=15 and 25) appears to decrease over n, which seems because σ_b^2 of the MPLE has the biggest bias among the estimates of parameters and the reduction of the bias is very slow as n increases while the ESE of σ_b^2 decreases rapidly. The reason σ_b^2 of the MPLE has the biggest bias is that $\widetilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s)$ of $|\boldsymbol{I}_{d_b} - \boldsymbol{\Sigma}_b \widetilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s)|$ in Eq. 3.7, which contributes to the estimation of σ_b^2, is further approximated as shown in Eq. 3.10. Note that the approximation of $\widetilde{l}''_{i|b_i}(\boldsymbol{\theta}, \Lambda_s)$ also affects the estimation of ψ but the influence is trivial since only part of the corresponding term in Eq. 3.10 is related to ψ.

Figure 1 shows the ratios of mean squared errors (MSEs) of the proposed MPLE to the MLE with sample sizes of 200 and 400 for the parameters of predictors in longitudinal and hazard models. This figure confirms the results provided in Tables 1 and 2 in that all plots indicate the ratios of MSEs are close to 1 which implies the proposed MPLE provides the MSE close to the MLE's.

Table 3 provides user times of the MLE and the MPLE to the MLE for running 1000 simulated data sets with sample sizes of 200 and 400 in the simultaneous modeling of binary longitudinal outcomes and survival time. The proposed MPLE appears to be more computationally efficient reducing

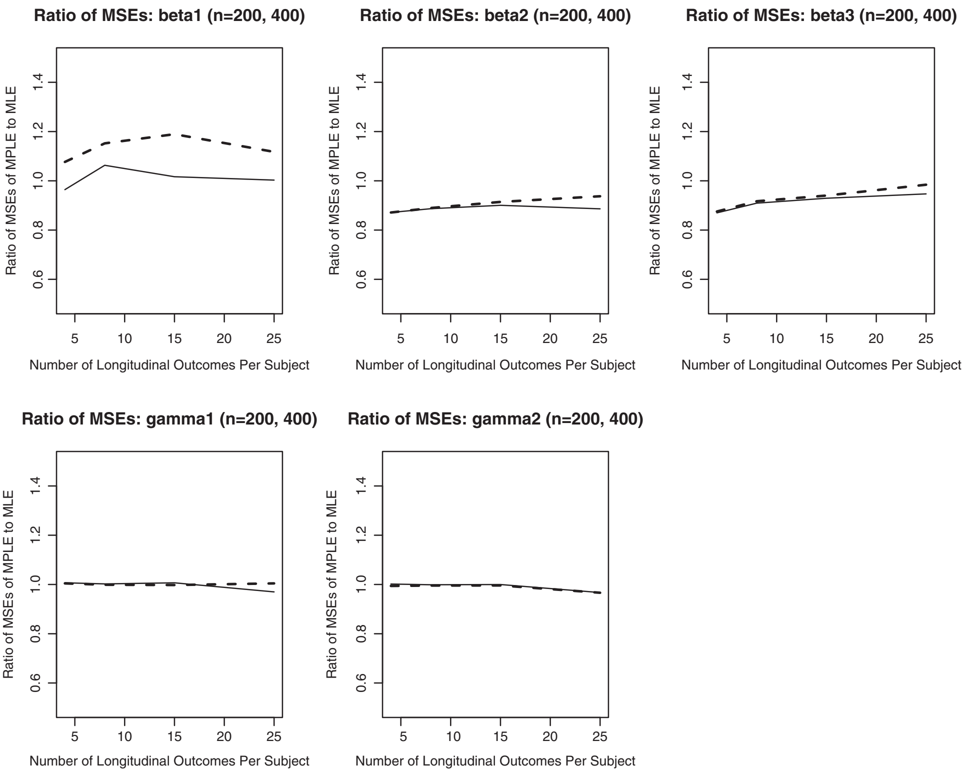

Figure 1: Plot of ratios of mean squared errors (MSEs) of maximum penalized likelihood estimator (MPLE) to maximum likelihood estimator (MLE) for parameters of predictors in longitudinal and hazard models (*solid line*: n=200, *dashed line*: n=400)

Table 3: Summary of user times of maximum likelihood estimation (MLE) and maximum penalized likelihood estimation (MPLE) for running 1000 simulated data sets in the simultaneous modeling of binary longitudinal outcomes and survival time

		User times (min)	
n	n_i	MLE	MPLE
200	4	39.93	10.56
	8	36.45	11.22
	15	54.15	15.72
	25	97.95	23.86
400	4	103.99	31.12
	8	113.03	30.80
	15	193.40	47.92
	25	344.14	94.46

about 70–75% of the computing time of the MLE over all different numbers of longitudinal outcomes per subject and sample sizes.

5. Application

The liver transplantation study of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) was a 7–year prospective study of 1563 candidates for liver transplantation at three major transplant centers. Among the 1563 candidates, 582 received the transplantation for the first time and these patients were evaluated at four months, one year, and annually afterwards till five years after their liver transplantation. At each evaluation, they were given questionnaires asking about their life satisfaction. By the end of the study, 76 patients were deceased. One goal of this study was to investigate whether factors such as patients' characteristic and disease history affect both quality of life and the risk of death. The life satisfaction evaluation and survival time can be correlated within the same subject. Appropriate analysis should take this dependency into consideration.

The longitudinal outcome is a binary measurement of the patients' quality of life. The original outcome of the patients' QoL is based on the question "Overall, how satisfied are you with health at the present time?". The response score ranges from 1 ("completely satisfied") to 7 ("completely dissatisfied"). We dichotomize this score into 0 ("satisfied"; QoL score<4) and

1 ("dissatisfied"; QoL score$\geq$4) and use this binary outcome in our analysis. The advantage for dichotomizing the measure is to reduce possible measurement error. There are 582 patients with 1382 complete post-transplantation QoL scores and the number of observations for each patient ranges from 1 to 6. The censoring rate is 87%.

We are interested in studying which variables, including gender, race, marriage status, age at liver transplantation, body mass index(BMI), and history of ascites, bone disease, cholangitis, and edema, predict the life satisfaction or the risk of death or both. Time at measurement is also included as a covariate for longitudinal outcomes. A random intercept b_0 for the dependence between the life satisfaction and the risk of death is included in both models, and assumed to follow a normal distribution with mean zero.

In Table 4, we compare the estimates and the estimated standard errors of the maximum likelihood estimation (MLE) and maximum penalized likelihood estimation (MPLE). From both "Est." and "ESE" columns, we see that MLE and MPLE provide similar estimates and estimated standard errors for the parameters of interest in longitudinal QoL and hazards models. On the other hand, the parameters of σ_b^2 and ψ, which denote the variance of random effects and the coefficient of random effects characterizing the dependence between longitudinal QoL and survival processes, respectively, have different estimates and estimated standard errors between the MLE and MPLE. This discrepancy of the MPLE from the MLE may be a numerical issue due to the small cluster size with the average of 1.93. In addition, the MPLE provides slightly bigger estimated standard errors than the MLE for the parameters in the longitudinal model while it appears in the reverse direction in hazards model. This also may be a numerical issue due to the small number of longitudinal outcomes per subject since the estimation in the longitudinal model is directly affected by the individual cluster size while the estimation in hazards model is not.

Comparing the computing time spent on producing the results in Table 4, the proposed MPLE took only a sixth of the time the MLE did (62.83 and 361.78 seconds for MPLE and MLE respectively). This analysis result indicates that, even for the small cluster size, the proposed MPLE provides the similar results to those of the MLE for the parameters of interest taking less computing time than the MLE. In the studies with larger number of longitudinal outcomes per subject, the results of the MPLE are expected to be close to those of the MLE for all parameters with much better efficiency on calculation.

Table 4: Analyses results from maximum likelihood estimation (MLE) and maximum penalized likelihood estimation (MPLE) for the Quality of Life and survival time for the liver transplantation data

Parameter		Est. MLE	Est. MPLE	ESE MLE	ESE MPLE
*Q*oL satisfaction longitudinal model					
Intercept	β_0	–3.964	–2.664	.950	.663
Center (ref=center1)					
– center2	β_1	1.432	1.003	.337	.237
– center3	β_2	1.341	.919	.328	.229
Gender (ref=male): female	β_3	.146	.131	.256	.184
Race (ref=non–Caucasian): Caucasian	β_4	.328	.230	.341	.248
Marriage (ref=single): married	β_5	–.336	–.265	.301	.216
ASC (ref=never had ascites): ever had ascites	β_6	–.463	–.325	.316	.228
BD (ref=never had bone disease): ever had bone disease	β_7	.457	.330	.411	.295
CHO (ref=never had cholangitis): ever had cholangitis	β_8	–.161	–.116	.421	.300
EDE (ref=never had edema): ever had edema	β_9	.046	.009	.305	.219
Age at liver transplantation	β_{10}	–.006	–.004	.011	.008
BMI	β_{11}	.063	.041	.025	.018
Time at measurement (years)	β_{12}	–.039	–.049	.080	.068
Variance of random effects	σ_b^2	3.545	1.195	.817	.314
*H*azards model					
Random effect coefficient	ψ	.425	.741	.123	.227
Center (ref=center1)					
– center2	γ_1	.626	.561	.336	.299
– center3	γ_2	.565	.483	.337	.289
Gender (ref=male): female	γ_3	.470	.468	.267	.248
Race (ref=non-Caucasian): Caucasian	γ_4	–.659	–.674	.308	.288
Marriage (ref=single): married	γ_5	.001	–.017	.314	.290
ASC (ref=never had ascites): ever had ascites	γ_6	–.215	–.238	.324	.300
BD (ref=never had bone disease): ever had bone disease	γ_7	–.230	–.245	.459	.425
CHO (ref=never had cholangitis): ever had cholangitis	γ_8	.213	.174	.407	.369
EDE (ref=never had edema): ever had edema	γ_9	.136	.139	.320	.299
Age at liver transplantation	γ_{10}	.028	.025	.012	.009
BMI	γ_{11}	.020	.014	.024	.017

6. Concluding Remarks

In this paper, we have developed a more computationally efficient estimation procedure adopting a penalized likelihood based on Laplace approximation for the simultaneous modeling of survival time and longitudinal outcome, particularly focusing on binary longitudinal data. Our proposed penalized likelihood estimation method is an effort to reduce the intensity on computation still providing the similar estimates to those by the EM algorithm of the maximum likelihood approach. Simulation studies indicated that the penalized likelihood approach performs as well as the EM algorithm of maximum likelihood approach, but only requires a fraction of the computing time. We also illustrated this comparison with the NIDDK liver transplantation study data.

In the simultaneous modeling considered in this paper, we assumed random effects to follow a Gaussian distribution with mean zero. However, it is unclear whether the normality assumption is truly satisfied in practice. Future work can include developing an approach to reduce computational intensity efficiently through the penalized likelihood approach for relaxing the normality assumption of random effects in the simultaneous modeling.

For longitudinal measurements, we assumed that their missingness is not informative. However, if the missing longitudinal data is non-ignorable, the MLE and MPLE could be biased. Non-ignorable missing data in the simultaneous modeling will be future research.

7. Supplementary Materials

R functions to perform the MLE and MPLE methods are provided and the usage of the functions is illustrated in Electronic Supplementary Materials.

Acknowledgments. This research was partially supported by the National Institutes of Health grants P01 CA142538, UL1 RR025747 and R01 ES021900.

Appendix

A.1. Implementation for continuous longitudinal data and survival time

Continuous longitudinal outcomes following a normal distribution has $A(D_i(t_j;\phi)) = \sigma_y^2$, $B_{ij}(\boldsymbol{\beta};\boldsymbol{b}_i) = \big(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\big)^2/2$, and $C(Y_{ij};D_i(t_j;\phi)) = -\big(y^2/\sigma_y^2 + \log(\sigma_y^2) + \log(2\pi)\big)/2$ in Eq. 2.1, where σ_y^2 is the variance of longitudinal outcomes given $\boldsymbol{b}_i$. Then, the $\boldsymbol{\kappa}_i'(\boldsymbol{b}_i)$ and $\boldsymbol{\kappa}_i''(\boldsymbol{b}_i)$, $i = 1,\ldots,n$, used in Step 1 of Section 3.2 are

$$\begin{aligned}
&\boldsymbol{\kappa}_i'(\boldsymbol{b}_i)\\
&= -\Bigg[\sum_{j=1}^{n_i}\frac{1}{\sigma_y^2}\big(Y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta} - \widetilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i\big)\widetilde{\boldsymbol{X}}_{ij} + \sum_{s=1}^{S} I(S_i = s)\Bigg(\Delta_i(\widetilde{\boldsymbol{Z}}_i(V_i)\circ\boldsymbol{\psi}^T)\\
&\qquad - \int_0^{V_i}\exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi}\circ\boldsymbol{b}_i) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}(\widetilde{\boldsymbol{Z}}_i(u)\circ\boldsymbol{\psi}^T)d\Lambda_s(u)\Bigg)\Bigg]\\
&\quad + \boldsymbol{b}_i^T\boldsymbol{\Sigma}_b^{-1}
\end{aligned}$$

and

$$\begin{aligned}
&\boldsymbol{\kappa}_i''(\boldsymbol{b}_i)\\
&= -\Bigg[\sum_{j=1}^{n_i}\Big(-\frac{1}{\sigma_y^2}\Big)\widetilde{\boldsymbol{X}}_{ij}^T\widetilde{\boldsymbol{X}}_{ij} + \sum_{s=1}^{S} I(S_i = s)\Big(-\int_0^{V_i} \exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi}\circ\boldsymbol{b}_i)\\
&\qquad +\boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}(\widetilde{\boldsymbol{Z}}_i^T(u)\circ\boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(u)\circ\boldsymbol{\psi}^T)d\Lambda_s(u)\Big)\Bigg]\\
&\quad + \boldsymbol{\Sigma}_b^{-1}.
\end{aligned}$$

In Step2, the penalized log-likelihood (3.10) has the following form for continuous longitudinal outcomes from a normal distribution and survival time,

$$\begin{aligned}
&l_P(\boldsymbol{\theta},\boldsymbol{\Lambda})\\
&=\sum_{i=1}^{n}\Bigg[-\frac{1}{2}\log\bigg|\boldsymbol{I}_{d_b} - \boldsymbol{\Sigma}_b\Big(\mathrm{E}\left[\widetilde{\boldsymbol{X}}_i^T\boldsymbol{W}_i\widetilde{\boldsymbol{X}}_i\right] - \sum_{s=1}^{S} I(S_i = s)\,\mathrm{E}\left[(\widetilde{\boldsymbol{Z}}_i^T(V_i)\circ\boldsymbol{\psi})\right.\\
&\qquad \left.\times(\widetilde{\boldsymbol{Z}}_i(V_i)\circ\boldsymbol{\psi}^T)\Delta_i\right]\Big)\bigg| - \sum_{j=1}^{n_i}\frac{1}{2\sigma_y^2}\big(Y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta} - \widetilde{\boldsymbol{X}}_{ij}\widetilde{\boldsymbol{b}}_i\big)^2 - \frac{n_i}{2}\log\left(2\pi\sigma_y^2\right)\\
&\qquad + \sum_{s=1}^{S} I(S_i = s)\Bigg[\Delta_i\log(\lambda_s(V_i)) + \Delta_i[\widetilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi}\circ\widetilde{\boldsymbol{b}}_i) + \boldsymbol{Z}_i(V_i)\boldsymbol{\gamma}]\\
&\qquad\qquad - \int_0^{V_i}\exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi}\circ\widetilde{\boldsymbol{b}}_i) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}d\Lambda_s(u)\Bigg]\\
&\qquad - \frac{1}{2}\widetilde{\boldsymbol{b}}_i^T\boldsymbol{\Sigma}_b^{-1}\widetilde{\boldsymbol{b}}_i\Big)\Bigg].
\end{aligned} \tag{A.1}$$

For the two expected values, $\mathrm{E}\left[\widetilde{\boldsymbol{X}}_i^T\boldsymbol{W}_i\widetilde{\boldsymbol{X}}_i\right]$ and $\mathrm{E}\left[(\widetilde{\boldsymbol{Z}}_i^T(V_i)\circ\boldsymbol{\psi})(\widetilde{\boldsymbol{Z}}_i(V_i)\circ\boldsymbol{\psi}^T)\Delta_i\right]$ in Eq. A.1, we evaluate $\boldsymbol{W}_i = (\sigma_y^2)^{-1}\boldsymbol{I}_i$, with a n_i-dimensional identity matrix $\boldsymbol{I}_i$, and $\Delta_i = \int_0^{V_i}\exp\{\widetilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi}\circ\boldsymbol{b}_i) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma}\}d\Lambda_s(u)$ at the estimates of parameters and cumulative hazards at the previous iteration and the estimates of random effects from Step 1 at the current iteration. That is, we use

$$\widehat{\boldsymbol{W}}_i = \frac{1}{\widehat{\sigma}_y^2}\boldsymbol{I}_i \quad \text{and} \quad \widehat{\Delta}_i = \exp\{\widetilde{\boldsymbol{Z}}_i(V_i)(\widehat{\boldsymbol{\psi}}\circ\widetilde{\boldsymbol{b}}_i) + \boldsymbol{Z}_i(V_i)\widehat{\gamma}\}\widehat{\Lambda}_s(V_i),$$

respectively. $S_P(\boldsymbol{\theta})$ is obtained by differentiating (A.1) with respect to $\boldsymbol{\theta}$, and $S_P'(\boldsymbol{\theta})$ is the derivative of $S_P(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

References

ALBERT, P.S. and FOLLMANN, D.A. (2000). Modeling Repeated Count Data Subject to Informative Dropout. *Biometrics* **56**, 667–677.

ALBERT, P. S. and FOLLMANN, D. A. (2007). Random Effects and Latent Processes Approaches for Analyzing Binary Longitudinal Data with Missingness: a Comparison of Approaches Using Opiate Clinical Trial Data. *Statistical Methods in Medical Research* **16**, 417–439.

BARTLETT, R. F. and SUTRADHAR, B. C. (1999). On estimating equations for parameters in generalized linear mixed models with application to binary data. *Environmetrics* **10**, 769–784.

BRESLOW, N. E. and CLAYTON, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of American Statistical Association* **88**, 9–25.

BRESLOW, N. E. and LIN, X. (1995). Bias Correction in Generalized Linear Mixed Models with a Single Component of Dispersion. *Biometrika* **82**, 81–91.

BROWN, E. R. and IBRAHIM, J. G. (2003). A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data. *Biometrics* **59**, 221–228.

CHAKRABORTY, A. and DAS, K. (2010). Inferences for joint modelling of repeated ordinal scores and time to event data. *Comput Math Methods Med* **11**, 281–295.

CHEN, W., GHOSH, D., RAGHUNATHAN, T. E. and SARGENT, D. J. (2009). Bayesian Variable Selection with Joint Modeling of Categorical and Survival Outcomes: An Application to Individualizing Chemotherapy Treatment in Advanced Colorectal Cancer. *Biometrics* **65**, 1030–1040.

CHOI, J., CAI, J., ZENG, D. and OLSHAN, A.F (2013) Joint Analysis of Survival Time and Longitudinal Categorical Outcomes. *Statistics in Biosciences* doi: 10.1007/s12561-013-9091-z.

DANG, Q. Y., MAZUMDAR, S. and HOUCK, P. R. (2008). Sample size and power calculations based on generalized linear mixed models with correlated binary outcomes. *Computer Methods and Programs in Biomedicine* **91**, 122–127.

DEAN, C. B., UGARTE, M. D. and MILITINO, A. F. (2004). Penalized Qusi-likelihood with Spatially Correlated Data. *Computational Statistics and Data Analysis* **45**, 235–248.

DIAZ, R. E. (2007). Comparison of PQL and Laplace 6 Estimates of Hierarchical Linear Models When Comparing Groups of Small Incident Rates in Cluster Randomised Trials. *Computational Statistics and Data Analysis* **51**, 2871–2888.

DING, J. and WANG, J. L. (2008). Modeling Longitudinal Data with Nonparametric Multiplicative Random Effects Jointly with Survival Data. *Biometrics* **64**, 546–556.

DUNSON, D. B. and HERRING, A. H. (2005). Bayesian latent variable models for mixed discrete outcomes. *Biostatistics* **6**, 11–25.

ELASHOFF, R. M., LI, G. and LI, N. (2007). An Approach to Joint Analysis of Longitudinal Measurements and Competing Risks Failure Time Data. *Statistics in Medicine* **26**, 2813–2835.

ELASHOFF, R. M., LI, G. and LI, N. (2008). A Joint Model for Longitudinal Measurements and Survival Data in the Presence of Multiple Failure Types. *Biometrics* **64**, 762–771.

FONG, Y., RUE, H. and WAKEFIELD, J. (2010). Bayesian Inference for Generalized Linear Mixed Models. *Biostatistics* **11**, 397–412.

GREEN, P. J. (1987). Penalized Likelihood for General Semi-parametric Regression Models. *International Statistical Review* **55**, 245–259.

HENDERSON, R., DIGGLE, P. and DOBSON, A. (2000). Joint Modeling of Longitudinal Measurements and Event Time Data. *Biometrics* **4**, 465–480.

HOGAN, J. and LAIRD, N. (1997). Mixture Models for the Joint Distribution of Repeated Measures and Event Times. *Statistics in Medicine* **16**, 239–257.

HU, W., LI, G. and LI, N. (2009). A Bayesian Approach to Joint Analysis of Longitudinal Measurements and Competing Risks Failure Time Data. *Journal of Computational and Graphical Statistics* **28**, 1601–1619.

HUANG, X., LI, G., ELASHFOFF, R. M. and PAN, J. (2011). A General Joint Model for Longitudinal Measurements and Competing Risks Survival Data with Heterogeneous Random Effects. *Journal of Computational and Graphical Statistics* **17**, 80–100.

HUBER, P., RONCHETTI, E. and VICTORIA-FESER, M. P. (2004). Estimation of generalized linear latent variable models. *Journal Royal Statistical Society B* **66**, 893–908.

JANG, W. and LIM, J. (2009). A Numerical Study of PQL Estimation Biases in Generalized Linear Mixed Models Under Heterogeneity of Random Effects. *Communications in Statistics-Simulation and Computation* **38**, 692–702.

KAUERMANN, G., XU, R. and VAIDA, F. (2008). Stacked Laplace-EM Algorithm for Duration Models with Time-varying and Random Effects. *Computational Statistics and Data Analysis* **52**, 2514–2528.

KRIVOBOKOVA, T., CRAINICEANU, C. M. and KAUERMANN, G. (2012). Fast Adaptive Penalized Splines. *Journal of Computational and Graphical Statistics* **17**, 1–20.

LAIRD, N. (1978). Nonparametric Maximum Likelihood Estimation of a Mixing Distribution. *Journal of American Statistical Association* **73**, 805–811.

LARSEN, K. (2004). Joint Analysis of Time-to-Event and Multiple Binary Indicators of Latent Classes. *Biometrics* **60**, 85–92.

LIN, X. (2007). Estimation Using Penalized Quasilikelihood and Qusi-pseudo-likelihood in Poisson Mixed Models. *Journal of Computational and Graphical Statistics* **13**, 533–544.

LIN, X. and BRESLOW, N. E. (1996). Bias Correction in Generalized Linear Mixed Models with Multiple Components of Dispersion. *Journal of American Statistical Association* **91**, 1007–1016.

LIU, L., MA, J. Z. and O'QUIGLEY, J. (2008). Joint analysis of multi-level repeated measures data and survival: an application to the end stage renal disease (ESRD) data. *Journal of Computational and Graphical Statistics* **27**, 5676–5691.

LOCALIO, A. R., BERLIN, J. A. and TEN HAVE, T. R. (2006). Longitudinal and repeated cross-sectional cluster-randomization designs using mixed effects regression for binary outcomes: Bias and coverage of frequentist and Bayesian methods. *Journal of Computational and Graphical Statistics* **25**, 2720–2736.

LOUIS, T. A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal Royal Statistical Society B* **44**, 226–233.

MASAOUD, E. and STRYHN, H. (2010). A simulation study to assess statistical methods for binary repeated measures data. *Preventive Veterinary Medicine* **93**, 81–97.

MAZROUI, Y., MATHOULIN-PELISSIER, S., SOUBEYRAN, P. and RONDEAU, V. (2012). General Joint Fraility Model for Recurrent Event Data with a Dependent Terminal Event: Application to follicular lymphoma data. *Journal of Computational and Graphical Statistics* **31**, 1162–1176.

NELSON, K. P. and LEROUX, B. G. (2008). Properties and comparison of estimation methods in a log-linear generalized linear mixed model. *Journal of Statistical Computation and Simulation* **78**, 367–384.

QIU, Z., SONG, P. X. K. and TAN, M. (2008). Simplex Mixed-Effects Models for Longitudinal Proportional Data. *Scandinavian Journal of Statistics* **35**, 577–596.

RAUDENBUSH, S. W., YANG, M. L. and YOSEF, M. (2000). Maximum Likelihood for Generalized Linear Models with Nexted Random Effects via High-order Multivariate Laplace Approximation. *Journal of Computational and Graphical Statistics* **9**, 141–157.

RIPATTI, S. and PALMGREN, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016–1022.

RIZOPOULOS, D., VERBEKE, G., LESAFFRE, E. and VANRENTERGHEM, Y. (2008). A Two-Part Joint Model for the Analysis of Survival and Longitudinal Binary Data with Excess Zeros. *Biometrics* **64**, 611–619.

RIZOPOULOS, D., VERBEKE, G. and MOLENBERGHS, G. (2008). Shared Parameter Models under Random Effects Misspecification. *Biometrika* **95**, 63–74.

ROBINSON, G. K. (1991). That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science* **6**, 15–51.

SONG, X., DAVIDIAN, M. and TSIATIS, A. A. (2002). A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data. *Biometrics* **58**, 742–753.

SONG, X. and WANG, C. Y. (2007). Semiparametric Approaches for Joint Modeling of Longitudinal and Survival Data with Time-Varying Coefficients. *Biometrics* **64**, 557–566.

TSENG, Y. K., HSIEH, R. and WANG, J. L. (2005). Joint Modelling of Accelerated Failure Time and Longitudinal Data. *Biometrika* **92**, 587–603.

TSIATIS, A. A., DEGRUTTOLA, V. and WULFSOHN, M. (1995). Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS. *Journal of American Statistical Association* **90**, 27–37.

TSIATIS, A. A. and DAVIDIAN, M. (2001). A Semiparametric Estimator for the Proportional Hazards Model with Longitudinal Covariates Measured with Error. *Biometrika* **88**, 447–458.

WANG, Y. and TAYLOR, J. M. G. (2001). Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome. *Journal of American Statistical Association* **96**, 895–905.

WOOD, S. N. (2011). Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimationof Semiparametric Generalized Linear Models. *Journal Royal Statistical Society B* **73**, 3–36.

WU, M. and CARROLL, R. (1988). Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modelling the Censoring Process. *Biometrics* **44**, 175–188.

WULFSOHN, M. and TSIATIS, A. A. (1997). A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics* **53**, 330–39.

XU, J. and ZEGER, S. (2001a). The Evaluation of Multiple Surrogate Endpoints. *Biometrics* **57**, 81–87.

XU, J. and ZEGER, S. (2001b). Joint Analysis of Longitudinal Data Comprising Repeated Measures and Times to Events. *Applied Statistics* **50**, 375–387.

YE, W., LIN, X. H. and TAYLOR, J. M. G. (2008a). Semiparametric modeling of longitudinal measurements and time-to-event data–a two-stage regression calibration approach. *Biometrics* **64**, 1238–1246.

YE, W., LIN, X. H. and TAYLOR, J. M. G. (2008b). A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. *Statistics and Its Interface* **1**, 33–45.

YU, X. and LIU, L. (2011). A Joint Model of Recurrent Events and a Terminal Event with a Nonparametric Covariate Function. *Journal of Computational and Graphical Statistics* **30**, 2683–2695.

YU, X., LIU, L., BRAVATA, D. M., WILLIAMS, L. S. and TEPPER, R. S. (2013). A Semiparametric Recurrent Events Model with Time-varying Coefficients. *Journal of Computational and Graphical Statistics* **32**, 1016–1026.

ZENG, D. and CAI, J. (2005a). Simultaneous Modelling of Survival and Longitudinal Data with an Application to Repeated Quality of Life Measures. *Journal of Computational and Graphical Statistics* **11**, 151–174.

ZENG, D. and CAI, J. (2005b). Asymptotic Results for Maximum Likelihood Estimators in Joint Analysis of Repeated Measurements and Survival Time. *The Annals of Statistics* **33**, 2132–2163.

Jaeun Choi
Department of Epidemiology and
Population Health,
Albert Einstein College of Medicine
1300 Morris Park Avenue, New York,
NY 10461, USA
E-mail: jaeun.choi@einstein.yu.edu

Jianwen Cai
Donglin Zeng
Department of Biostatistics,
University of North Carolina at
Chapel Hill,
McGavran-Greenberg Hl, 135 Dauer
Drive, CB 7420, North Carolina, NC
27599, USA

Paper received: 30 April 2014.