Joint Modeling of Survival Time and Longitudinal Outcomes with Flexible Random Effects

Jaeun Choi,

Department of Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

Donglin Zeng,

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

Andrew F. Olshan, and

Department of Epidemiology, University of North Carolina at Chapel Hill, McGavran-Greenberg Hl, 135 Dauer Drive, CB 7435, Chapel Hill, NC 27599, USA

Jianwen Cai

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

Abstract

Joint models with shared Gaussian random effects have been conventionally used in analysis of longitudinal outcome and survival endpoint in biomedical or public health research. However, misspecifying the normality assumption of random effects can lead to serious bias in parameter estimation and future prediction. In this paper, we study joint models of general longitudinal outcomes and survival endpoint but allow the underlying distribution of shared random effect to be completely unknown. For inference, we propose to use a mixture of Gaussian distributions as an approximation to this unknown distribution and adopt an Expectation-Maximization (EM) algorithm for computation. Either AIC and BIC criteria are adopted for selecting the number of mixtures. We demonstrate the proposed method via a number of simulation studies. We illustrate our approach with the data from the Carolina Head and Neck Cancer Study (CHANCE).

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 can 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. Dr. Jack Kalbeisch has done important and influential work in the area of joint modeling of longitudinal data and survival time.

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, De Gruttola, and Wulfsohn (1995), Wulfsohn and Tsiatis (1997), Henderson, Diggle and Dobson (2000), Tsiatis and Davidian (2001), Xu and Zeger (2001a,b), Song, Davidian and Tsiatis (2002), Larsen (2004), Tseng, Hsieh and Wang (2005), Hsieh, Tseng and Wang (2006), Song and Wang (2007), Ye, Lin and Taylor (2008), Huang, Stefanski and Davidian (2009) 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), Wang, Wang and Wang (2000), Albert and Follmann (2000, 2007), Liu, Wolfe and Kalbeisch (2007), and Ding and Wang (2008) among others. On the other hand, simultaneous models of longitudinal outcome and survival time were proposed by Xu and Zeger (2001b) and Zeng and Cai (2005a, 2005b) and further studied by Elashfoff, Li and Ni (2007, 2008), Liu, Ma and O'Quigley (2008), Rizopoulos, Verbeke and Molenberghs (2008), Rizopoulos, Verbeke, Lesaffre and Vanrenterghem (2008), Choi, Cai, Zeng, and Olshan (2015), and Choi, Cai, and Zeng (2017). Wang and Taylor (2001), Brown and Ibrahim (2003), Dunson and Herring (2005), Chen, Ghosh, Raghunathan, and Sargent (2009), Hu, Li and Li (2009), Ghosh, Ghosh and Tiwari (2011), Huang, Li, Elashfoff and Pan (2011) and Baghfalaki, Ganjali and Verbeke (2016) studied simultaneous modeling from the Bayesian perspective.

In most of existing methods, random effects are incorporated to accommodate latent dependence between survival time and longitudinal data. Furthermore, random effects are conventionally assumed to be normally distributed and this assumption plays a vital role in parameter estimation and inference. However, the latter is not testable using observed data and moreover, it is well documented that misspecifying normality assumption can lead to serious bias in estimation (Neuhaus, Hauck, and Kalbeisch, 1992; Kleinman and Ibrahim, 1998; Heagerty and Kurland, 2001; Agresti, Caffo, and Ohman-Strickland, 2004). This concern was also noted in joint models (Wulfsohn and Tsiatis, 1997; Wang, Wang and Wang, 2000) and the assumption was relaxed by Song, Davidian and Tsiatis (2002) in a proportional hazard model depending on longitudinal process requiring only that random effects have density belonging to a class of smooth densities studied by Gallant and Nychka (1987) who suggested a seminonparametric (SNP) density estimator. In a similar setting of joint models for a time-to-event endpoint, Tsiatis and Davidian (2001) proposed conditional score estimators (CSEs) that require no assumption on the distribution of the random effects. Li, Zhang, and Davidian (2004) also considered CSEs in joint models for a simple endpoint of a generalized linear model with covariates that are subject-specific random effects in a linear mixed effect model for measurements. The issue of robustness of joint models to the distributional assumption on random effects was further discussed by some authors. Hsieh, Tseng and Wang (2006) suggested that the maximum likelihood estimators (MLEs) in joint models with a primary time-to-event endpoint and a longitudinal covariate process are robust to the violation of the random effect model assumption when there is rich enough information available from the longitudinal data (i.e. the longitudinal data should not be too

sparse or carry too large measurement errors). Rizopoulos, Verbeke, and Molenberghs (2008) concluded that the effect of misspecifying the random effects distribution in joint models of survival and longitudinal processes becomes minimal (converging to zero) as the number of repeated longitudinal measurements per individual increases. Huang, Stefanski and Davidian (2009) presented diagnostic tools that can reveal adverse effects of random effect model misspecification in joint models of a primary endpoint and a longitudinal process by improving the remeasurement method for structural measurement error models (Huang, Stefanski and Davidian, 2006) which was derived from the simulation-extrapolation (SIMEX) method developed by Cook and Stefanski (1994) and Stefanski and Cook (1995). As an effort to resolve the issue, instead of using Gaussian random effects and errors, Dirichlet process (DP) priors were assumed to model the distribution of individual random effects and error distribution in a fully Bayesian approach by Ghosh, Ghosh and Tiwari (2011) who considered a multiple-changepoint model for longitudinal process and a proportional hazards model for dropout time.

Alternatively, some studies considered latent class memberships to be shared between longitudinal marker trajectory and risk of event under the structure of assuming a heterogeneous population of subjects who can be divided into latent homogenous subgroups and modeling an individual's probability of belonging to a latent class via a multinomial logistic regression. The joint latent class model for a longitudinal biomarker and an event-time outcome subject to censoring was proposed by Lin, Turnbull, McCulloch and Slate (2002) who generalized the latent class models of Muthén and Shedden (1999) and Lin, McCulloch, Turnbull, Slate and Clark (2000) for a longitudinal biomarker and a binary outcome in the setting of complete follow-up. Garre, Zwinderman, Geskus and Sijpkens (2008) used a Bayesian approach to fit a joint latent class changepoint model for survival prediction with longitudinal biomarker readings. Recently, Proust-Lima, Séne, Taylor and Jacqmin-Gadda (2014) studied the joint latent class models in details in comparison with the joint shared random effects models. On the other hand, to allow the heterogeneous population in joint shared random effects models of continuous longitudinal measurements and event time data, Rizopoulos, Verbeke and Molenberghs (2008) and Huang, Li and Elashoff (2010) proposed parameterizations of normal random effects and most recently Baghfalaki, Ganjali and Verbeke (2016) considered a finite mixture of normal distributions as the distribution for random effects. Heterogeneity of the shared random effects was also assessed in Baghfalaki, Ganjali and Verbeke (2016) by adopting the graphical method of Verbeke and Molenberghs (2013) which presented the gradient function as an exploratory diagnostic tool for the assumed distribution of random effects in mixed models.

In this paper, we seek to alleviate the problems due to violation of normality of random effects when considering simultaneous models in the joint analysis of general longitudinal outcomes (continuous or categorical) and survival time, by assuming the underlying distribution of random effects is completely unknown. For estimating model parameters, we propose to use a mixture of Gaussian distributions as an approximation for the unknown random effect distribution. Using a finite mixture of parametric distributions to approximate an unknown distribution has been mostly studied in other context, including linear mixed effect models (Verbeke and Lesaffre, 1996; Verbeke and Molenberghs, 2000; Xu and Hedeker, 2001; Zhang and Davidian, 2001; Lemenuel-Diot, Mallet, Laveille, and Bruno,

2005; Cheon, Albert, and Zhang, 2012) and generalized linear mixed effect models (Komárek and Verbeke, 2002; Verbeke and Lesaffre, 1996; Fieuws, Spiessens, Draney, 2004; Caffo, Ming-Wen and Rohde, 2007; Cagnone and Viroli, 2012). Also, finite normal mixture models were studied by many authors and in particular the work by Dr. Kalbeisch and his colleagues includes Lesperance and Kalbeisch (1992), Neuhaus, Hauck and Kalbeisch (1992) and Chen and Kalbeisch (1996, 2005). In the joint modeling framework, Baghfalaki, Ganjali and Verbeke (2016) used a finite mixture of normal distributions for the shared random effects in survival time and continuous longitudinal data processes and developed a Bayesian procedure for estimation and inference. Therefore, this is the first time we extend this method to joint models of general longitudinal data – incorporating both continuous and categorical types – and survival time. With the approximation, an expectation-maximization (EM) algorithm can be used for parameter estimation in the joint models. We also adopt the Akaike's information criterion (AIC) (Akaike, 1973) and the Bayesian information criterion (BIC) (Schwarz, 1978) in this paper for selecting the number of mixtures.

The outline of this paper is as follows. In Section 2, we present a simultaneous modeling for longitudinal outcomes and survival time with random effects from an unknown distribution, and describe the inference procedure. Asymptotic properties of the proposed estimators are investigated in Section 3, and numerical results from simulation studies are given in Section 4. Our proposed method is illustrated with the data from the Carolina Head and Neck Cancer Study (CHANCE) in Section 5. In Section 6, we discuss some further consideration.

2 Models and Inference Procedure

2.1 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 exponential family in order to incorporate both continuous and categorical measurements. Let T denote survival time, and suppose that the survival time T is possibly right censored. Suppose a set of n subjects are followed over an interval $[0, \tau]$, where τ is the study end time. Denote b_i^*, $i = 1, \ldots, n$, as a vector of subject-specific random effects of dimension d_b and b_i^*'s are mutually independent. Different from the traditional joint models, we assume the underlying distribution of b_i^* is completely unknown and denote its density as $f(b_i^*)$.

Given the random effects 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,

$$\exp \left\{ \frac{y_i \eta_i(t) - B(\eta_i(t))}{A(D_i(t;\phi))} + C(y_i, D_i(t;\phi)) \right\} \tag{1}$$

with $\mu_i(t) = \mathrm{E}(Y_i(t)|b_i^*) = B'(\eta_i(t))$ and $\upsilon_i(t) = \mathrm{Var}(Y_i(t)|b_i^*) = B''(\eta_i(t))A(D_i(t;\phi))$, satisfying

$$\eta_i(t) = g(\mu_i(t)) = \boldsymbol{X}_i(t)\boldsymbol{\beta} + \tilde{\boldsymbol{X}}_i(t)\boldsymbol{b}_i^*$$

and $\upsilon_i(t) = \upsilon(\mu_i(t))A(D_i(t, \boldsymbol{\phi}))$, where $g(\cdot)$ and $\upsilon(\cdot)$ are known link and variance functions respectively, $\boldsymbol{X}_i(t)$ and $\tilde{\boldsymbol{X}}_i(t)$ are the row vectors of the observed covariates for subject i and may include external time-dependent covariates, and $\boldsymbol{\beta}$ is a column vector of coefficients for $\boldsymbol{X}_i(t)$. $\boldsymbol{X}_i(t)$ does not include intercept and it does not contain any covariates in $\tilde{\boldsymbol{X}}_i(t)$ because the intercept and any potential common covariates for fixed effects are combined with the corresponding random effects in $\tilde{\boldsymbol{X}}_i(t)$ so that mean of random effects does not have any restriction. The random effect $\boldsymbol{b}_i^*$ is allowed to differ for different individuals.

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,

$$\lambda_s(t) \, \exp\{\tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\psi} \circ \boldsymbol{b}_i^*) + \boldsymbol{Z}_i(t)\boldsymbol{\gamma}\}, \quad (2)$$

where, 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; $\boldsymbol{Z}_i(t)$ and $\tilde{\boldsymbol{Z}}_i(t)$ 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 $\boldsymbol{Z}_i(t)$ and $\tilde{\boldsymbol{Z}}_i(t)$ do not include dummy variables for strata since baseline hazard rate is stratum-specific. For notation, we use common fixed effects and random effects across strata in both hazard and longitudinal models. However, the model is flexible and allows for possibly different covariate effects for different strata, which can be accommodated 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 addition, $\tilde{\boldsymbol{X}}_i(t)$ and $\tilde{\boldsymbol{Z}}_i(t)$ have the same dimensions as $\boldsymbol{b}_i^*$'s.

Under models (1) and (2), the two outcomes $Y(t)$ and T are independent conditional on the covariates and the random effects. The parameter $\boldsymbol{\psi}$ in model (2) characterizes the dependence between the longitudinal outcomes and the survival time due to latent random effect: When the m-th component of $\boldsymbol{\psi}$ is 0 (i.e. $\psi_m = 0$), it implies that the dependence between the survival time and longitudinal responses is not due to the corresponding latent variable b_{im}^*; $\psi_m \neq 0$ implies that such dependence may be due to the corresponding latent variable b_{im}^*.

Let n_i be the number of 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 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 $(n_i, Y_{ij}, X_{ij}, \tilde{X}_{ij})$, $j=1, \ldots, n_i$, $i=1, \ldots, n$, and $(V_i, \Delta_i, S_i, \{(\boldsymbol{Z}_i(t), \tilde{\boldsymbol{Z}}_i(t)) : t \leq V_i\})$, $i=1, \ldots, n$, where for subject i, $(Y_{ij}, X_{ij}, \tilde{X}_{ij})$ is the j-th observation of

$(Y_i(t), X_i(t), \tilde{X}_i(t))$, C_i is the right-censoring time and assumed to be independent of T_i, $Y_i(t)$, and random effects conditional on all covariates, $V_i = \min(T_i, C_i)$, S_i denotes the stratum, and $\Delta_i = I(T_i \leq C_i)$. 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, the likelihood function of the complete data $(\boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b}^*)$ has the form,

$$
\begin{aligned}
L_c(\boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b}^*) &= \prod_{i=1}^n f(\boldsymbol{Y}_i | \boldsymbol{b}_i^*) \left(\prod_{s=1}^S [f(V_i | \boldsymbol{b}_i^*)]^{I(S_i = s)} \right) f(\boldsymbol{b}_i^*) \\
&= \prod_{i=1}^n \ \exp \left\{ \sum_{j=1}^{n_i} \left[\frac{Y_{ij}(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \tilde{\boldsymbol{X}}_{ij}\boldsymbol{b}_i^*) - B(\boldsymbol{\beta}; \boldsymbol{b}_i^*)}{A(D_i(t_j; \phi))} + C(Y_{ij}; D_i(t_j; \phi)) \right] \right\} \\
&\quad \times \left(\prod_{s=1}^S \left[\lambda_s(V_i)^{\Delta_i} \ \exp \left\{ \Delta_i [\tilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\psi} \circ \boldsymbol{b}_i^*) + \boldsymbol{Z}_i(V_i)\boldsymbol{\gamma}] \right. \right. \right. \\
&\quad \left. \left. \left. - \int_0^{V_i} \exp \left\{ \tilde{\boldsymbol{Z}}_i(u)(\boldsymbol{\psi} \circ \boldsymbol{b}_i^*) + \boldsymbol{Z}_i(u)\boldsymbol{\gamma} \right\} d\Lambda_s(u) \right\} \right]^{I(S_i = s)} \right) \\
&\quad \times f(\boldsymbol{b}_i^*),
\end{aligned}
\tag{3}
$$

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

$$
L_f(\boldsymbol{Y}, \boldsymbol{V}) = \int_{\boldsymbol{b}^*} L_c(\boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b}^*) d\boldsymbol{b}^*.
\tag{4}
$$

2.2 Inference Procedure

Since the distribution of the random effects is completely unknown, it is necessary to estimate this distribution nonparametrically. However, since there are no observations associated with such latent random effects, a fully nonparametric estimation can be numerically unstable. Instead, we propose to estimate this unknown distribution via an approximation by a series of parametric distributions. Particulary, we choose to use a finite mixture of normal distributions to approximate this unknown distribution where the number of mixtures will be chosen based on data.

For the subject-specific random effects $\boldsymbol{b}_i^*$ in Section 2.1, we approximate the distribution of $\boldsymbol{b}_i^*$ with a mixture of a finite number of d_b-dimensional multivariate normal distributions. That is, the distribution of $\boldsymbol{b}_i^*$ is approximated by $\sum_{k=1}^K w_k \mathcal{N}(\boldsymbol{\mu}_k, \sum_b)$, where K is the number of mixture components. We denote the probability of belonging to component k by w_k, such that $\sum_{k=1}^K w_k = 1$. $\boldsymbol{\mu}_k$ is the mean of the k-th component and it is assumed that each component has the same covariance matrix Σ_b. This constraint is needed to avoid infinite likelihoods (Böhning, 1999). We write $\boldsymbol{w} = (w_1, \ldots, w_{K-1})^T$, the vector of $K-1$ component probabilities, and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \ldots, \boldsymbol{\mu}_K^T)^T$, the vector of all component means which are ordered from the largest to the smallest ($\boldsymbol{\mu}_1 > \boldsymbol{\mu}_2 > \cdots > \boldsymbol{\mu}_K$) for identifiability of component labels.

We introduce b_i and $a_i = k$, ($k = 1, \ldots, K$), as the i-th subject's random effects following the mixture distribution and the k-th component of the mixture from which b_i is sampled, respectively. The distribution of a_i is then described by $P(a_i = k) = w_k$ and, given $a_i = k$, b_{ik} $\sim \mathcal{N}(\mu_k, \Sigma_b)$. Thus, $b_i = \sum_{k=1}^{K} I(\alpha_i = k) b_{ik}$, where $I(a_i = k)$ is the indicator of belonging to component k. For n subjects, $b = (b_1^T, \ldots, b_n^T)^T$ and $a = (a_1, \ldots, a_n)^T$.

Now we estimate and make inferences on the parameters $\theta = (\beta^T, \phi^T, \mathrm{Vech}(\Sigma_b)^T, \mu^T, w^T, \psi^T, \gamma^T)^T$, where $\mathrm{Vech}(\cdot)$ operator creates a column vector from a matrix by stacking the diagonal and upper-triangle elements of the matrix, and the baseline cumulative hazard functions with S strata, $\Lambda(t) = (\Lambda_1(t), \ldots, \Lambda_S(t))^T$, where $\Lambda_s(t) = \int_0^t \lambda_s(u) \mathrm{d}u$, $s = 1, \ldots S$. The parameters β and ϕ are from the longitudinal model, ψ and γ are from the hazard model, and μ, w, and Σ_b are associated with the random effects. The likelihood function (3) of the complete data (Y, V, b, a) and the full likelihood function (4) of the observed data (Y, V) for (θ, Λ) have the following forms respectively,

$$L_c(\theta, \Lambda; Y, V, b, \alpha)$$

$$= \prod_{i=1}^{n} \prod_{k=1}^{K} \left[\exp\left\{ \sum_{j=1}^{n_i} \left[\frac{Y_{ij}(X_{ij}\beta + \tilde{X}_{ij}b_{ik}) - B(\beta; b_{ik})}{A(D_i(t_j; \phi))} + C(Y_{ij}; D_i(t_j; \phi)) \right] \right\} \times \left(\prod_{s=1}^{S} \left[\lambda_s(V_i)^{\Delta_i} \exp\left\{ \Delta_i [\tilde{Z}_i(V_i)(\psi \circ b_{ik}) + Z_i(V_i)\right. \right. \right.$$

and

$$L_f(\theta, \Lambda; Y, V) = \sum_{\alpha} \int_b L_c(\theta, \Lambda; Y, V, b, \alpha) db.$$

The proposed estimation method is to calculate the maximum likelihood estimates for ($\theta, \Lambda(t)$) over a set of θ and $\Lambda(t)$. We let each $\Lambda_s(t)$ of $\Lambda(t)$, $s = 1, \ldots, S$, be a non-decreasing and right-continuous step function with jumps only at the observed failure times belonging to stratum s.

EM-algorithm is used for calculating the maximum likelihood estimates. In the EM-algorithm, b_i and a_i are considered as missing data for $i = 1, \ldots, n$. Therefore, the M-step solves the conditional score equations from complete data given observations, where the conditional expectation can be evaluated in E-step. The procedure involves iterating between the following two steps until convergence is achieved: at the m-th iteration,

> _(1) E-step_ Calculate the conditional expectations of some known functions of b_i and a_i, needed in the next M-step, for subject i with $S_i = s$ given observations and the current estimate ($\theta^{(m)}, \Lambda_s^{(m)}$). The conditional expectation is calculated using the Gauss-Hermite Quadrature numerical approximation, denoted as $E[q(b_i, \alpha_i) | \theta^{(m)}, \Lambda_s^{(m)}]$ for a known function $q(b_i, a_i)$.

> _(2) M-step_ After differentiating the conditional expectation of complete data log-likelihood function given observations and the current estimate ($\theta^{(m)}, \Lambda^{(m)}$), the updated estimator ($\theta^{(m+1)}, \Lambda^{(m+1)}$) can be obtained as follows: ($\beta^{(m+1)}, \phi^{(m+1)}$) solves

the conditional expectation of complete data log-likelihood score equation using one-step Newton-Raphson iteration; For the covariance matrix of random effects,

$$\sum\nolimits_{b}^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{S}\sum_{k=1}^{K}\mathrm{E}[\,I(\alpha_i{=}k)(\boldsymbol{b}_{\mathrm{ik}}-\boldsymbol{\mu}_k)(\boldsymbol{b}_{\mathrm{ik}}-\boldsymbol{\mu}_k)^T|\boldsymbol{\theta}^{(m)},\Lambda_s^{(m)}]I(S_i{=}s);$$

For the k-th mixture component ($k = 1, \ldots, K$),

$$\mu_k^{(m+1)} = \frac{\sum_{i=1}^{n}\sum_{s=1}^{S}\mathrm{E}[\,I(\alpha_i{=}k)\boldsymbol{b}_{\mathrm{ik}}|\boldsymbol{\theta}^{(m)},\Lambda_s^{(m)}]I(S_i{=}s)}{\sum_{i=1}^{n}\sum_{s=1}^{S}\mathrm{E}[\,I(\alpha_i{=}k)|\boldsymbol{\theta}^{(m)},\Lambda_s^{(m)}]I(S_i{=}s)}$$

and

$$w_k^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{S}\mathrm{E}[\,I(\alpha_i{=}k)|\boldsymbol{\theta}^{(m)},\Lambda_s^{(m)}]I(S_i{=}s);$$

($\boldsymbol{\psi}^{(m+1)}$, $\boldsymbol{\gamma}^{(m+1)}$) solves the partial likelihood score equation from the full data using one-step Newton-Raphson iteration,

$$\sum_{i=1}^{n}\sum_{s=1}^{S}\Delta_i\left\{\begin{pmatrix} E[(\tilde{\boldsymbol{Z}}_i^T(V_i)\circ\boldsymbol{b}_i)|\boldsymbol{\theta}^{(m)},\Lambda_s^{(m)}] \\ \boldsymbol{Z}_i \end{pmatrix} - \frac{\sum_{l:V_l\geq V_i}\begin{pmatrix} E[(\tilde{\boldsymbol{Z}}_l^T(V_i)\circ\boldsymbol{b}_l)\ \exp\{\tilde{\boldsymbol{Z}}_l(V_i)(\boldsymbol{\psi}\circ\boldsymbol{b}_l)+\boldsymbol{Z}_l(V_i)\boldsymbol{\gamma}\}|\boldsymbol{\theta}^{(m)},\Lambda_s^{(m)}] \\ E[\boldsymbol{Z}_l(V_i)\ \exp\{\tilde{\boldsymbol{Z}}_l(V_i)(\boldsymbol{\psi}\circ\boldsymbol{b}_l)+\boldsymbol{Z}_l(V_i)\boldsymbol{\gamma}\}|\boldsymbol{\theta}^{(m)},\Lambda_s^{(m)}] \end{pmatrix}I(S_l{=}s)}{\sum_{l:V_l\geq V_i}E[\exp\{\tilde{\boldsymbol{Z}}_l(V_i)(\boldsymbol{\psi}\circ\boldsymbol{b}_l)+\boldsymbol{Z}_l(V_i)\boldsymbol{\gamma}\}|\boldsymbol{\theta}^{(m)},\Lambda_s^{(m)}]I(S_l{=}s)}\right\}I(S_i{=}s)=\boldsymbol{0};$$

$\Lambda_s^{(m+1)}$ is obtained as an empirical function with jumps only at the observed failure time,

$$\Lambda_s^{(m+1)}(t) = \sum_{i:V_i\leq t}\frac{\Delta_i I(S_i{=}s)}{\sum_{l:V_l\geq V_i}\mathrm{E}[\exp\ \{\tilde{\boldsymbol{Z}}_l(V_i)(\boldsymbol{\psi}^{(m+1)}\circ\boldsymbol{b}_l)+\boldsymbol{Z}_l(V_i)\boldsymbol{\gamma}^{(m+1)}\}|\boldsymbol{\theta}^{(m)},\Lambda_s^{(m)}]I(S_l{=}s)}.$$

The expressions of the conditional expectation and the conditional score equations calculated in the E- and M-steps for continuous longitudinal outcomes following a normal distribution and for binary longitudinal outcomes with survival time are given in Supplementary Materials (Web Appendix A).

The observed information matrix via Louis (1982) formula is adopted to obtain the variance estimate for ($\hat{\boldsymbol{\theta}}$, $\hat{\Lambda}(t)$). The variance of $\sqrt{n}\ \hat{\boldsymbol{\theta}}$ is asymptotically equal to the corresponding sub-matrix of the inverse of the calculated observed information matrix.

3 Asymptotic Properties

In this section, we provide asymptotic properties of the proposed estimator $(\hat{\boldsymbol{\theta}}, \hat{\Lambda}(t))$ with $\hat{\boldsymbol{\theta}}$ = $(\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\phi}}^T, \mathrm{Vech}(\hat{\Sigma}_b)^T, \boldsymbol{\mu}^T, \boldsymbol{w}^T, \hat{\boldsymbol{\psi}}^T, \hat{\boldsymbol{\gamma}}^T)^T$ and $\hat{\Lambda}(t) = (\hat{\Lambda}_1(t), \ldots, \hat{\Lambda}_S(t))^T$, when assuming that random effects $\boldsymbol{b}_i$ follow a finite mixture of normal distributions. We need the following conditions.

(A1)
   The true parameter $\boldsymbol{\theta}_0 = (\beta_0^T, \phi_0^T, \mathrm{Vech}(\sum_{b0})^T, \boldsymbol{\mu}^T, \boldsymbol{w}^T, \psi_0^T, \gamma_0^T)^T$ belongs to a known compact set Θ which lies in the interior of the domain for $\boldsymbol{\theta}$.

(A2)   The distribution of random effects $\boldsymbol{b}_i$ is a mixture of a finite number of d_b-dimensional multivariate normal distributions with means $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \ldots, \boldsymbol{\mu}_K^T)^T$ and a common covariance matrix $\boldsymbol{\Sigma}_b$. i.e. $\boldsymbol{b}_i \sim \sum_{k=1}^{K} w_k \mathcal{N}(\boldsymbol{\mu}_k, \sum_b)$, where K is the number of mixture components.

(A3)   The true baseline hazard rate function $\boldsymbol{\lambda}_0(t) = (\lambda_{10}(t), \ldots, \lambda_{S0}(t))$ is continuous and positive in $[0, \tau]$, where τ is the time of study end.

(A4)   For the censoring time C, $P(C \ \ \tau | \boldsymbol{Z}, \tilde{\boldsymbol{Z}}, \boldsymbol{X}, \tilde{\boldsymbol{X}}) = P(C = \tau | \boldsymbol{Z}, \tilde{\boldsymbol{Z}}, \boldsymbol{X}, \tilde{\boldsymbol{X}}) > 0$.

(A5)   For the number of observed longitudinal measurements per subject n_N, $P(n_N > d_b | \boldsymbol{X}, \tilde{\boldsymbol{X}}) > 0$ with probability one, and $P(n_N \ \ n_0) = 1$ for some integer n_0.

(A6)   Both $\boldsymbol{X}^T \boldsymbol{X}$ and $\tilde{\boldsymbol{X}}^T \tilde{\boldsymbol{X}}$ are full rank with positive probability. Moreover, if there exist constant vectors $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ such that, with positive probability, for any t, $\boldsymbol{Z}(t)\boldsymbol{c}_1 = a_0(t)$ and $\tilde{\boldsymbol{Z}}(t) \circ \boldsymbol{c}_2 = 0$ for a deterministic function $a_0(t)$, then $\boldsymbol{c}_1 = 0$, $\boldsymbol{c}_2 = 0$, and $a_0(t) = 0$.

Assumption (A4) means that, by the end of the study, some proportion of the subjects will still be alive and censored at the study end time τ, and thus the maximum right censoring time is equal to τ. Assumption (A5) implies that some proportion of the subjects have at least d_b longitudinal observations, and there exists an integer n_0 such that all subjects have a finite number of longitudinal observations which are not larger than n_0. Consistency and asymptotic distribution of the proposed estimator are summarized in the following two theorems.

Theorem 1

Under the assumptions (A1)~(A6), as $n \to \infty$, the maximum likelihood estimator $(\hat{\boldsymbol{\theta}}, \hat{\Lambda}(t))$ is consistent under the product norm of the Euclidean distance and the supremum norm on $[0, \tau]$. That is, $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| + \sup_{t \in [0, \tau]} \| \hat{\Lambda}(t) - \Lambda_0(t)\| \to 0$, a.s., where

$$\| \hat{\Lambda}(t) - \Lambda_0(t) \| = \sum_{s=1}^{S} |\hat{\Lambda}_s(t) - \Lambda_{s0}(t)|.$$

Theorem 2

Under the assumptions (A1)~(A6), as, $n \to \infty$, $\sqrt{n}((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T, (\hat{\Lambda}(t) - \Lambda_0(t))^T)^T$ weakly converges to a Gaussian random element in $R^{d\theta} \times \ell^\infty[0, \tau] \times \cdots \times \ell^\infty[0, \tau]$, and the estimator

$\hat{\boldsymbol{\theta}}$ is asymptotically efficient, where d_θ is the dimension of $\boldsymbol{\theta}$ and $\ell^\infty[0, \tau]$ is the normed space containing all the bounded functions in $[0, \tau]$.

The proofs of Theorems 1 and 2 follow similar steps as in Choi et al. (2013) and Zeng and Cai (2005b). However, since the distribution for random effects in our method is a finite mixture of normal distributions, some regularity conditions such as parameter identifiability and invertibility of information operators need treatment specific to our models. The latter are non-trivially different from Choi et al. (2013) and Zeng and Cai (2005b). The technical proofs are provided in the Supplementary Materials (Web Appendix B).

4 Simulation Studies

In this section, we present the results from our simulation studies. First, to assess finite sample properties of the proposed maximum likelihood estimators, two sets of simulations with different generalized linear mixed models for the longitudinal outcomes are performed. Continuous and binary data are considered for longitudinal process in the simulations in Sections 4.1 and 4.2, respectively. Then, we conduct simulation studies for examining the robustness of the assumed mixture distribution in Section 4.3. Selection procedures for the number of mixtures by AIC and BIC criteria are assessed through simulation studies in Section 4.4.

4.1 Continuous Longitudinal Outcomes and Survival Time

In this section, we assume Y_{ij} follow a Gaussian distribution given a subject-specific random intercept. Specifically we have

$$Y_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_i + \varepsilon_{ij} = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3ij} + b_i + \varepsilon_{ij},$$

for $j = 1, \ldots, n_i$, where $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_y^2)$, and

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

where $b_i \sim \sum_{k=1}^{K} w_k \mathcal{N}(\mu_k, \sigma_b^2)$, K is the number of mixture components, and $K = 2$ and $K = 3$ are simulated. $X_{1i} \equiv Z_{1i}$ 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, which is a time-dependent variable included in the longitudinal model. We suppose the longitudinal data are observed for every 0.1 unit of time, and thus X_{3ij} has the value of every 0.1 unit ranging over 0 through 2.4. The average number of longitudinal observations (n_i) per subject is 7–8 with the range of 1 to 24. 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 summarizing

the performance 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.

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 $\beta_1 = 1$, $\beta_2 = -0.5$, $\beta_3 = -0.2$, $\sigma_y^2 = 0.5$, $\psi = -0.1$, $\gamma_1 = -0.1$, $\gamma_2 = 0.1$, and $\lambda(t) = 1$. The parameters in the mixture distribution for random effects are $\mu_1 = -1.5$, $\mu_2 = 1.5$, and $w_1 = 0.4$ for $K = 2$ and $\mu_1 = -3$, $\mu_2 = 0$, $\mu_3 = 3$, $w_1 = 0.4$, and $w_2 = 0.3$ for $K = 3$. The weight of the last mixture component (w_2 and w_3 for $K = 2$ and $K = 3$ respectively) is determined from the restriction $\sum_{k=1}^{K} w_k = 1$. The variance of random effects σ_b^2 is chosen as 0.3. Different sample sizes (n=400, 800) are simulated with 1000 replications. The results of the maximum likelihood estimates for $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_y^2, \boldsymbol{\mu}^T, \boldsymbol{w}^T, \sigma_b^2, \psi, \boldsymbol{\gamma}^T)^T$ and the baseline cumulative hazards at the three time points and their respective standard error estimates are reported in Table 1. The simulation study is conducted using R. In Table 1, "True" gives the true values of parameters; the averages of the maximum likelihood estimates from the EM algorithm are in "Est."; 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; "CP" is the coverage proportion of 95% confidence intervals based on the estimated standard error "ESE". Satterthwaite (1946) method is used for the coverage probabilities of σ_y^2 and σ_b^2.

From Table 1, we can see that even for the smaller sample size (n=400), the bias of the estimates from EM algorithm is negligible for most cases. The estimated standard errors calculated from the observed information matrix are close to the sample standard deviations from the 1000 estimates, and the 95% confidence interval coverage rates are close to 0.95 except for weights of the mixture components. The coverage rates of weights are improved for larger sample size in both 2 and 3 mixtures. The estimates for the parameters in the longitudinal and hazards models ($\boldsymbol{\beta}$, σ_y^2, ψ, $\boldsymbol{\gamma}$ and $\Lambda(t)$) perform well for different mixtures.

4.2 Binary Longitudinal Outcomes and Survival Time

In this section, we assume that Y_{ij} is a binary outcome following

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

with $\eta_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3ij} + b_i$ for $j = 1, \ldots, n_i$, and we consider the same hazards model and simulation setting as those used in Section 4.1 except the followings. The parameters in the mixture distribution for random effects are $\mu_1 = -3$, $\mu_2 = 3$, and $w_1 = 0.4$ for $K = 2$ and $\mu_1 = -6$, $\mu_2 = 0$, $\mu_3 = 6$, $w_1 = 0.4$, and $w_2 = 0.3$ for $K = 3$. The binary longitudinal data are generated for every 0.1 and 0.05 units of time for the mixture of 2 and 3 normal distributions, respectively, and X_{3ij}, the time at measurement, has the values of every 0.1 and 0.05 units corresponding to the mixture distributions ranging over 0 through 2.4. Thus, the average numbers of longitudinal observations (n_i) are 7–8 with the range of 1

to 24 and 15–16 with the range of 1 to 48 for the mixture of 2 and 3 distributions, respectively.

The results of the maximum likelihood estimates for $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\mu}^T, \boldsymbol{w}^T, \sigma_b^2, \psi, \boldsymbol{\gamma}^T)^T$ and baseline cumulative hazards at the given three time points and their respective standard error estimates are reported in Table 2. Similar to the results for the continuous longitudinal outcomes, Table 2 shows that overall the estimates perform well even for the smaller sample size $n = 400$ with small biases. The parameters of interest in longitudinal and hazards models have the estimated standard errors which are close to the sample standard deviations. Meanwhile, the estimated standard errors of the parameters of mixture components which are means of random effects and weights appear to be overestimated being larger than their sample standard deviations, which leads to the wide confidence interval.

4.3 Sensitivity for Model-Misspecification

In this section, we conduct simulation studies to examine the sensitivity of the assumed mixture distribution. We consider continuous longitudinal outcomes and survival time with the same setting used in Section 4.1 except for the true distribution of random effects. Random effects are generated from a mixture of a t-distribution with 10 degrees of freedom and non-centrality of −1 and a Gamma distribution with shape and scale parameters of 7 and 1/8 respectively. We assume equal probability for the two distributions. We fit 5 sets of simultaneous models assuming different mixtures for random effects which are 1 normal distribution without mixture and the mixtures of 2, 3, 4 and 5 normal distributions, and we compare the results for the parameters of interest in longitudinal and hazards models and the estimated density plots of random effects. Table 3 shows the results of longitudinal and hazards models from assuming the 5 different models for random effects. As the number of mixtures increases, the changes in bias and coverage rate are more pronounced in the longitudinal model than in the hazards model; it is clear that bias gets smaller and coverage rates become closer to the 95% nominal level in the longitudinal model while biases are similarly small and coverage rates are close to the nominal level over all assumed distributions. From the table, we also find that bigger number of mixtures produces estimates that are closer to the true values in the longitudinal model while estimates in hazards model are less sensitive to the number of distributions in mixture. In other words, when the true distribution of random effects is not a Gaussian distribution, the use of mixture is effective in longitudinal model while the inference on hazards model is reasonable regardless of mixture.

Figure 1 shows the true and estimated density plots of random effects. From these density plots, all the mixture models of 2, 3, 4 and 5 normal distributions produces similar shapes to the true distribution while one normal distribution does not. The mixture of 5 normal distribution appears to be close to the true density. Figure 2 shows the relative bias plot of the parameters in longitudinal and hazard models which are denoted with thin and thick lines respectively. The relative biases are calculated from the median absolute biases divided by their absolute true values. Figure 2 confirms what we observe in Table 3.

For further investigation of the sensitivity to model-misspecification, we conducted additional simulations under another true mixture distribution of random effects – the mixture of non-central $t_{20}(-2)$ and *Gamma*(7, 1/8) – which is more deviated from normal distribution, with heavier tailed and more left-shifted t-distribution in mixture. We fit the same 5 sets of simultaneous models assuming different mixtures for random effects which are 1 normal distribution with no mixture and the mixtures of 2, 3, 4 and 5 normal distributions. Their results, density plots and relative bias plots are provided in the Supplementary Materials (Table 1 and Figures 1 and 2 of Web Appendix D.1). Although the results show slightly bigger biases and inconsistent coverage rates when compared to those for the original true mixture, overall trends appear to be similar. The overall conclusion appears to be similar.

4.4 Selection of the Number of Mixture Distributions

We adopt AIC and BIC for selection of the number of normal distribution in mixture and assess these selection procedures through simulation studies in this section. AIC gives a penalty to a model with more parameters and BIC gives a penalty to a model with more parameters and larger sample size. Given a data set, competing models are ranked according to their AIC (or BIC), with the one having the lowest AIC (or BIC) being the best. Chen and Kalbeisch (1996), who proposed a method for consistent estimation for the mixing distribution and the number of mixture components, mentioned that different penalty methods will provide similar results in many instances and also the application results by their method were consistent to those by AIC and BIC. Thus, the use of AIC and BIC will be a reasonable choice.

Continuous longitudinal outcomes and survival time are considered with the same setting used in Section 4.1. Random effects are generated from a mixture of 3 normal distributions. We fit 5 sets of simultaneous models with different mixtures for random effects which are 1 normal distribution without mixture and the mixtures of 2, 3, 4 and 5 normal distributions. AIC and BIC values are calculated for all 5 fitted mixture models in each data set and we report frequencies of mixture models selected as best by AIC and BIC among 1000 data sets. Sample sizes of 200 and 800 are considered.

The result shows that for the sample size of 200 both AIC and BIC mostly select the true distribution of a mixture of 3 normal distributions as best – 969 and 990 out of 1000 simulated data sets, respectively. For the large sample size of 800, the mixture of 3 normal distributions is selected by both AIC and BIC for all 1000 simulated data sets. This demonstrates that the number of mixture distributions is properly selected by AIC and BIC even for small sample sizes.

5 Analysis of the CHANCE Study

The Carolina Head and Neck Cancer Study (CHANCE) is a population based epidemiologic study conducted at 60 hospitals in 46 counties in North Carolina from 2002 through 2006 (Divaris et al. 2010). Patients were diagnosed with head and neck cancer (oral, pharynx, and larynx cancer) from 2002–2006. Their survival status was collected up to 2007 and their Quality of Life (QoL) was evaluated over time for three years after diagnosis. QoL

information was collected through questionnaires. Based on summary scores of the five domains of self-perceived quality of life including Physical Well-Being (PWB), Social/Family Well-Being (SWB), Emotional Well-Being (EWB), Functional Well-Being (FWB) and Head and Neck Cancer Specific symptoms (HNCS), patient's QoL information was classified into satisfaction or dissatisfaction with life. Survival time is defined as the time to death from diagnosis. Demographic and life style characteristics, medical histories and clinical factors are also collected. Ending in December 2007, information on QoL has been obtained from 554 head and neck cancer patients in the analysis. Based on the death information through 2007 available from the National Death Index (NDI), 85 of 554 patients died and the censoring rate is 85%. All censoring was due to the termination of study and thus the noninformative censoring assumption is appropriate for this study. The number of observations per patient ranges 1 to 3 with average of 1.93. It is of interest to elucidate the variables which are associated with both QoL satisfaction and survival time for patients with head and neck cancer. In particular, we are interested in the comparison between African-Americans and Whites since it is known that African-Americans have a higher incidence of head and neck cancer and worse survival than Whites. The longitudinal QoL satisfaction outcomes and survival time are correlated within a patient, and this dependency should be taken into account in the analysis.

We apply our proposed method to Head and Neck Cancer Specific symptoms (HNCS) among QoL domains with survival time. Longitudinal HNCS QoL outcomes are binary measurements with 1 ("satisfied") and 0 ("dissatisfied"). We are interested in investigating which factors are related to QoL satisfaction and the risk of death. In the full models for both longitudinal QoL and survival time, we consider race (African-Americans, Whites), the number of 12 oz. beers consumed per week (None, <1, 1–4, 5–14, 15–29, 30), household income (0–10K, 20–30K, 40–50K, 60K), surgery (Yes/No), radiation therapy (Yes/No), chemotherapy (Yes/No), primary tumor site (Oral & Pharyngeal, Laryngeal) and tumor stage (I, II, III, IV) as categorical, and age at diagnosis (range: 24–80), the number of persons supported by household income (range: 1–5), body mass index (BMI) (range: 15.66–56.28) and the total number of medical conditions reported (range: 0–6) as continuous. Additionally, 2 interactions with race, i.e. race × the total number of medical conditions reported and race × tumor site, are included in both models since we are particularly interested in the difference of QoL and survival between African American and White. Time at survey measurement is included as a time-dependent covariate for the longitudinal QoL outcome. A random intercept for the dependence between the QoL satisfaction and the risk of death is included in both models and assumed to follow an unknown distribution.

For the full model, we first considered 5 different distributions for random effects which are 1 normal distribution without mixture and the mixtures of 2, 3, 4 and 5 normal distributions, and both AIC and BIC selected a mixture of 3 normal distributions with their lowest values as best. Then, we conducted backward variable selection based on the Likelihood Ratio Test (LRT) from the full model using a mixture of 3 normal distributions for approximating the random effect. Table 4 gives the results from the final models after removing non-significant covariates by LRT. From the "Simultaneous" columns, we see the number of 12 oz. beers consumed per week, household income and tumor stage are significantly associated with both patients' HNCS QoL satisfaction and hazard of death. Using 30 or more of 12 oz. beers

consumed per week as the reference group, all categories of the smaller amount are in general associated with higher odds of being satisfied while the categories of 'none' and '5 to 14' of 12 oz. beers consumed per week are associated with lower risk of death. Higher household income is generally associated with higher odds of being satisfied and lower risk of death. Both patients' HNCS QoL satisfaction and risk of death are significantly different for patients in different tumor stages. On the other hand, race (African-American), radiation therapy, the number of persons supported by household income, and BMI are selected only in the HNCS QoL longitudinal model while the number of medical conditions reported is significant only in the hazard model. The results indicate that African-Americans, patients not treated with radiation therapy, patients in the family with the smaller number of persons supported by household income, or patients with higher BMI are associated with higher odds of being satisfied, but the risk of death is not affected by these factors. On the other hand, higher number of reported medical conditions is associated with higher risk of death, but it is not associated with HNCS QoL satisfaction. Furthermore, time at survey measurement is statistically significant in the HNCS QoL longitudinal model implying that patients have higher odds to be satisfied over time. The parameter ψ for the dependence between longitudinal HNCS QoL and survival time is negative and is statistically significant with p-value as 0.008. This means the longitudinal HNCS QoL and survival time are correlated and some latent factors which increase HNCS QoL satisfaction also decrease the risk of death. Although not provided in Table 4, we have additional parameters of the mixture distribution for random effects in the simultaneous modeling. The obtained estimates of three means of random effects are -3.146, 0.376 and 1.730 with their estimated mixing probabilities of 0.147, 0.105 and 0.748, respectively, and the common variance estimate of random effects is 0.637. In particular, the mixing probabilities are significant at significance level 0.05, which strengthens the mixture of 3 normal distributions with the estimated 3 means of random effects.

For the purpose of comparison, we also conducted separate analyses for longitudinal HNCS QoL and survival time whose results are given in the last three columns of Table 4. Comparing the results from the simultaneous and separate analyses in Table 4, we can see our simultaneous analysis identifies two additional factors (the number of persons supported by household income and BMI) in the HNCS QoL longitudinal model and one additional factor (the number of 12 oz. beers consumed per week) in the hazard model.

Figure 3(a) shows the estimated baseline cumulative hazard rates over follow-up time with 95% confidence interval. The estimated baseline cumulative hazard rates look at at the very early time within a year, but soon appear to be linearly increasing. Figure 3(b) shows the predicted conditional longitudinal trend of HNCS QoL satisfaction probabilities based on the simultaneous models (solid line) and the empirical longitudinal trend of HNCS QoL satisfaction probabilities (dotted line) based on the empirical longitudinal HNCS QoL satisfaction probabilities (dots). The predicted conditional probability of HNCS QoL satisfaction is calculated as the conditional expectation of the conditional probability of HNCS QoL satisfaction given the subject is alive at time t. That is, $E_{b,a} [P(Y(t) = 1 | T > t) | \hat{\theta}, \hat{\Lambda}]$ using model notations in Section 2. The empirical probability of HNCS QoL satisfaction is calculated for every 0.05 unit of time at survey measurements. From Figure

3(b), the longitudinal trend of HNCS QoL satisfaction probabilities appears to be increasing over time and the empirical probabilities also gradually increase over time.

We also applied the rest four unselected distributions for random effects – one normal distribution without mixture and the mixtures of 2, 4 and 5 normal distributions – to the final simultaneous models derived under the mixture of 3 normal distributions and compared their results (provided in the Supplementary Materials – Tables 2–5, respectively, of the Web Appendix D.2) to those in Table 4. Most of the covariates in the final models yielded same conclusions under different distributions assumed for random effects except one covariate. The detail is given in Paragraph 1 of the Web Appendix D.2. On the other hand, overall, the estimates for same variables are similar under all mixtures but slightly different from those under one normal distribution.

In addition, we conducted simulations under the settings similar to the CHANCE data with the high censoring rate = 85% and the low average number of longitudinal observations per patient (n_i) = 1.93, and we compared the results to the simulation studies presented in Section 4.2. The results given in the Supplementary Materials (Table 6 of Web Appendix D. 2) are for the mixture of 2 distributions, sample size (n) of 400, and $\sigma_b^2 = 0.5$ from 1000 data sets. Due to very sparse events, about 17% of the simulated datasets did not converge or encountered problems with variance estimation. Thus, the table only reports the convergent cases and the results show that the bias is reasonably small and that the coverage probabilities are reasonable, although conservative for the coefficients in the longitudinal model. The numerical issues on convergence and variance estimation occurred more for the larger number of mixture components but were recovered for the increased n and n_i and the decreased censoring rate, σ_b^2 and number of predictors included in the joint models.

6 Concluding Remarks

We have relaxed normality assumption of random effects in the simultaneous modeling of longitudinal outcomes and survival time. Assuming the underlying distribution of random effects to be unknown, we used a mixture of Gaussian distributions as an approximation for the random effect distribution. We developed a maximum likelihood estimation method for the proposed simultaneous models and presented asymptotic properties of the proposed estimators. The proposed estimation procedure using EM algorithm has been assessed via simulation studies for both continuous and binary longitudinal data with survival time. The proposed estimates performed well in finite samples. The variance estimates based on the observed information matrix approximate the true variance well in finite samples. Simulation studies indicated that, when the true distribution of random effects is not normal, mixture distributions yield less biased estimates than no mixture and all the estimated density plots of random effects based on mixture distributions appear to have similar shapes to the true distribution. Furthermore, simulation studies also showed that the number of mixture distributions is properly selected by AIC and BIC. The proposed method was applied to data from the CHANCE study.

Consideration of general distributions other than normal distribution in joint modelling is novel. Our method demonstrates the better fit of the real data using 3 mixture distributions, as compared to existing approaches which rely on one single normal random effect.

Alternatively, one may consider the seminonparameteric (SNP) method which is another way to approximate non-normal distributions. However, using mixture normal distributions has a better computational advantage. As described in Section 2, we can treat random effects as a mixture of multiple independent normal variables so the EM algorithm can be easily coined to facilitate computation. In contrast, SNP density does not have this property so the computation has to directly maximize the likelihood function, which is highly nonlinear, over a large number of parameters. Developing the SNP method can be a worthwhile effort for future work.

One generalization of the proposed model is to allow both random intercept and random slope. In this case, we can consider a bivariate mixture normal distribution to approximate their joint distribution. However, the computation will be much more intense due to the higher dimensional numerical integration in the E-step and the increased number of the mixing components. Alleviating the computational intensity of the method for high dimensional random effects will be an interesting topic of investigation of further study.

When estimating variances by the Louis method (1982) as employed in this paper, the information matrix may not be positive definite. This could be due to relatively sparse events because of high censoring or large number of covariates in the model. When there is a problem in estimating the variance, one alternative approach is to use bootstrap method, although bootstrap method is computationally intensive.

7 Supplementary Materials

EM algorithm referenced in Sections 2.2 and technical proofs for Theorem 1 and Theorem 2 referenced in Section 3 are provided in Electronic Supplementary Materials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Agresti A, Caffo B, Ohman-Strickland P. Examples in Which Mis-specification of a Random Effects Distribution Reduces Efficiency, and Possible Remedies. Comput Stat Data Anal. 2004; 47:639–653.

2. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, BN., Csáki, F., 2nd, editors. International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2–8, 1971. Budapest: Akadémiai Kiad'o; 1973. p. 267-281.

3. Albert PS, Follmann DA. Modeling Repeated Count Data Subject to Informative Dropout. Biometrics. 2000; 56:667–677. [PubMed: 10985201]

4. Albert PS, Follmann DA. Random Effects and Latent Processes Approaches for Analyzing Binary Longitudinal Data with Missingness: a Comparison of Approaches Using Opiate Clinical Trial Data. Stat Methods Med Res. 2007; 16:417–439. [PubMed: 17656452]

5. Baghfalaki T, Ganjali M, Verbeke G. A Shared Parameter Model of Longitudinal Measurements and Survival Time with Heterogeneous Random-effects Distribution. J Appl Stat. 2016; doi: 10.1080/02664763.2016.1266309

6. Bickel, PJ., Klaassen, CAJ., Ritov, Y., Wellner, JA. Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press; Baltimore: 1993.

7. Böhning, D. Computer-Assisted Analysis of Mixtures and Applications: Meta-analysis, Disease Mapping and Others, Number 81 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC; 1999.

8. Brown ER, Ibrahim JG. A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data. Biometrics. 2003; 59:221–228. [PubMed: 12926706]

9. Caffo B, Ming-Wen A, Rohde C. Flexible Random Intercept Models for Binary Outcomes Using Mixtures of Normals. Comput Stat Data Anal. 2007; 51:5220–5235. [PubMed: 18628822]

10. Cagnone S, Viroli C. A Factor Mixture Analysis Model for Multivariate Binary Data. Stat Modelling. 2012; 12(3):257–277.

11. Chakraborty A, Das K. Inferences for joint modelling of repeated ordinal scores and time to event data. Comput Math Methods Med. 2010; 11:281–295. [PubMed: 20721765]

12. Chen W, Ghosh D, Raghunathan TE, Sargent DJ. Bayesian Variable Selection with Joint Modeling of Categorical and Survival Outcomes: An Application to Individualizing Chemotherapy Treatment in Advanced Colorectal Cancer. Biometrics. 2009; 65:1030–1040. [PubMed: 19210736]

13. Chen JH, Kalbeisch JD. Penalized Minimum-Distance Estimates in Finite Mixture Models. Can J Stat. 1996; 24:167–175.

14. Chen JH, Kalbeisch JD. Modified likelihood ratio test in finite mixture models with a structural parameter. J Stat Plan Inference. 2005; 129:93–107.

15. Cheon K, Albert PS, Zhang ZW. The impact of random-effect misspecification on percentile estimation for longitudinal growth data. Stat Med. 2012; 31:3708–3718. [PubMed: 22786788]

16. Choi J, Cai J, Zeng D. Penalized Likelihood Approach for Simultaneous Analysis of Survival Time and Binary Longitudinal Outcome. Sankhya Ser B. 2017; doi: 10.1007/s13571-017-0132-3

17. Choi J, Cai J, Zeng D, Olshan AF. Joint Analysis of Survival Time and Longitudinal Categorical Outcomes. Stat Biosci. 2015; 7:19–47. [PubMed: 26052353]

18. Cook J, Stefanski LA. Simulation Extrapolation Estimation in Parametric Measurement Error Models. J Amer Statist Assoc. 1994; 89:1314–1328.

19. Ding J, Wang JL. Modeling Longitudinal Data with Nonparametric Multiplicative Random Effects Jointly with Survival Data. Biometrics. 2008; 64:546–556. [PubMed: 17888040]

20. Divaris K, Olshan AF, Smith J, Bell ME, Weissler MC, Funkhouser WK, Bradshaw PT. Oral Health and Risk for Head and Neck Squamous Cell Carcinoma: the Carolina Head and Neck Cancer Study. Cancer Cause Control. 2010; 21:567–575.

21. Dunson DB, Herring AH. Bayesian latent variable models for mixed discrete outcomes. Biostatistics. 2005; 6:11–25. [PubMed: 15618524]

22. Elashoff RM, Li G, Li N. An Approach to Joint Analysis of Longitudinal Measurements and Competing Risks Failure Time Data. Stat Med. 2007; 26:2813–2835. [PubMed: 17124698]

23. Elashoff RM, Li G, Li N. A Joint Model for Longitudinal Measurements and Survival Data in the Presence of Multiple Failure Types. Biometrics. 2008; 64:762–771. [PubMed: 18162112]

24. Fieuws, S., Spiessens, B., Draney, K. Mixture Models. In: De Boeck, P., Wilson, M., editors. Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach. Springer-Verlag; New York: 2004. p. 317-340.Ch. 11

25. Gallant AR, Nychka DW. Seminonparametric Maximum Likelihood Estimation. Econometrica. 1987; 55:363–390.

26. Garre FG, Zwinderman AH, Geskus RB, Sijpkens YWJ. A Joint Latent Class Changepoint Model to Improve the Prediction of Time to Graft Failure. J Roy Statist Soc Ser A (Statistics in Society). 2008; 171(1):299–308.

27. Ghidey W, Lesaffre E, Eilers P. Smooth Random Effects Distribution in a Linear Mixed Model. Biometrics. 2004; 60:945–953. [PubMed: 15606415]

28. Ghosh P, Ghosh K, Tiwari RC. Joint modeling of longitudinal data and informative dropout time in the presence of multiple changepoints. Stat Med. 2011; 30(6):611–626. [PubMed: 21337357]

29. Henderson R, Diggle P, Dobson A. Joint Modeling of Longitudinal Measurements and Event Time Data. Biometrics. 2000; 4:465–480.

30. Heagerty PJ, Kurland BF. Misspecified Maximum Likelihood Estimates and Generalised Linear Mixed Models. Biometrika. 2001; 88:973–985.

31. Hogan J, Laird N. Mixture Models for the Joint Distribution of Repeated Measures and Event Times. Stat Med. 1997; 16:239–257. [PubMed: 9004395]

32. Hsieh F, Tseng YK, Wang JL. Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited. Biometrics. 2006; 62:1037–1043. [PubMed: 17156277]

33. Hu W, Li G, Li N. A Bayesian Approach to Joint Analysis of Longitudinal Measurements and Competing Risks Failure Time Data. Stat Med. 2009; 28:1601–1619. [PubMed: 19308919]

34. Huang X, Li G, Elashfoff RM. A Joint Model of Longitudinal and Competing Risks Survival Data with Heterogeneous Random Effects and Outlying Longitudinal Measurements. Statistics and Its Interface. 2010; 3:185–195. [PubMed: 21892381]

35. Huang X, Li G, Elashfoff RM, Pan J. A General Joint Model for Longitudinal Measurements and Competing Risks Survival Data with Heterogeneous Random Effects. Lifetime Data Anal. 2011; 17:80–100. [PubMed: 20549344]

36. Huang X, Stefanski LA, Davidian M. Latent-model Robustness in Structural Measurement Error Models. Biometrika. 2006; 93:53–64.

37. Huang X, Stefanski LA, Davidian M. Latent-model Robustness in Joint Models for a Primary Endpoint and a Longitudinal Process. Biometrics. 2009; 65(3):719–727. [PubMed: 19173697]

38. Kleinman KP, Ibrahim JG. A Semiparametric Bayesian Approach to the Random Effects Model. Biometrics. 1998; 54:921–938. [PubMed: 9750242]

39. Komárek A, Lesaffre E. Generalized Linear Mixed Model with a Penalized Gaussian Mixture as a Random Effects Distribution. Comput Stat Data Anal. 2008a; 52:3441–3458.

40. Komárek A, Lesaffre E. Bayesian Accelerated Failure Time Model with Multivariate Doubly Interval-Censored Data and Flexible Distributional Assumptions. J Amer Statist Assoc. 2008b; 103:523–533.

41. Komárek A, Lesaffre E. The Regression Analysis of Correlated Interval-censored Data: Illustration Using Accelerated Failure Time Models with Flexible Distributional Assumptions. Stat Modelling. 2009; 9:299–319.

42. Lange N, Ryan L. Assessing normality in random effects models. Ann Stat. 1989; 17:624–642.

43. Larsen K. Joint analysis of time-to-event and multiple binary indicators of latent classes. Biometrics. 2004; 60:85–92. [PubMed: 15032777]

44. Lemenuel-Diot A, Mallet A, Laveille C, Bruno R. Estimating Heterogeneity in Random Effects Models for Longitudinal Data. Biometr J. 2005; 47:329–345.

45. Lesperance ML, Kalbeisch JD. An Algorithm for Computing the Nonparametric MLE of a Mixing Distribution. J Amer Statist Assoc. 1992; 87:120–126.

46. Lin H, McCulloch CE, Turnbull BW, Slate EH, Clark LC. A Latent Class Mixed Model for Analyzing Biomarker Trajectories in Longitudinal Data With Irregularly Scheduled Observations. Stat Med. 2000; 19:1303–1318. [PubMed: 10814979]

47. Lin H, Turnbull BW, McCulloch CE, Slate EH. Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data: Application to Longitudinal Prostate-specific Antigen Readings and Prostate Cancer. J Amer Statist Assoc. 2002; 97(457):53–65.

48. Liu L, Ma JZ, O'Quigley J. Joint analysis of multi-level repeated measures data and survival: an application to the end stage renal disease (ESRD) data. Stat Med. 2008; 27:5676–5691.

49. Liu L, Wolfe RA, Kalbeisch JD. A shared random effects model for censored medical costs and mortality. Stat Med. 2007; 26:139–155. [PubMed: 16526006]

50. Louis TA. Finding the Observed Information Matrix when Using the EM Algorithm. J Roy Statist Soc Ser B. 1982; 44:226–233.

51. Muthén B, Shedden K. Finite Mixture Modeling With Mixture Outcome Using the EM Algorithm. Biometrics. 1999; 55:463–469. [PubMed: 11318201]

52. Neuhaus JM, Hauck WW, Kalbeisch JD. The Effects of Mixture Distribution Misspecification When Fitting Mixed-Effects Logistic Models. Biometrika. 1992; 79:755–762.

53. Parner E. Asymptotic Theory for the Correlated Gamma-frailty Model. Ann Stat. 1998; 26:183–214.

54. Pinheiro, JC., Bates, DM. Mixed Effects Models in S and S-Plus. Springer-Verlag; New York: 2000.

55. Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint Latent Class Models for Longitudinal and Time-to-event Data: A Review. Stat Methods Med Res. 2014; 23(1):74–90. [PubMed: 22517270]

56. Rizopoulos D, Verbeke G, Lesaffre E, Vanrenterghem Y. A Two-Part Joint Model for the Analysis of Survival and Longitudinal Binary Data with Excess Zeros. Biometrics. 2008; 64:611–619. [PubMed: 17725808]

57. Rizopoulos D, Verbeke G, Molenberghs G. Shared Parameter Models under Random Effects Misspecification. Biometrika. 2008; 95:63–74.

58. Satterthwaite FW. An Approximate Distribution of Estimates of Variance Components. Biometrics. 1946; 2:110–114. [PubMed: 20287815]

59. Schwarz GE. Estimating the dimension of a model. Ann Stat. 1978; 6(2):461–464.

60. Song X, Davidian M, Tsiatis AA. A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data. Biometrics. 2002; 58:742–753. [PubMed: 12495128]

61. Song X, Wang CY. Semiparametric Approaches for Joint Modeling of Longitudinal and Survival Data with Time-Varying Coefficients. Biometrics. 2007; 64:557–566. [PubMed: 17725812]

62. Stefanski LA, Cook J. Simulation Extrapolation: The Measurement Error Jackknife. J Amer Statist Assoc. 1995; 90:1247–56.

63. Tseng YK, Hsieh R, Wang JL. Joint Modelling of Accelerated Failure Time and Longitudinal Data. Biometrika. 2005; 92:587–603.

64. Tsiatis AA, Degruttola V, Wulfsohn M. Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS. J Amer Statist Assoc. 1995; 90:27–37.

65. Tsiatis AA, Davidian M. A Semiparametric Estimator for the Proportional Hazards Model with Longitudinal Covariates Measured with Error. Biometrika. 2001; 88:447–458.

66. van der Vaart, AW. Asymptotic Statistics. Cambridge University Press; 1998.

67. van der Vaart, AW., Wellner, JA. Weak Convergence and Empirical Processes. New York: Springer-Verlag; 1996.

68. Verbeke G, Lesaffre E. A Linear Mixed-effects Model with Heterogeneity in the Random-effects Model with Heterogeneity in the Random-effects Population. J Amer Statist Assoc. 1996; 91:217–221.

69. Verbeke, G., Molengerghs, G. Linear Mixed Models for Longitudinal Data, Springer Series in Statistics. Springer-Verlag; New-York: 2000.

70. Verbeke G, Molengerghs G. The Gradient Function as an Exploratory Goodness-of-fit Assessment of the Random-effects Distribution in Mixed Models. Biostatistics. 2013; 14:477–490. [PubMed: 23376427]

71. Wang Y, Taylor JMG. Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome. J Amer Statist Assoc. 2001; 96:895–905.

72. Wang CY, Wang N, Wang S. Regression Analysis When Covariates Are Regression Parameters of a Random Effects Model for Observed Longitudinal Measurements. Biometrics. 2000; 56:487–495. [PubMed: 10877308]

73. Wu M, Carroll R. Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modelling the Censoring Process. Biometrics. 1988; 44:175–188.

74. Wulfsohn M, Tsiatis AA. A Joint Model for Survival and Longitudinal Data Measured with Error. Biometrics. 1997; 53:330–39. [PubMed: 9147598]

75. Xu W, Hedeker D. A Random-effects Mixture Model for Classifying Treatment Response in Longitudinal Clinical Trials. J Biopharm Stat. 2001; 11:253–273. [PubMed: 12018779]

76. Xu J, Zeger S. The Evaluation of Multiple Surrogate Endpoints. Biometrics. 2001a; 57:81–87. [PubMed: 11252622]

77. Xu J, Zeger S. Joint Analysis of Longitudinal Data Comprising Repeated Measures and Times to Events. Appl Stat. 2001b; 50:375–387.

78. Ye W, Lin XH, Taylor JMG. Semiparametric modeling of longitudinal measurements and time-to-event data-a two-stage regression calibration approach. Biometrics. 2008; 64:1238–1246. [PubMed: 18261160]

79. Zeng D, Cai J. Simultaneous Modelling of Survival and Longitudinal Data with an Application to Repeated Quality of Life Measures. Lifetime Data Anal. 2005a; 11:151–174. [PubMed: 15940822]

80. Zeng D, Cai J. Asymptotic Results for Maximum Likelihood Estimators in Joint Analysis of Repeated Measurements and Survival Time. Ann Stat. 2005b; 33:2132–2163.

81. Zhang D, Davidian M. Linear Mixed Models with Flexible Distributions of Random Effects for Longitudinal Data. Biometrics. 2001; 57:795–802. [PubMed: 11550930]
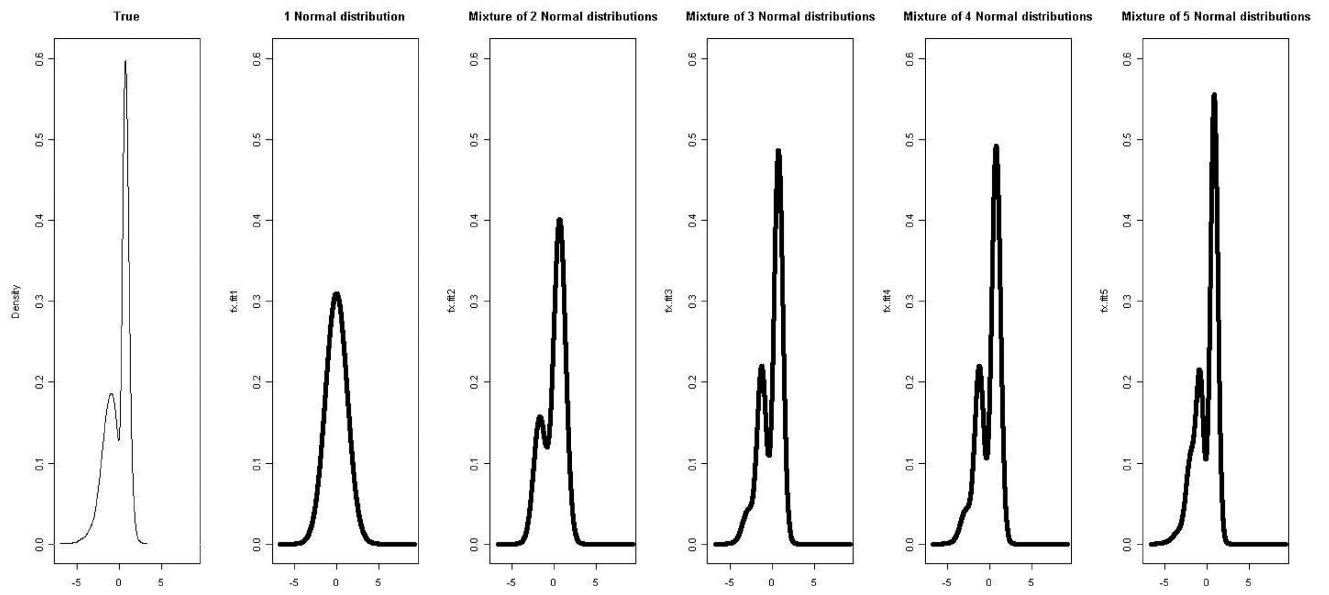
Fig. 1.
Density plots from simulation results of sensitivity for model-misspecification
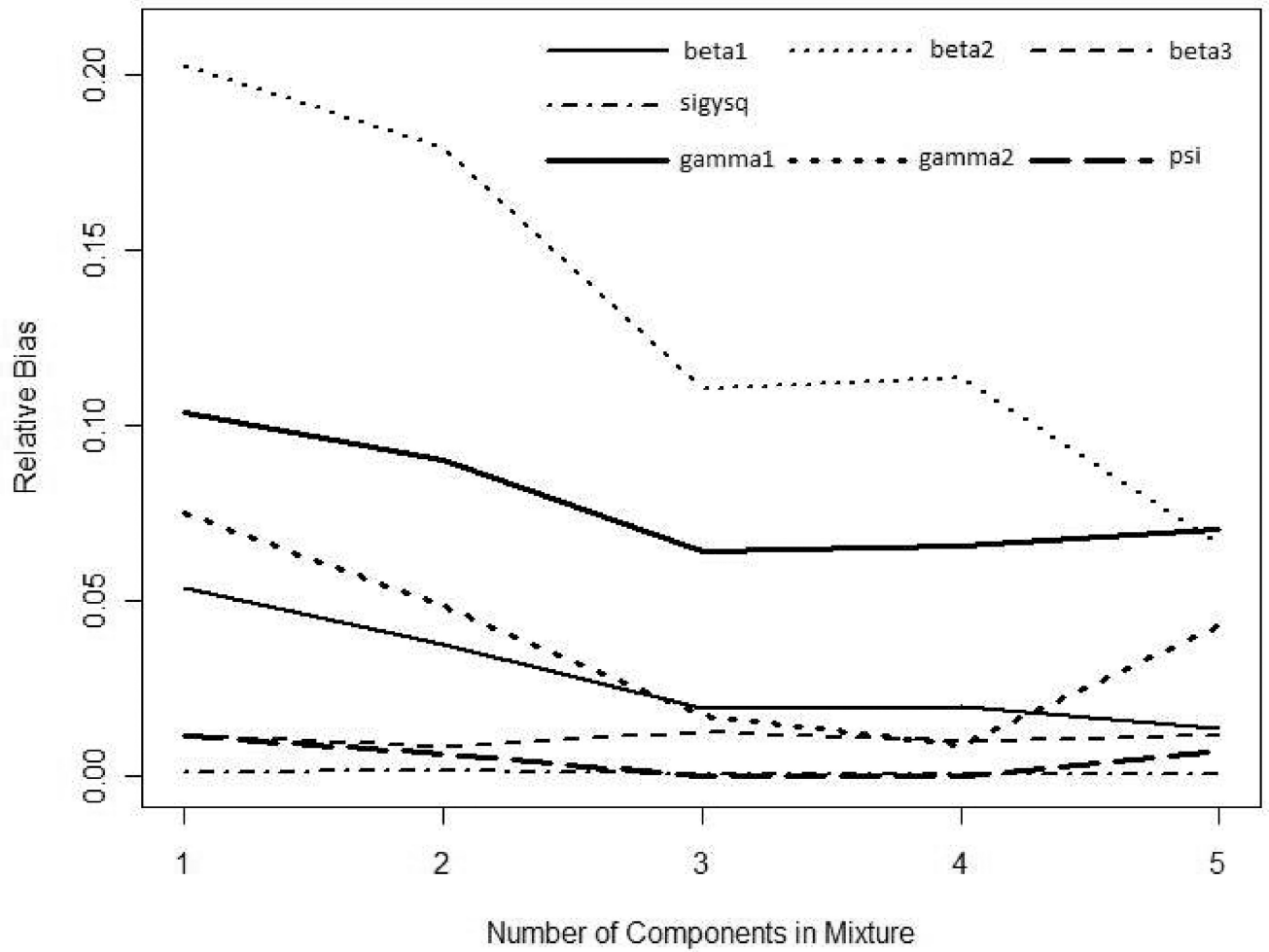
Fig. 2.
Relative bias plot of parameters in longitudinal and hazard models (thin and thick lines respectively) from simulation results of sensitivity for model-misspecification
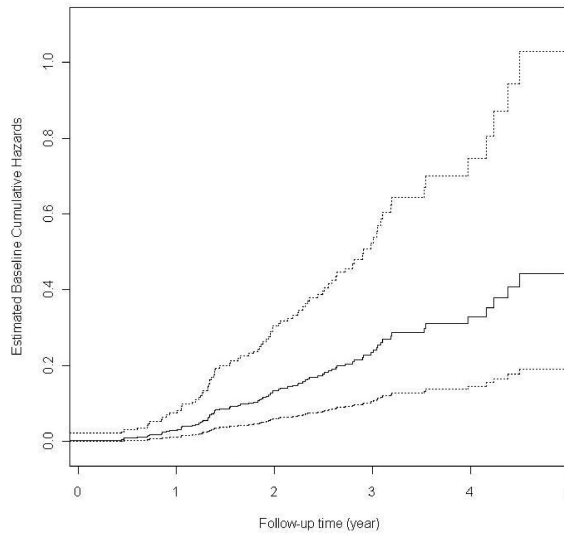
(a) Estimated baseline cumulative hazards (solid line) with 95 % confidence interval (dotted lines) by the simultaneous analysis of HNCS QoL longitudinal outcome and survival time
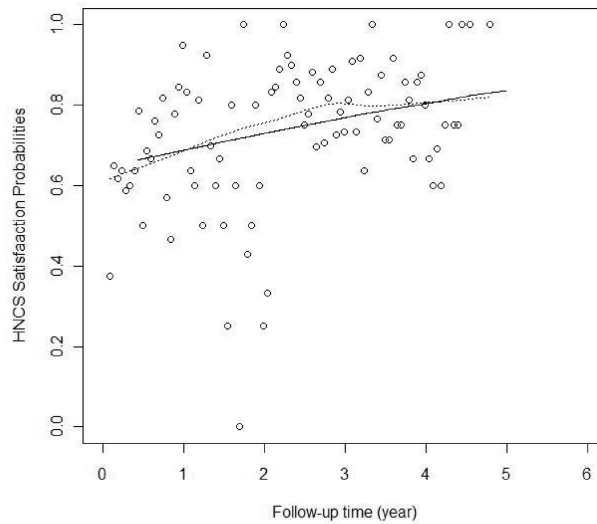
(b) The predicted conditional longitudinal trend based on the simultaneous models (solid line) and the empirical longitudinal trend (dotted line) based on the empirical longitudinal HNCS QoL satisfaction probabilities (dots)

Fig. 3.
Plots of the CHANCE study analysis

Table 1

Summary of simulation results of maximum likelihood estimation using mixtures of Gaussian distributions for random effects in the joint modeling of continuous longitudinal outcomes and survival time.

Mixture	Par.	True	n=400				n=800			
			Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
2	β_1	1.0	.983	.066	.068	.958	.985	.047	.048	.947
	β_2	−.5	−.529	.107	.119	.969	−.540	.079	.084	.947
	β_3	−.2	−.203	.033	.033	.955	−.203	.024	.024	.952
	σ_y^2	.5	.500	.014	.014	.954	.500	.010	.010	.948
	μ_1	−1.5	−1.478	.081	.088	.962	−1.469	.060	.062	.938
	μ_2	1.5	1.524	.075	.082	.966	1.530	.055	.058	.940
	w_1	.4	.400	.025	.033	.991	.401	.018	.023	.981
	σ_b^2	.3	.296	.029	.029	.955	.298	.020	.020	.958
	ψ	−.1	−.102	.040	.039	.950	−.100	.028	.028	.946
	γ_1	−.1	−.101	.123	.121	.945	−.105	.085	.085	.952
	γ_2	.1	.102	.209	.210	.954	.096	.144	.147	.950
	$\Lambda(.9)$	.9	.911	.130	.128	.950	.909	.087	.090	.955
	$\Lambda(1.4)$	1.4	1.421	.206	.202	.942	1.415	.139	.141	.952
	$\Lambda(1.9)$	1.9	1.939	.304	.295	.953	1.924	.205	.205	.950
3	β_1	1.0	.983	.070	.071	.947	.984	.049	.050	.956
	β_2	−.5	−.543	.116	.123	.952	−.543	.085	.087	.922
	β_3	−.2	−.203	.034	.034	.949	−.204	.024	.024	.960
	σ_y^2	.5	.500	.014	.014	.957	.500	.010	.010	.950
	μ_1	−3.0	−2.970	.084	.090	.954	−2.968	.064	.063	.909
	μ_2	.0	.028	.093	.097	.954	.032	.069	.068	.933
	μ_3	3.0	3.030	.089	.094	.954	3.034	.063	.066	.925
	w_1	.4	.400	.025	.033	.992	.400	.018	.023	.983
	w_2	.3	.299	.024	.029	.980	.300	.017	.020	.977

Par.	True	n=400				n=800			
		Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
Mixture									
σ_b^2	.3	.295	.029	.029	.956	.298	.021	.021	.946
ψ	-.1	-.101	.024	.024	.956	-.101	.017	.017	.941
γ_1	-.1	-.091	.112	.119	.963	-.096	.085	.084	.950
γ_2	.1	.088	.215	.207	.946	.114	.146	.146	.944
$\Lambda(.9)$	.9	.913	.125	.127	.948	.897	.088	.088	.951
$\Lambda(1.4)$	1.4	1.417	.202	.200	.949	1.402	.141	.140	.949
$\Lambda(1.9)$	1.9	1.928	.297	.292	.946	1.908	.206	.204	.948

Table 2

Summary of simulation results of maximum likelihood estimation using mixtures of Gaussian distributions for random effects in the joint modeling of binary longitudinal outcomes and survival time.

Mixture	Par.	True	n=400				n=800			
			Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
2	β_1	1.0	1.029	.193	.201	.960	1.015	.143	.141	.942
	β_2	−.5	−.508	.292	.323	.966	−.495	.205	.227	.965
	β_3	−.2	−.200	.166	.180	.966	−.203	.116	.127	.968
	μ_1	−3.0	−3.046	.241	.275	.968	−3.034	.164	.193	.970
	μ_2	3.0	3.016	.211	.253	.976	3.011	.142	.177	.984
	w_1	.4	.401	.025	.033	.993	.400	.017	.023	.991
	σ_b^2	.3	.329	.133	.195	.940	.332	.092	.136	.956
	ψ	−.1	−.099	.021	.021	.949	−.099	.015	.015	.955
	γ_1	−.1	−.103	.121	.122	.959	−.098	.087	.086	.947
	γ_2	.1	.091	.210	.211	.944	.104	.142	.149	.958
	$\Lambda(.9)$	.9	.910	.131	.130	.955	.900	.088	.091	.956
	$\Lambda(1.4)$	1.4	1.421	.209	.206	.934	1.402	.142	.143	.956
	$\Lambda(1.9)$	1.9	1.932	.310	.299	.941	1.899	.205	.207	.948
3	β_1	1.0	.988	.167	.171	.953	.993	.123	.121	.947
	β_2	−.5	−.519	.268	.287	.960	−.516	.189	.203	.967
	β_3	−.2	−.208	.126	.128	.957	−.206	.091	.091	.951
	μ_1	−6.0	−5.844	.353	.483	.967	−5.872	.260	.342	.963
	μ_2	.0	.023	.172	.194	.970	.018	.127	.138	.966
	μ_3	6.0	6.024	.397	.504	.984	6.006	.303	.349	.971
	w_1	.4	.402	.025	.035	.995	.402	.018	.024	.989
	w_2	.3	.298	.025	.034	.986	.298	.017	.024	.985
	σ_b^2	.3	.277	.095	.100	.977	.289	.070	.072	.966
	ψ	−.1	−.102	.014	.015	.955	−.101	.011	.010	.946

Mixture	Par.	True	n=400				n=800			
			Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
	γ_1	-.1	-.103	.121	.120	.955	-.107	.085	.084	.948
	γ_2	.1	.104	.201	.208	.961	.099	.147	.146	.949
	$\Lambda(.9)$	.9	.909	.128	.130	.950	.911	.094	.092	.930
	$\Lambda(1.4)$	1.4	1.421	.202	.207	.960	1.420	.147	.146	.946
	$\Lambda(1.9)$	1.9	1.926	.297	.302	.958	1.929	.220	.213	.946

Table 3

Summary of simulation results of sensitivity for model-misspecification

Par.	TRUE	1 Normal distribution				Mixture of 2 Normal distributions			
		Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
Longitudinal model									
β_1	1.0	.947	.085	.095	.931	.965	.074	.074	.928
β_2	−.5	−.591	.122	.164	.975	−.582	.118	.128	.933
β_3	−.2	−.203	.024	.025	.964	−.202	.024	.025	.961
σ_y^2	.5	.501	.010	.010	.948	.501	.010	.010	.948
Hazards model									
ψ	−.1	−.102	.034	.033	.942	−.101	.034	.033	.943
γ_1	−.1	−.093	.083	.085	.949	−.095	.083	.084	.953
γ_2	.1	.107	.144	.147	.949	.105	.144	.147	.949
$\Lambda(.9)$	.9	.906	.089	.089	.943	.907	.089	.089	.944
$\Lambda(1.4)$	1.4	1.408	.139	.140	.945	1.409	.139	.140	.945
$\Lambda(1.9)$	1.9	1.911	.202	.203	.956	1.911	.202	.203	.956

Par.	TRUE	Mixture of 3 Normal distributions				Mixture of 4 Normal distributions			
		Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
Longitudinal model									
β_1	1.0	.977	.061	.060	.935	.978	.061	.060	.934
β_2	−.5	−.557	.101	.105	.930	−.555	.101	.104	.928
β_3	−.2	−.202	.024	.024	.961	−.202	.024	.024	.959
σ_y^2	.5	.501	.010	.010	.950	.501	.010	.010	.952
Hazards model									
ψ	−.1	−.101	.034	.033	.941	−.101	.034	.033	.940
γ_1	−.1	−.096	.083	.084	.954	−.096	.083	.084	.951
γ_2	.1	.103	.143	.146	.951	.102	.143	.146	.951

Par.	TRUE	Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
$\Lambda(.9)$	.9	.906	.089	.089	.945	.907	.089	.089	.943
$\Lambda(1.4)$	1.4	1.408	.139	.140	.947	1.409	.139	.140	.946
$\Lambda(1.9)$	1.9	1.910	.200	.203	.958	1.911	.202	.203	.956

		Mixture of 5 Normal distributions			
Par.	TRUE	Est.	SSD	ESE	CP
Longitudinal model					
β_1	1.0	.985	.055	.054	.943
β_2	−.5	−.536	.092	.095	.928
β_3	−.2	−.202	.024	.024	.955
σ_y^2	.5	.500	.010	.010	.946
Hazards model					
ψ	−.1	−.102	.034	.033	.945
γ_1	−.1	−.098	.085	.084	.948
γ_2	.1	.108	.144	.146	.952
$\Lambda(.9)$	.9	.904	.090	.089	.944
$\Lambda(1.4)$	1.4	1.403	.137	.139	.947
$\Lambda(1.9)$	1.9	1.905	.200	.202	.954

Table 4

Results from final models of simultaneous and separate analyses for the Quality of Life and survival time for the CHANCE study

Parameter		Simultaneous			Separate[†]		
		Est.	ESE	P-value	Est.	ESE	P-value
HNCS QoL longitudinal model							
Intercept	β_0				1.190	.390	.002
Race (ref= White): African American	β_1	.900	.399	.024	.511	.256	.047
# of 12 oz. beers consumed per week (ref=30 or more)							
– None	β_2	.858	.428	.045	.622	.300	.038
– less than 1	β_3	1.119	.600	.062	.735	.396	.064
– 1 to 4	β_4	1.588	.563	.005	1.268	.326	<.001
– 5 to 14	β_5	1.450	.428	.001	1.018	.279	<.001
– 15 to 29	β_6	1.007	.531	.058	.547	.327	.095
Household income (ref= level1: 0–10K)							
– level2: 20–30K	β_7	– .337	.358	.346	– .328	.258	.204
– level3: 40–50K	β_8	.633	.440	.151	.250	.282	.376
– level4: 60K	β_9	1.960	.509	<.001	1.045	.286	<.001
Radiation therapy (ref= No) : Yes	β_{10}	–1.668	.608	.006	–1.048	.280	<.001
Tumor stage (ref= I)							
– II	β_{11}	– .683	.554	.218	– .352	.330	.286
– III	β_{12}	–2.012	.534	<.001	–1.198	.314	<.001
– IV	β_{13}	–1.826	.507	<.001	–1.057	.277	<.001
# of persons supported by household income	β_{14}	– .388	.140	.006			
BMI	β_{15}	.061	.026	.021			
Time at survey measurement (years)	β_{16}	.354	.093	<.001	.254	.067	<.001
Hazards model							
Random effect coefficient	ψ	– .206	.078	.008			
# of 12 oz. beers consumed per week (ref=30 or more)							
– None	γ_1	– .705	.347	.042			
– less than 1	γ_2	– .156	.393	.692			

Parameter		Simultaneous			Separate[†]		
		Est.	ESE	P-value	Est.	ESE	P-value
− 1 to 4	γ_3	−.712	.385	.064			
− 5 to 14	γ_4	−.991	.348	.004			
− 15 to 29	γ_5	−.579	.370	.117			
Household income (ref= level1: 0–10K)							
− level2: 20–30K	γ_6	−.206	.274	.453	−.219	.263	.406
− level3: 40–50K	γ_7	−.884	.341	.010	−.928	.331	.005
− level4: 60K	γ_8	−1.401	.374	<.001	−1.393	.358	<.001
Tumor stage (ref= I)							
− II	γ_9	−.255	.443	.564	−.295	.435	.498
− III	γ_{10}	.168	.403	.677	.136	.389	.727
− IV	γ_{11}	.950	.306	.002	.914	.295	.002
Total # of medical conditions reported	γ_{12}	.207	.095	.030	.205	.091	.025

P-value for testing σ_b^2 being zero is based on a mixture of 0 and χ^2 distribution with 1 degree of freedom with equal mixing probabilities.

[†] Standard packages assuming one normal component for random effects were used for separate model fit.