Lifetime Data Analysis, 11, 151–174, 2005

Simultaneous Modelling of Survival and Longitudinal Data with an Application to Repeated Quality of Life Measures

DONGLIN ZENG dzeng@bios.unc.edu
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420, USA

JIANWEN CAI cai@bios.unc.edu
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420, USA

Received July 24, 2003; Revised September 22, 2004; Accepted November 15, 2004

Abstract. In biomedical studies, interest often focuses on the relationship between patient's characteristics or some risk factors and both quality of life and survival time of subjects under study. In this paper, we propose a simultaneous modelling of both quality of life and survival time using the observed covariates. Moreover, random effects are introduced into the simultaneous models to account for dependence between quality of life and survival time due to unobserved factors. EM algorithms are used to derive the point estimates for the parameters in the proposed model and profile likelihood function is used to estimate their variances. The asymptotic properties are established for our proposed estimators. Finally, simulation studies are conducted to examine the finite-sample properties of the proposed estimators and a liver transplantation data set is analyzed to illustrate our approaches.

Key words: cox proportional hazard model, EM algorithm, maximum likelihood estimator, mixed model, profile likelihood

1. Introduction

In biomedical studies, researchers are often interested in determining the relationship between patient's characteristics or some risk factors and both quality of life and survival time of subjects. One example, which motivated our research, stems out from a liver transplantation study of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). This study was a 7-year prospective study of 1563 candidates for liver transplantation at three major transplant centers. Among the entire 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 deceased. One goal of this study is to investigate whether factors such as patients' marriage status or disease history affected both quality of life and the risk of death.

In working with studies involving information on both longitudinal measurements and survival time, Robins, Rotnitzky and colleagues proposed to treat the longitudinal measurements as repeated measures subject to informative drop-out due to death and construct inverse probability weighted estimating equations to estimate regression parameters (Robins et al., 1994; Rotnitzky and Robins, 1995; Rotnitzky et al., 1998). In their approach, the drop-out due to death is assumed to be predictable using the observed covariates. Another useful existing approach is based on the quality-adjusted survival time (Zhao and Tsiatis, 1999). However, both approaches do not directly answer our question of interest in the NIDDK liver transplantation study. Moreover, both approaches cannot adjust for latent variables or latent processes, which can be associated with both longitudinal measurements and survival time simultaneously. For example, in the NIDDK liver transplantation study, the quality of donor's liver could potentially affect both the patients' life satisfaction and death. However, the liver's quality is not directly available from the data and such unmeasured factor induces unobserved heterogeneity among the patients. To adjust for such unobserved heterogeneity, one good way is to simultaneously model quality of life and survival time by introducing subject-specific effects in both models.

Joint analysis of repeated measurements and survival time has been intensively studied in recent literature. The models used in such analysis can be categorized into a selection model or a pattern mixture model. Let $\mathbf{Y}$ denote the longitudinal outcomes, for example, quality of life, then $\mathbf{Y}$ are realizations of a latent process $\tilde{\mathbf{Y}}$ measured with errors. Let T denote survival time. A selection model focuses on estimating the distribution of T given $\tilde{\mathbf{Y}}$. Such a selection model has been studied by many authors, for example, Tsiatis et al. (1995), Wulfsohn and Tsiatis (1997), Hu et al. (1998), Huang et al. (2001), and Xu and Zeger (2001a, b). Usually, $\tilde{\mathbf{Y}}$ is modelled as a function of observed covariates and subject-specific random effects; then it is fed into the model of T given $\tilde{\mathbf{Y}}$ as a linear predictor. In the pattern mixture model, a model is assumed for longitudinal outcome $\mathbf{Y}$ conditional on survival time T (Wu and Carroll, 1988; Wu and Bailey, 1989; Hogan and Laird, 1997) and interest focuses on estimating parameters in the model for longitudinal outcome. Although both selection model and pattern-mixture model are useful in some other contexts, they cannot directly answer our question of interest in the NIDDK liver transplant study. The selection model would answer the question regarding how one's quality of life affects death and the pattern-mixture model would describe the pattern of quality of life given one's death time, while we are interested in finding which factor or treatment can simultaneously improve the patient's quality of life and reduce the risk of death.

Simultaneous modelling will serve the purpose. In this paper, we model both the process for quality of life, $\mathbf{Y}$, and survival time, T, given observed covariates $\mathbf{X}$. These two outcomes are modelled by observed covariates as well as by some unobserved factors. Heterogeneity caused by unobserved factors is represented using individual random effects in both models. It is noted that our approach is different from either selection model or pattern-mixture model, although mathematically, all three models can be regarded as different ways of writing the distribution of $(T, \mathbf{Y})$ given covariates. Xu and Zeger (2001b) proposed a similar model. However, in their model, a

common latent process is shared by both $\mathbf{Y}$ and T, while our model allows individual random effects to affect the quality of life and the survival time very differently.

Our estimation approach is likelihood-based. The EM-algorithm is employed for parameter estimation and profile likelihood function is used for variance estimation. Particularly, we suggest an efficient algorithm based on the EM-algorithm to calculate the profile likelihood function. Furthermore, we provide the asymptotic results for the maximum likelihood estimators of the parameters in the joint models.

The rest of the paper is organized as follows. A general framework and model assumptions for modelling quality of life and survival time are proposed in Section 2. Section 3 describes the EM-algorithms for the maximum likelihood estimation of the parameters. We then propose an innovative simple method in computing the asymptotic variances. Section 4 gives the asymptotic results. Small-sample properties of the proposed estimators are examined via simulation studies in Section 5. In Section 6, we apply our proposed method to the liver transplantation data. Further discussion and generalization are given in Section 7. The outline of the proofs of the asymptotic results are provided in Appendices A.2 and A.3.

2. General Model

We model quality of life and survival time through parametric and semiparametric models, respectively. We assume a linear mixed effect model for the longitudinal outcomes of quality of life and assume a multiplicative hazards model for survival time. In both models, observed covariates, such as subjects' baseline variables, disease status, environmental information over time, are included as predictors and they are assumed to be either time-independent or external time-dependent variables. Unobserved factors enter the models as subject-specific random effects so as to account for unobserved heterogeneity.

Specifically, for subject i, given $T > t$ and the observed history till time t, the quality of life at time t, denoted by $Y_i(t)$, follows

$$Y_i(t) = \mathbf{X}_i(t)\boldsymbol{\beta} + \tilde{\mathbf{X}}_i(t)\mathbf{a}_i + \epsilon_i(t), \tag{1}$$

where $\mathbf{X}_i(t)$ and $\tilde{\mathbf{X}}_i(t)$ are the row vectors of the observed covariates, $\epsilon_i(t)$ is a white noise process with mean zero and variance σ_y^2, $\mathbf{a}_i$ denotes a vector of subject-specific random effect of dimension k_0 following a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_a$, and $\boldsymbol{\beta}$ is a column vector of coefficients for $\mathbf{X}_i(t)$. The random effect $\mathbf{a}_i$ reflects the unobserved heterogeneity and is allowed to differ for different levels of covariates $\tilde{\mathbf{X}}_i(t)$. Additionally, in model (1), $\mathbf{X}_i(t)$ and $\tilde{\mathbf{X}}_i(t)$ can be completely different or share some components. For example, in a clinical trial with two treatment arms, $\mathbf{X}_i(t)$ can contain both column of 1's and treatment status for subject i; while $\tilde{\mathbf{X}}_i(t)$ is a column of 1's, which implies a random intercept is used in the model; or, $\tilde{\mathbf{X}}_i(t)$ is the same as $\mathbf{X}_i(t)$ which allows the two different treatment arms to have different random effects. For the survival time T_i, the conditional hazard rate function given the observed covariates, the observed

history before time t, and random effect $\mathbf{a}_i$, is assumed to follow a multiplicative hazards model:

$$\lambda(t)e^{\tilde{\mathbf{W}}_i(t)(\boldsymbol{\phi}\circ\mathbf{a}_i)+\mathbf{W}_i(t)\boldsymbol{\gamma}}, \tag{2}$$

where $\mathbf{W}_i(t)$ and $\tilde{\mathbf{W}}_i(t)$ are the row vectors of the observed covariates and may share the same components, $\boldsymbol{\phi}$ is a vector of parameters, $\lambda(t)$ is the baseline hazard rate function, and $\boldsymbol{\gamma}$ is a column vector of coefficients for $\mathbf{W}_i(t)$. Here, for any vectors v_1 and v_2 of the same dimension, $v_1 \circ v_2$ denotes the component-wise product.

In models (1) and (2), one may use $\mathbf{a}_i$ to denote the random effect in the model for the quality of lifetime but use another different random effect, say $\mathbf{b}_i$, to denote the subject-specific effect in the hazard model. Then under the assumptions that $\mathbf{a}_i$ and $\mathbf{b}_i$ are jointly normal, we can write $\mathbf{b}_i = \boldsymbol{\theta}'\mathbf{a}_i + \mathbf{c}_i$, where $\boldsymbol{\theta}$ is a constant vector and $\mathbf{c}_i$ is independent of $\mathbf{a}_i$. In fact, $\mathbf{c}_i$ represents the subject-specific effect which affects the survival time but not the quality of life. However, $\mathbf{c}_i$ is usually set to be zero for the following two reasons: first, $\mathbf{c}_i$ is not identifiable from the observed data; second, in practice, any unobserved factors affecting the risk of death are believed to affect the quality of life as well. When $\mathbf{c}_i = 0$, the hazards model for the survival time becomes model (2) after suitable reparameterization. In other words, although sharing the same random effect in models (1) and (2) appears to be restrictive, models (1) and (2) are in fact very general.

In models (1) and (2), the parameter $\boldsymbol{\phi}$ characterizes the dependence between the quality of life and the survival time. In particular, when $\boldsymbol{\phi} = 0$, the dependence between the survival time and the quality of life can be fully attributed to the observed covariates; when $\boldsymbol{\phi} \neq 0$, it implies that such dependence may also be due to some latent variables.

We suppose that the survival time is possibly right censored and that right-censored time is completely random. We let N_i denote the number of the observed quality of life measurements for subject i and assume that N_i is not informative about parameters of interest. Furthermore, we write the observed data from n subjects as

$$(N_i, Y_i^j, \mathbf{X}_i^j, \tilde{\mathbf{X}}_i^j),\ j = 1, \ldots, N_i, \quad i = 1, \ldots, n$$

and

$$(Z_i, \Delta_i, \{(\mathbf{W}_i(t), \tilde{\mathbf{W}}_i(t)) : t \leq Z_i\}), \quad i = 1, \ldots, n,$$

where for subject i, $(Y_i^j, \mathbf{X}_i^j, \tilde{\mathbf{X}}_i^j)$ is the j-th observation of $(Y_i(t), \mathbf{X}_i(t), \tilde{\mathbf{X}}_i(t))$, C_i is the right-censoring time, $Z_i = min(T_i, C_i)$, and $\Delta_i = I(T_i \leq C_i)$. We are interested in estimating and making inference on the parameters $\boldsymbol{\theta} = (\sigma_y, \boldsymbol{\Sigma}_a, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\gamma})$ and the baseline cumulative hazard function $\Lambda(t) = \int_0^t \lambda(s)ds$.

3. Maximum Likelihood Estimation

We integrate over the random effects in the joint models (1) and (2) and note that by the Gaussian property of $\mathbf{Y}_i(t)$, the observed quality of life, $\mathbf{Y}_i$, follows a multivariate normal distribution given random effects. The observed likelihood function for the parameters $(\boldsymbol{\theta}, \Lambda)$ can be expressed as

$$\prod_{i=1}^{n}\int_{\mathbf{a}}\Big[(2\pi\sigma_y^2)^{-N_i/2}\exp\{-(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}-\tilde{\mathbf{X}}_i\mathbf{a})'(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}-\tilde{\mathbf{X}}_i\mathbf{a})/2\sigma_y^2\}$$
$$\lambda(Z_i)^{\Delta_i}\exp\{\Delta_i(\tilde{\mathbf{W}}_i(Z_i)(\boldsymbol{\phi}\circ\mathbf{a})+\mathbf{W}_i(Z_i)\boldsymbol{\gamma})$$
$$-\int_0^{Z_i}e^{\tilde{\mathbf{W}}_i(s)(\boldsymbol{\phi}\circ\mathbf{a})+\mathbf{W}_i(s)\boldsymbol{\gamma}}d\Lambda(s)\}(2\pi)^{-k_0/2}|\boldsymbol{\Sigma}_a|^{-1/2}\exp\{-\mathbf{a}'\boldsymbol{\Sigma}_a^{-1}\mathbf{a}/2\}\Big]d\mathbf{a},$$

where $\mathbf{Y}_i$ denotes the vector of $(Y_i^1,\ldots,Y_i^{N_i})'$, $\mathbf{X}_i$ denotes the matrix of $((\mathbf{X}_i^1)',\ldots,(\mathbf{X}_i^{N_i})')'$, $\tilde{\mathbf{X}}_i$ denotes $((\tilde{\mathbf{X}}_i^1)',\ldots,(\tilde{\mathbf{X}}_i^{N_i})')'$, and k_0 is the dimension of $\mathbf{a}$.

Our estimation method is to calculate the maximum likelihood estimates for $(\boldsymbol{\theta},\Lambda)$ over a set in which $\boldsymbol{\theta}$ is in a bounded set and Λ belongs to a space consisting of all the increasing functions with $\Lambda(0)=0$. It is clear that the maximum likelihood estimate for Λ can be chosen as a step function with jumps only at the observed failure times. Specifically, the EM algorithm can be used to calculate the maximum likelihood estimates. In the EM algorithm, $\mathbf{a}_i$ is considered as the missing statistics for $i=1,\ldots,n$. Therefore, the M-step solves the conditional score equation from the complete data given the observations, where the conditional expectation can be evaluated in the E-step. We iterate between E-step and M-step until the estimates converge. The details of the EM algorithms can be found in the Appendix A.1. We denote the final maximum likelihood estimate for $(\boldsymbol{\theta},\Lambda)$ by $(\hat{\boldsymbol{\theta}},\hat{\Lambda})$.

The profile likelihood function is used to obtain the variance estimate for $\hat{\boldsymbol{\theta}}$. We define the logarithm of the profile likelihood function of $\boldsymbol{\theta}$ as $pl_n(\boldsymbol{\theta})=\max_\Lambda n^{-1}\sum_{i=1}^n q_i(\boldsymbol{\theta},\Lambda)$ where $q_i(\boldsymbol{\theta},\Lambda), i=1,\ldots,n$, is the logarithm of the observed likelihood function for the i-th subject. Then the second-order numerical difference of $pl_n(\boldsymbol{\theta})$ at $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$ can be used to approximate the asymptotic variance of $\hat{\boldsymbol{\theta}}$. In particular,

$$-\frac{pl_n(\hat{\boldsymbol{\theta}}+h_n\mathbf{e})-2pl_n(\hat{\boldsymbol{\theta}})+pl_n(\hat{\boldsymbol{\theta}}-h_n\mathbf{e})}{h_n^2}$$

approximates $\mathbf{e}'I_{\boldsymbol{\theta}}\mathbf{e}$ for any norm-1 vector $\mathbf{e}$ and any constant $h_n=O(1/\sqrt{n})$. $I_{\boldsymbol{\theta}}$ is the efficient information matrix for $\boldsymbol{\theta}$, which is also equal to the inverse of the asymptotic variance of $\sqrt{n}\hat{\boldsymbol{\theta}}$. We denote the estimated standard error for $\hat{\boldsymbol{\theta}}$ based on this second-order numerical difference by se(II). An alternative estimate of the variance is based on the equality $I_{\boldsymbol{\theta}}=E[l_{\boldsymbol{\theta}}^* l_{\boldsymbol{\theta}}^{*\prime}]$, where $l_{\boldsymbol{\theta}}^*$ denotes the efficient score function for $\boldsymbol{\theta}$. Then we can estimate $\mathbf{e}'I_{\boldsymbol{\theta}}\mathbf{e}$ using the expression:

$$\frac{1}{nh_n^2}\sum_{i=1}^{n}[q_i(\hat{\boldsymbol{\theta}}+h_n\mathbf{e},\hat{\Lambda}_{\hat{\boldsymbol{\theta}}+h_n\mathbf{e}})-q_i(\hat{\boldsymbol{\theta}},\hat{\Lambda})]^2,$$

where $\hat{\Lambda}_{\boldsymbol{\theta}}$ indicates the cumulative hazard function which maximizes the observed log-likelihood function for a given $\boldsymbol{\theta}$. The estimated standard error based on this first-order numerical difference is denoted by se(I). We use both se(I) and se(II) to estimate the variance of $\hat{\boldsymbol{\theta}}$ in the subsequent analysis.

In the above variance estimation, we need to compute $\hat{\Lambda}_{\boldsymbol{\theta}}$ for a fixed $\boldsymbol{\theta}$ in the neighborhood of $\hat{\boldsymbol{\theta}}$. Due to the complicated expression of the likelihood function in the presence of the missing data, calculating $\hat{\Lambda}_{\boldsymbol{\theta}}$ via direct maximization is not efficient. Interestingly, however, such calculation can be done via the EM algorithm as well. The only change from the previous EM algorithm is that we always hold $\boldsymbol{\theta}$ fixed in both E-step and M-step so the only updated parameters in each iteration is Λ. When the iteration converges, the final Λ approximates $\hat{\Lambda}_{\boldsymbol{\theta}}$ so the profile likelihood can be calculated by evaluating the observed likelihood function at $(\boldsymbol{\theta}, \hat{\Lambda}_{\boldsymbol{\theta}})$. We abbreviate such an EM-based algorithm as the PEME algorithm (partial expectation, maximization and evaluation). A general procedure and properties of PEME will be given in Appendix A.4. Our simulation studies indicate that the PEME algorithm is much faster than the direct optimization.

4. Asymptotic Properties

The asymptotic properties of $(\hat{\boldsymbol{\theta}}, \hat{\Lambda})$ are given in this section. We need the following assumptions:

(A1) The true parameter $\boldsymbol{\theta}_0 = (\sigma_{0y}, \boldsymbol{\Sigma}_{0a}, \boldsymbol{\beta}_0, \boldsymbol{\phi}_0, \boldsymbol{\gamma}_0)$ belongs to a known compact set, which lies in the interior of the domain for $\boldsymbol{\theta}$.
(A2) The true baseline hazard rate function $\lambda_0(t)$ is positive and bounded in $[0, \tau]$.
(A3) $P(C \geq \tau) = P(C = \tau) > 0$, i.e., by the end of the study, some proportion of the subjects will still be alive and censored at τ.
(A4) $P(N > k_0) > 0$. In other words, some proportion of the subjects have at least k_0 longitudinal observations (k_0 is the number of random effects in the mixed model). Moreover, $P(N \leq n_0) = 1$ for some integer n_0.
(A5) $\tilde{\mathbf{W}}$ is time-independent.
(A6) With positive probability, $\mathbf{X}'\mathbf{X}$, $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$,$\tilde{\mathbf{W}}'\tilde{\mathbf{W}}$ are full rank. Moreover, if for any t, $\mathbf{W}(t)\boldsymbol{\gamma} = c_0(t)$ for a deterministic function $c_0(t)$, then $\boldsymbol{\gamma} = 0$ and $c_0(t) = 0$.

Under the assumptions (A1)–(A6), the maximum likelihood estimator $(\hat{\boldsymbol{\theta}}, \hat{\Lambda})$ is consistent under the product norm of the Euclidean distance and the supreme norm on $[0, \tau]$. That is,

$$|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| + \sup_{t \in [0,\tau]} |\hat{\Lambda}(t) - \Lambda_0(t)| \to 0, \quad a.s.$$

The outline of the proof for the consistency is given in the Appendix A.2. Additionally, in the Appendix A.3., we show that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \hat{\Lambda} - \Lambda_0)$ weakly converges to a Gaussian random element in $R^k \times l^\infty[0, \tau]$, where k is the dimension of $\boldsymbol{\theta}$ and $l^\infty[0, \tau]$ is the normed space containing all the bounded functions in $[0, \tau]$. Especially, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ weakly converges to a multivariate normal distribution with mean zero; and $\{\sqrt{n}(\hat{\Lambda}(t) - \Lambda_0(t)) : t \in [0, \tau]\}$ weakly converges to a Gaussian process indexed by t.

When both the consistency and the asymptotic normality are obtained, we can verify all the conditions for the profile likelihood theory given in Murphy and van der Vaart (2000). Specifically, the logarithm of the profile likelihood function, $pl_n(\boldsymbol{\theta})$, has an approximate parabolic shape around the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ and its curvature approximates the negative information matrix for $\boldsymbol{\theta}_0$. Conclusively, $-h_n^{-2}\left\{pl_n(\hat{\boldsymbol{\theta}} + h_n\mathbf{e}) - 2pl_n(\hat{\boldsymbol{\theta}}) + pl_n(\hat{\boldsymbol{\theta}} - h_n\mathbf{e})\right\}$ converges to $\mathbf{e}^T I(\boldsymbol{\theta}_0)\mathbf{e}$ for any norm-1 vector $\mathbf{e}$ and constant $h_n = O(n^{-1/2})$.

5. Simulation Study

Two sets of simulations are performed in this section. In the first set of simulations, we consider a binary time-independent covariate for both the model of quality of life and the survival model and a random intercept is used in both models. Specifically, we have $Y_i^j = \beta_0 + \beta_1 X_i + a_i + \epsilon_{ij}$ for $j = 1, \ldots, N_i$ and $h(t|X_i, a_i) = \lambda_0(t)\exp\{\phi a_i + \gamma X_i\}$, where $a_i \sim N(0, \sigma_a^2)$, $\epsilon_{ij} \sim N(0, \sigma_y^2)$, and X_i is simulated from a Bernoulli distribution with success probability being 0.5. Censoring time is generated from an exponential distribution with a constant hazard rate C_0. The data for quality of life are generated for every 0.2 unit of time. The parameters in the two models are chosen as $\beta_0 = -1, \beta_1 = 1, \sigma_y = 0.4, \sigma_a = 1, \phi = -0.2, \gamma = 0.4, \lambda_0(t) = 1$. We consider different sample sizes ($n = 50, 100, 200$) and censoring proportions (0%, 30%, 50%). The latter two censoring rates correspond to $C_0 = 1$ and $C_0 = 4$, respectively. For most of the subjects, the number of longitudinal observations (N_i) is around 2–4. Each simulation is repeated 500 times. The results of the maximum likelihood estimates for $\boldsymbol{\theta}$ and their respective standard error estimates using se(I) and se(II) are reported in Table 1.

In our second set of simulations, the covariates include a time-dependent covariate, which is indeed a time-independent covariate modelled to have a time-varying effect. Specifically, we generate the data from the models $Y_i^j = \beta_0 X_i + \beta_1 X_i t_{ij} + a_i + \epsilon_{ij}$ and $h(t|a_i, X_i) = \lambda_0(t)\exp\{\phi a_i + \boldsymbol{\gamma} X_i t\}$, where $a_i \sim N(0, \sigma_a^2)$, $\epsilon_{ij} \sim N(0, \sigma_y^2)$, X_i is simulated from a Bernoulli distribution with success probability being 0.5, and $t_{ij} = 0.2(j-1)$ is the time for observing Y_i^j. Censoring time is generated from an exponential distribution with constant hazard rate C_0. The choice of the parameters is $\beta_0 = -1, \beta_1 = 1, \sigma_y = 0.4, \sigma_a = 1$, $\phi = -0.5, \gamma = 0.2, \lambda_0(t) = 1$. We consider different sample sizes ($n = 50, 100, 200$) and censoring proportions (0%, 30%, 50%). The latter two censoring rates correspond to $C_0 = 1$ and $C_0 = 3.5$, respectively. For most of the subjects, the number of longitudinal observations (N_i) is around 2–4. We present the simulation results in Table 2, where each simulation is repeated 500 times. Both simulations were coded and conducted using Splus 2000.

In both Tables 1 and 2, "θ_0" column gives the true values of each parameter; the averages of the maximum likelihood estimates from the EM algorithm are in "MLE" column and the sample standard deviations from 500 simulations are reported in the column "emp.se"; "se(I)" is the average of the 500 standard error estimates based on the first-order numerical difference of the profile likelihood function and "se(II)" is the average of the standard error estimates based on the

Table 1. Results from simulation study with time-independent covariate.

n	cen%	$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	MLE	emp.se.	se(I)	se(II)	CP(I)	CP(II)
$n = 50$	0%	β_0	−1.00	−0.989	0.193	0.218	0.201	0.960	0.946
		β_1	1.00	0.987	0.278	0.310	0.286	0.970	0.952
		σ_y^2	0.16	0.159	0.017	0.019	0.016	0.966	0.946
		σ_a^2	1.00	0.965	0.208	0.241	0.206	0.926	0.898
		ϕ	−0.20	−0.205	0.177	0.198	0.162	0.974	0.936
		γ	0.40	0.408	0.309	0.341	0.309	0.970	0.944
	30%	β_0	−1.00	−0.987	0.209	0.217	0.202	0.956	0.948
		β_1	1.00	0.990	0.294	0.311	0.287	0.966	0.946
		σ_y^2	0.16	0.160	0.021	0.024	0.020	0.958	0.948
		σ_a^2	1.00	0.967	0.219	0.244	0.208	0.932	0.904
		ϕ	−0.20	−0.223	0.202	0.222	0.192	0.974	0.948
		γ	0.40	0.411	0.366	0.390	0.360	0.966	0.952
	50%	β_0	−1.00	−0.987	0.195	0.221	0.204	0.966	0.956
		β_1	1.00	0.983	0.294	0.312	0.289	0.952	0.930
		σ_y^2	0.16	0.159	0.025	0.030	0.025	0.954	0.938
		σ_a^2	1.00	0.971	0.210	0.249	0.211	0.944	0.920
		ϕ	−0.20	−0.213	0.253	0.264	0.229	0.964	0.926
		γ	0.40	0.404	0.479	0.463	0.432	0.950	0.934
$n = 100$	0%	β_0	−1.00	−1.004	0.143	0.149	0.143	0.958	0.950
		β_1	1.00	1.001	0.205	0.212	0.203	0.950	0.942
		σ_y^2	0.16	0.160	0.012	0.013	0.012	0.960	0.946
		σ_a^2	1.00	0.973	0.146	0.161	0.147	0.940	0.926
		ϕ	−0.20	−0.208	0.119	0.123	0.110	0.956	0.940
		γ	0.40	0.394	0.227	0.226	0.213	0.958	0.942
	30%	β_0	−1.00	−0.995	0.141	0.150	0.144	0.964	0.950
		β_1	1.00	1.003	0.199	0.213	0.205	0.966	0.960
		σ_y^2	0.16	0.161	0.014	0.015	0.014	0.962	0.956
		σ_a^2	1.00	0.980	0.153	0.163	0.149	0.948	0.922
		ϕ	−0.20	−0.213	0.133	0.141	0.130	0.970	0.952
		γ	0.40	0.434	0.239	0.260	0.250	0.964	0.950
	50%	β_0	−1.00	−0.999	0.144	0.152	0.146	0.958	0.954
		β_1	1.00	1.014	0.201	0.215	0.207	0.958	0.954
		σ_y^2	0.16	0.159	0.018	0.020	0.018	0.948	0.934
		σ_a^2	1.00	0.984	0.159	0.167	0.153	0.938	0.916
		ϕ	−0.20	−0.213	0.164	0.167	0.156	0.960	0.960
		γ	0.40	0.374	0.284	0.307	0.296	0.974	0.966
$n = 200$	0%	β_0	−1.00	−0.993	0.092	0.104	0.102	0.970	0.968
		β_1	1.00	0.994	0.143	0.148	0.145	0.956	0.950
		σ_y^2	0.16	0.160	0.008	0.008	0.008	0.950	0.946
		σ_a^2	1.00	0.984	0.103	0.109	0.105	0.952	0.938
		ϕ	−0.20	−0.198	0.076	0.081	0.076	0.970	0.952
		γ	0.40	0.404	0.147	0.154	0.149	0.954	0.950
	30%	β_0	−1.00	−1.002	0.108	0.105	0.103	0.940	0.938
		β_1	1.00	0.999	0.149	0.149	0.146	0.942	0.942
		σ_y^2	0.16	0.160	0.010	0.010	0.010	0.956	0.952
		σ_a^2	1.00	0.993	0.105	0.111	0.107	0.944	0.942
		ϕ	−0.20	−0.204	0.087	0.094	0.090	0.978	0.970
		γ	0.40	0.407	0.180	0.177	0.174	0.954	0.940

Table 1. (Continued).

n	cen%	$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	MLE	emp.se.	se(I)	se(II)	CP(I)	CP(II)
	50%	β_0	−1.00	−1.004	0.101	0.105	0.103	0.964	0.962
		β_1	1.00	1.005	0.142	0.149	0.147	0.950	0.948
		σ_y^2	0.16	0.160	0.013	0.013	0.013	0.948	0.938
		σ_a^2	1.00	0.991	0.110	0.113	0.108	0.946	0.942
		ϕ	−0.20	−0.204	0.109	0.112	0.107	0.950	0.940
		γ	0.40	0.422	0.193	0.211	0.207	0.964	0.962

second-order numerical difference of the profile likelihood function. In both standard error estimation, the profile likelihood functions are calculated using the PEME algorithms; "$CP(I)$" and "$CP(II)$" are the coverage proportions of the 95% nominal confidence intervals by using the corresponding se(I) and se(II).

From Tables 1 and 2, we can see that even for a small sample size ($n = 50$), the bias of the estimates from EM algorithm is negligible. Both se(I) and se(II) calculated from the PEME algorithms are close to the sample standard deviations from the 500 estimates and the 95% confidence interval coverage rates are close to 0.95. In addition, the simulations show that as the sample size (n) increases, the variances of the estimators decrease. As the right-censoring becomes heavier, the estimated variances for the parameters in the survival models increase while the estimators of the variances for the coefficients in the longitudinal model are fairly robust to this change. This might be due to that conditional on random effects, the variance estimate for the coefficients in the longitudinal model depends on the number of the observations obtained by the ending time and it is not relevant to whether the ending time is a survival time or a censoring time.

To study the sensitivity of the parameter estimates to the mis-specification in the random component structure, we conduct an additional simulation study, where the quality of life data are generated from the mixed model in our second simulation and the survival time is generated from the following proportional hazards model

$$h(t|a_i, b_i, X_i) = \lambda_0(t) \exp\{\phi a_i + b_i + \boldsymbol{\gamma} X_i t\},$$

where b_i is another random effect independent of a_i and b_i follows a normal distribution with mean zero and variance σ_b^2. Thus, the only difference from our second simulation is that there is an extra random effect b_i affecting the survival time but not the quality of life. We calculate the maximum likelihood estimates by ignoring b_i; in other words, we mis-specify the model for the survival time. We wish to study how the parameter estimates vary with the values of σ_b^2. We use the same parameter setting as our second simulation and we generate right-censoring time from an exponential distribution with constant hazard rate 1. Moreover, the values of σ_b^2 vary from 0 to 2. The results from 500 repetitions with sample size 100 are given in Table 3. From Table 3, we find that the regression coefficients in the model for the quality of life are very robust to this extra random effect; the regression coefficient for the

Table 2. Results from simulation study with time-dependent covariate.

n	cen%	$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	MLE	emp.se	se(I)	se(II)	CP(I)	CP(II)
$n = 50$	0%	β_0	−1.00	−0.999	0.081	0.101	0.092	0.970	0.954
		β_1	1.00	0.999	0.032	0.045	0.033	0.980	0.940
		σ_y^2	0.16	0.158	0.016	0.019	0.016	0.974	0.940
		σ_a^2	1.00	0.996	0.217	0.249	0.213	0.956	0.926
		ϕ	−0.50	−0.529	0.180	0.216	0.178	0.980	0.946
		γ	0.20	0.230	0.389	0.416	0.320	0.972	0.928
	30%	β_0	−1.00	−0.997	0.077	0.100	0.092	0.982	0.978
		β_1	1.00	1.001	0.063	0.078	0.061	0.980	0.940
		σ_y^2	0.16	0.159	0.019	0.023	0.020	0.960	0.942
		σ_a^2	1.00	0.971	0.213	0.246	0.209	0.940	0.918
		ϕ	−0.50	−0.537	0.226	0.249	0.211	0.976	0.938
		γ	0.20	0.225	0.649	0.751	0.626	0.976	0.956
	50%	β_0	−1.00	−0.994	0.091	0.102	0.093	0.954	0.940
		β_1	1.00	0.994	0.104	0.125	0.101	0.966	0.950
		σ_y^2	0.16	0.159	0.025	0.028	0.024	0.946	0.924
		σ_a^2	1.00	0.979	0.208	0.249	0.213	0.954	0.932
		ϕ	−0.50	−0.539	0.252	0.286	0.248	0.976	0.954
		γ	0.20	0.220	1.213	1.364	1.160	0.978	0.964
$n = 100$	0%	β_0	−1.00	−1.001	0.055	0.068	0.065	0.978	0.968
		β_1	1.00	1.000	0.022	0.027	0.022	0.968	0.952
		σ_y^2	0.16	0.160	0.011	0.013	0.011	0.954	0.946
		σ_a^2	1.00	0.992	0.153	0.163	0.150	0.942	0.924
		ϕ	−0.50	−0.516	0.122	0.138	0.121	0.966	0.948
		γ	0.20	0.220	0.243	0.238	0.205	0.956	0.936
	30%	β_0	−1.00	−0.999	0.057	0.068	0.065	0.980	0.960
		β_1	1.00	1.000	0.043	0.048	0.042	0.970	0.950
		σ_y^2	0.16	0.160	0.015	0.015	0.014	0.950	0.934
		σ_a^2	1.00	0.979	0.148	0.160	0.149	0.936	0.926
		ϕ	−0.50	−0.515	0.154	0.156	0.143	0.952	0.944
		γ	0.20	0.211	0.447	0.460	0.412	0.956	0.938
	50%	β_0	−1.00	−0.999	0.061	0.070	0.067	0.962	0.960
		β_1	1.00	0.998	0.063	0.077	0.070	0.972	0.968
		σ_y^2	0.16	0.159	0.017	0.019	0.017	0.960	0.950
		σ_a^2	1.00	0.997	0.163	0.165	0.153	0.934	0.912
		ϕ	−0.50	−0.508	0.178	0.179	0.166	0.956	0.956
		γ	0.20	0.241	0.762	0.824	0.749	0.976	0.958
$n = 200$	0%	β_0	−1.00	−1.003	0.036	0.047	0.046	0.982	0.984
		β_1	1.00	0.999	0.016	0.017	0.016	0.954	0.940
		σ_y^2	0.16	0.160	0.008	0.008	0.008	0.936	0.928
		σ_a^2	1.00	0.991	0.107	0.110	0.106	0.944	0.930
		ϕ	−0.50	−0.506	0.088	0.090	0.084	0.946	0.944
		γ	0.20	0.220	0.145	0.152	0.137	0.956	0.942
	30%	β_0	−1.00	−1.001	0.037	0.048	0.047	0.982	0.982
		β_1	1.00	1.003	0.031	0.031	0.029	0.938	0.940
		σ_y^2	0.16	0.160	0.010	0.010	0.010	0.944	0.942
		σ_a^2	1.00	0.993	0.107	0.110	0.107	0.938	0.940
		ϕ	−0.50	−0.508	0.099	0.103	0.098	0.962	0.954
		γ	0.20	0.218	0.273	0.291	0.273	0.970	0.950

Table 2. (Continued).

n	cen%	$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	MLE	emp.se	se(I)	se(II)	CP(I)	CP(II)
	50%	β_0	−1.00	−0.998	0.039	0.048	0.047	0.982	0.972
		β_1	1.00	1.002	0.052	0.053	0.030	0.932	0.916
		σ_y^2	0.16	0.159	0.013	0.013	0.012	0.952	0.928
		σ_a^2	1.00	0.998	0.105	0.112	0.108	0.956	0.964
		ϕ	−0.50	−0.511	0.118	0.120	0.114	0.954	0.948
		γ	0.20	0.212	0.500	0.533	0.506	0.974	0.950

survival time are robust when σ_b^2 is small; however, when σ_b^2 is large, its bias becomes significant, although the inference is correct. Furthermore, the parameter estimate for ϕ is sensitive to the change of σ_b^2 and the bias is large when $\sigma_b^2 = 2$. This sensitivity analysis indicates that even when some extra latent effect may affect the survival time, the inference for the covariate coefficients in the model for quality of life can be good; however, the inference for the parameters in the model for the survival time may be incorrect.

6. Application

We apply our proposed method to the data from the NIDDK liver transplantation study. We select one measurement of the patients' quality of life as the longitudinal outcome. This measurement 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 treat this score as continuous in our analysis. There are 582 patients with 1382 complete post-transplantation quality of life scores and the range of the number of observations for each patient is 1–5. The censoring rate is 87%. The solid lines in Figures 1 and 2 are the average dissatisfaction scores and the Kaplan–Meier curves for the Caucasian patients with different marriage status and the ascites history. The plot shows that the life dissatisfaction scores did not change over time for all the groups and the patients who were married and had ascites appeared to be slightly more satisfied with their life; some difference is observed in the survival curves between the single patients and the marriage ones.

We are interested in studying which variables, including age, race, gender, marriage status and disease history of ascites, bone disease, cholangitis and edema, predict the life satisfaction score or the risk of death or both. Let $\mathscr{H}(t)$ denote the failure, censoring, and covariates information up to time t. The joint models for the life satisfaction score and the survival time are:

$$\begin{aligned} Y(t) =& \beta_0 + \beta_1 I(Center = 2) + \beta_2 I(Center = 3) + \beta_3 age + \beta_4 sex \\ &+ \beta_5 race + \beta_6 marriage + \beta_7 ASC + \beta_8 BD + \beta_9 CHOL + \beta_{10} EDE \\ &+ \beta_{11} BMI + \beta_{12} t + a_0 + a_1 t + \epsilon(t), \end{aligned}$$

Table 3. Results from sensitivity analysis in simultaneous modelling with $n = 100$.

σ_b^2	$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	MLE	emp.se	se(I)	se(II)	CP(I)	CP(II)
0	β_0	−1.00	−0.999	0.057	0.068	0.065	0.980	0.960
	β_1	1.00	1.000	0.043	0.048	0.042	0.970	0.950
	σ_y^2	0.16	0.160	0.015	0.015	0.014	0.950	0.934
	σ_a^2	1.00	0.979	0.148	0.160	0.149	0.936	0.926
	ϕ	−0.50	−0.515	0.154	0.156	0.143	0.952	0.944
	γ	0.20	0.211	0.447	0.460	0.412	0.956	0.938
0.25	β_0	−1.00	−1.000	0.056	0.069	0.066	0.978	0.974
	β_1	1.00	0.999	0.041	0.046	0.040	0.966	0.960
	σ_y^2	0.16	0.159	0.014	0.015	0.014	0.946	0.930
	σ_a^2	1.00	0.994	0.155	0.162	0.151	0.954	0.948
	ϕ	−0.50	−0.468	0.142	0.153	0.140	0.950	0.942
	γ	0.20	0.160	0.451	0.462	0.418	0.948	0.938
1	β_0	−1.00	−0.997	0.055	0.069	0.066	0.970	0.968
	β_1	1.00	1.001	0.038	0.044	0.038	0.966	0.942
	σ_y^2	0.16	0.159	0.014	0.015	0.014	0.944	0.936
	σ_a^2	1.00	1.000	0.157	0.165	0.152	0.940	0.926
	ϕ	−0.50	−0.360	0.137	0.147	0.138	0.834	0.804
	γ	0.20	0.178	0.481	0.494	0.448	0.964	0.944
2	β_0	−1.00	−1.003	0.057	0.069	0.066	0.962	0.956
	β_1	1.00	1.002	0.034	0.042	0.036	0.972	0.954
	σ_y^2	0.16	0.160	0.014	0.015	0.014	0.936	0.930
	σ_a^2	1.00	1.005	0.158	0.166	0.153	0.956	0.954
	ϕ	−0.50	−0.301	0.138	0.146	0.137	0.704	0.670
	γ	0.20	0.097	0.485	0.526	0.478	0.976	0.956

(σ_b^2 is the variance of the extra random effect in the proportional hazards model).

$$\begin{aligned} h(t|\mathscr{H}(t)) =& \lambda(t)\exp\{\phi a_0 + \gamma_1 I(Center = 2) + \gamma_2 I(Center = 3) + \gamma_3 age \\ &+ \gamma_4 sex + \gamma_5 race + \gamma_6 marriage + \gamma_7 ASC + \gamma_8 BD \\ &+ \gamma_9 CHOL + \gamma_{10} EDE + \gamma_{11} BMI\}, \end{aligned}$$

where $\epsilon(t)$ is normally distributed with mean zero and variance σ_y^2, (a_0, a_1) follows a bivariate normal distribution with mean zero. $Y(t)$ is the dissatisfaction score at time t, "$I(Center = 2)$" and "$I(Center = 3)$" are indicators for the center identity with center 1 being the reference group, "*age*" is the age of the patient at liver transplantation, *BMI* is the body mass index, and the other covariates are all binary: "*sex*" (0: male, 1: female), "*race*" (0: Caucasian, 1: non-Caucasian), "*marriage*" (0: single, 1: married), "*ASC*" (0: never had ascites, 1: ever had ascites), "*BD*" (0: never had bone disease, 1: ever had bone disease), "*CHOL*" (0: never had cholangitis, 1: ever had cholangitis), "*EDE*" (0: never had edema, 1: ever had edema). The shared random element in both models is the random intercept a_0, which accounts for the unobserved heterogeneity in explaining the dependence between the longitudinal quality of life outcome and the hazard of death.

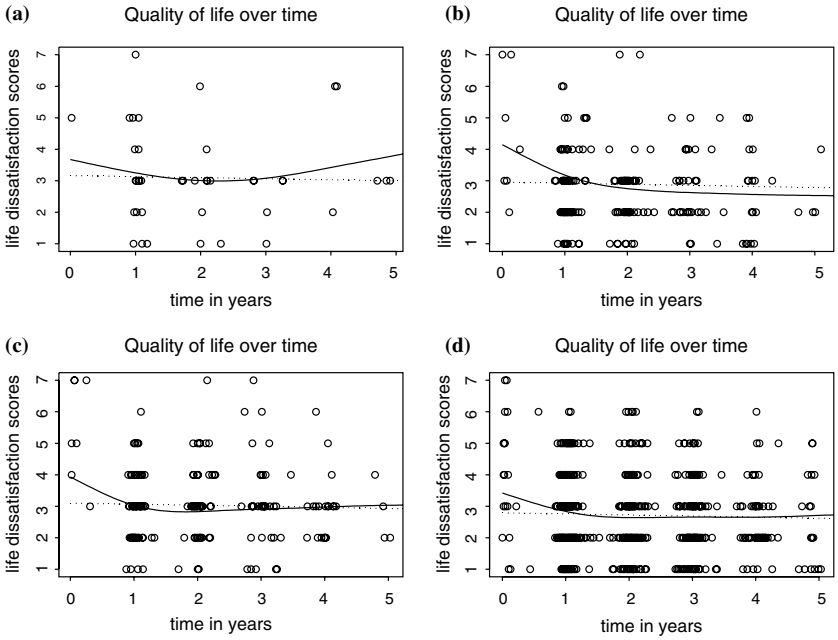

Figure 1. Solid lines are the fitted curves using smoothing splines and the dotted lines are the fitted curves from the simultaneous modelling. (a) is the plot for the Caucasian patients who were single and never had ascites; (b) is the plot for the Caucasian patients who were married and never had ascites; (c) is the plot for the Caucasian patients who were single and ever had ascites; (d) is the plot for the Caucasian patients who were married and ever had ascites.

After fitting the joint model, we find that the variables of gender, BD, CHOL and EDE are not statistically significant in either the model for the quality of life or the model for the survival time. We remove these variables and refit the joint model. The likelihood ratio test indicates that after removing these four variables, the log-likelihood function decreases from -2270.8 to -2274.5, which is not statistically significant when compared with 95-percentile of the χ^2-distribution with 4 degrees of freedom. Table 4 gives the result from our simultaneous modelling. From the table, the patients' dissatisfaction with the current health status significantly varied among three centers. An overall likelihood ratio test for the significance of the centers gives the test statistics value 13 (p-value less than 0.001). The hazard of death in these centers were similar; the patients who were currently married or who ever had ascites before the liver transplantation were more satisfied with their health after transplantation than those who were not; but these factors were not statistically significant in predicting death; the elder and Caucasian patients tended to have a higher risk of death but their quality of life was not significantly different from other groups. We further find that the patients with lower BMI tended to be more satisfied

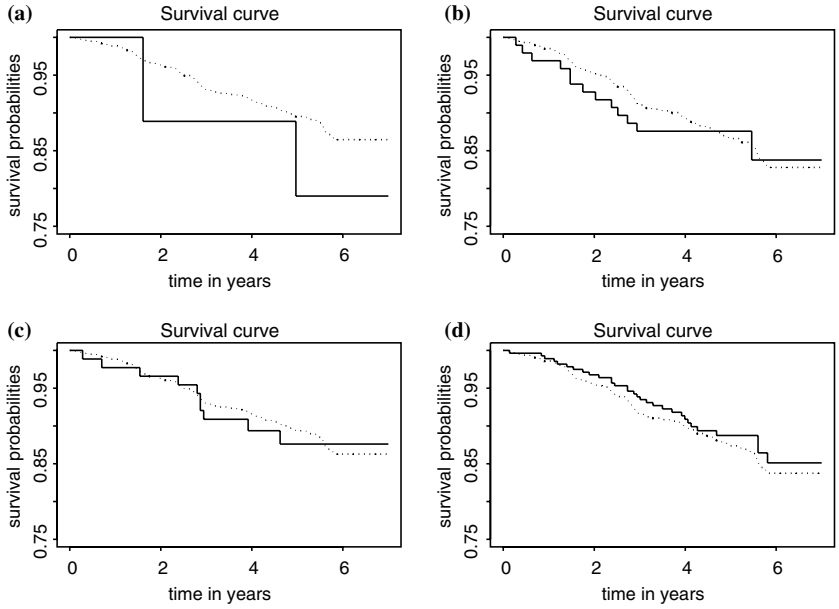

Figure 2. Solid lines are the Kaplan–Meier curves and the dotted lines are the fitted curves from the simultaneous modelling. (a) is the plot for the Caucasian patients who were single and never had ascites; (b) is the plot for the Caucasian patients who were married and never had ascites; (c) is the plot for the Caucasian patients who were single and ever had ascites; (d) is the plot for the Caucasian patients who were married and ever had ascites.

with health but their risk of death was not affected by it. We also note that the patients' dissatisfaction did not change significantly within five years post transplantation. Finally, the parameter ϕ is significantly positive, which implies that some latent factor causing life dissatisfaction also increases the risk of death.

For comparison, we also analyze the data using separate models for fitting the life dis-satisfaction score and the survival time. The results from this separate analysis, which are not reported here, indicate that only marriage status and EDE were marginally significant in the model for the quality of life and the life dissatisfaction score decreased significantly over years; while only age and race significantly affected the survival. Both models did not find that the BMI index was significant.

We also use the results from the simultaneous modelling to predict the mean quality of life scores and the survival function for a given group. The predicted survival function can be obtained using the expression

$$\int_{\mathbf{a}} \exp\left\{-\int_0^t e^{\widetilde{\mathbf{W}}(s)(\boldsymbol{\phi}\circ\mathbf{a})+\mathbf{W}(s)\gamma} d\Lambda(s)\right\} (\sqrt{2\pi|\boldsymbol{\Sigma}_a|})^{-k_0} \exp\left\{-\frac{\mathbf{a}'\boldsymbol{\Sigma}_a^{-1}\mathbf{a}}{2}\right\} d\mathbf{a},$$

Table 4. Results from analyzing liver transplantation study.

Covariate	$\hat{\beta}$	$\hat{se}_1$	$\hat{\beta}/\hat{se}_1$	$\hat{se}_2$	$\hat{\beta}/\hat{se}_2$
Quality of life satisfaction score					
Intercept	2.3745	0.2826	8.40	0.2662	8.92
I(Center = 2)	0.3435	0.1223	2.81	0.1146	3.00
I(Center = 3)	0.3873	0.1186	3.26	0.1065	3.63
Age	0.0032	0.0048	0.65	0.0022	1.42
race	0.1119	0.1282	0.87	0.1226	0.91
marriage	−0.2695	0.1115	−2.42	0.1189	−2.27
ASC	−0.1924	0.0993	−1.94	0.1067	−1.80
BMI	0.0247	0.0091	2.72	0.0038	6.41
Time	−0.0338	0.0255	−1.33	0.0256	−1.32
Survival endpoint					
ϕ	0.5774	0.1371	4.21	0.1419	4.07
I(Center = 2)	0.6042	0.3377	1.79	0.3289	1.84
I(Center = 3)	0.5074	0.3262	1.55	0.3249	1.56
Age	0.0253	0.0123	2.05	0.0115	2.20
race	−0.5582	0.3052	−1.83	0.2929	−1.91
marriage	0.0725	0.3037	0.24	0.2994	0.24
ASC	−0.1463	0.2889	−0.51	0.2813	−0.52
BMI	0.0202	0.0202	1.00	0.0220	0.92

$\hat{se}_1$ and $\hat{se}_2$ are the estimated standard errors based on the methods of se(I) and se(II), respectively. ϕ is the coefficient of the random effect in the multiplicative hazard model. The estimate for σ_y^2 is 0.6387 with $\hat{se}_1(\hat{\sigma}_y^2) = 0.0262$ and $\hat{se}_2(\hat{\sigma}_y^2) = 0.0377$ and the estimate for the covariance of (a_0, a_1) is equal to $\begin{pmatrix} 1.1265 & -0.1276 \\ -0.1276 & 0.0508 \end{pmatrix}$ with $\hat{se}_1(\hat{\Sigma}_a) = \begin{pmatrix} 0.1358 & 0.0417 \\ 0.0417 & 0.0163 \end{pmatrix}$ and $\hat{se}_2(\hat{\Sigma}_a) = \begin{pmatrix} 0.1487 & 0.0501 \\ 0.0501 & 0.0200 \end{pmatrix}$.

where the parameters $(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}_a)$ and $\Lambda(\cdot)$ are substituted with the estimates from the simultaneous modelling. As one example, in Figures 1 and 2, we plot the predicted quality of life and the survival function for the groups with different marriage status and ascites history, respectively, where both predicted curves are obtained as the average over all other covariates. For comparison, we also plot the fitted quality of life scores using smoothing spline and the Kaplan–Meier curves for both figures. Both figures indicate that the model prediction fits the data pretty well. Small difference among the predictive curves of the quality of life is due to the small effect of the marriage status and the ascites history.

Finally, we study the sensitivity of the estimates to the structure of the random components. We use either random effect a_1 or the summation of a_0 and a_1 in the multiplicative hazards model for the survival time. The former gives the log-likelihood value -2276.7 and the latter gives -2274.6. The parameter estimates from these two different models are very similar to the ones given in Table 4. It indicates that our previous model for fitting the transplantation data is not very sensitive to different formulation of random components.

7. Remarks

We have proposed a simultaneous model for modelling both quality of life and survival time. We have also presented the asymptotic properties of the maximum likelihood estimators when both models only share the random effects of time-independent covariates. Moreover, we have proposed an efficient algorithm to evaluate the profile likelihood function in a neighborhood of the maximum likelihood estimates. We conclude that, compared with separate analyses of survival and quality of life, joint modelling can utilize data information to a maximal extent to give more efficient and less biased estimation. Our real data example indicates that joint modelling identifies some risk factors which are not picked up using separate analyses.

Two formulae of estimating the asymptotic variance of the maximum likelihood estimator are given and denoted by $se(I)$ and $se(II)$. The simulation studies show that the algorithm provides a reasonable estimate of the variance and the convergence is rapid. Moreover, we observe from the simulation studies that the formula of $se(I)$ often gives larger values than the ones given by $se(II)$. This might be due to the fact that we need to compute the first-order numerical difference n times (n is the sample size) for the approximations in $se(I)$ so it would accumulate more numerical errors; comparatively, we only need to calculate the second-order numerical difference once for the approximation in $se(II)$. However, the computing time for $se(I)$ is substantially shorter when there are many elements in $\boldsymbol{\theta}$. This is because in obtaining the information matrix using our algorithms, $se(II)$ requires calculating the profile likelihood function $k(k+1)/2$ times, while $se(I)$ only needs such calculations $2k$ times, where k is the dimension of $\boldsymbol{\theta}$.

The quantity $\boldsymbol{\phi}$ in the hazards model can be used to represent the dependence between quality of life and survival time due to the unobserved factors. Base on $\boldsymbol{\phi}$, we can define the correlation between the random term in model (1) and the random term in model (2), which is equal to

$$r(t) = \frac{\tilde{\mathbf{X}}(t)\boldsymbol{\Sigma}_a(\boldsymbol{\phi} \circ \tilde{\mathbf{W}}(t)')}{\sqrt{(\tilde{\mathbf{X}}(t)\boldsymbol{\Sigma}_a\tilde{\mathbf{X}}(t)')((\boldsymbol{\phi} \circ \tilde{\mathbf{W}}(t)')'\boldsymbol{\Sigma}_a(\boldsymbol{\phi} \circ \tilde{\mathbf{W}}(t)'))}}$$

as a quantity characterizing the dependence between these two outcomes due to latent variables at time t. If larger value of $Y(t)$ implies lower quality of life, then positive $r(t)$ implies that high quality of life is associated with decreased risk of death at time t.

In the model for quality of time, we assumed that the measurement error is a white noise process. In fact, one can assume $\epsilon(t)$ to be a gaussian process with serial correlation, for example, an autoregressive process. Such generalization will result in more complicated computation in the E-step. However, our approach and results can be generalized to include this situation. One alternative to our maximum likelihood estimation is Bayesian approach, which was implemented in Xu and Zeger (2001b) for their joint models with common random effect. Comparison

of the two approaches could be of interest. Finally, simultaneous modelling can be further generalized to consider longitudinal models which can incorporate categorical or ordinal data and can be used to model multiple processes or multiple survival events.

Acknowledgement

This work was partially supported by National Institutes of Health Grant R01 HL69720. We also thank the anonymous reviewers for valuable suggestions.

Appendix A

A.1. EM Algorithms

We provide the details of the EM-algorithms in the following: at the kth iteration, *E-step*: we calculate the conditional expectation of $g(\mathbf{a}_i)$ given the observations and the current parameters $(\boldsymbol{\theta}^{(k)}, \Lambda^{(k)})$ for some known function $g(\cdot)$. We denote the result by $E[g(\mathbf{a}_i)|\boldsymbol{\theta}^{(k)}, \Lambda^{(k)}]$. Particularly, given the current parameter $(\boldsymbol{\theta}^{(k)}, \Lambda^{(k)})$,

$$E\left[g(\mathbf{a}_i)|\boldsymbol{\theta}^{(k)}, \Lambda^{(k)}\right] = \frac{\int_{\mathbf{z}} g\left(\sqrt{2}\left(\tilde{\boldsymbol{\Sigma}}_i^{(k)}\right)^{1/2}\mathbf{z}+\tilde{\mu}_i^{(k)}\right)\exp\left\{-\int_0^{Z_i} e^{\tilde{\mathbf{W}}_i(s)\left(\boldsymbol{\phi}^{(k)}\circ\left(\sqrt{2}\left(\tilde{\boldsymbol{\Sigma}}_i^{(k)}\right)^{1/2}\mathbf{z}+\tilde{\mu}_i^{(k)}\right)\right)+\mathbf{W}_i(s)\boldsymbol{\gamma}^{(k)}} d\Lambda^{(k)}(s)\right\}\exp\{-\mathbf{z}'\mathbf{z}\}d\mathbf{z}}{\int_{\mathbf{z}}\exp\left\{-\int_0^{Z_i} e^{\tilde{\mathbf{W}}_i(s)\left(\boldsymbol{\phi}^{(k)}\circ\left(\sqrt{2}\left(\tilde{\boldsymbol{\Sigma}}_i^{(k)}\right)^{1/2}\mathbf{z}+\tilde{\mu}_i^{(k)}\right)\right)+\mathbf{W}_i(s)\boldsymbol{\gamma}^{(k)}} d\Lambda^{(k)}(s)\right\}\exp\{-\mathbf{z}'\mathbf{z}\}d\mathbf{z}}$$

where $\tilde{\boldsymbol{\Sigma}}_i^{(k)} = (\tilde{\mathbf{X}}_i{}'\tilde{\mathbf{X}}_i/(\sigma_y^{(k)})^2 + (\boldsymbol{\Sigma}_a^{(k)})^{-1})^{-1}$ and $\tilde{\mu}_i^{(k)} = \tilde{\boldsymbol{\Sigma}}_i^{(k)}[(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(k)})'\tilde{\mathbf{X}}_i/(\sigma_y^{(k)})^2 + \Delta_i\boldsymbol{\phi}^{(k)}\circ\tilde{\mathbf{W}}_i(Z_i)']$. Both the numerator and the denominator can be calculated using the Gauss–Hermite Quadrature numerical approximation. The latter approximates an integration $\int_{\mathbf{z}}\exp\{-\mathbf{z}'\mathbf{z}\}f(\mathbf{z})d\mathbf{z}$ using a summation like $\sum_{s=1}^m \omega_s f(x_s)$ where ω_s is the weight used in the approximation and x_s is the interpolated points.

M-step: after differentiating the full likelihood function, the updated estimators $(\boldsymbol{\theta}^{(k+1)}, \Lambda^{(k+1)})$ can be obtained as follows.

- $\boldsymbol{\beta}^{(k+1)}$ is the linear regression coefficients of regressing $\{\mathbf{Y}_i - E[\tilde{\mathbf{X}}_i\mathbf{a}_i|\boldsymbol{\theta}^{(k)}, \Lambda^{(k)}], i = 1, \ldots, n\}$ on $\{\mathbf{X}_i, i = 1, \ldots, n\}$.
- $(\sigma_y^2)^{(k+1)}$ is equal to $\{\sum_{i=1}^n N_i\}^{-1}\sum_{i=1}^n[\mathbf{R}_i'\mathbf{R}_i + E[(\tilde{\mathbf{X}}_i\mathbf{a}_i)^2|\boldsymbol{\theta}^{(k)}, \Lambda^{(k)}] - (E[\tilde{\mathbf{X}}_i\mathbf{a}_i|\boldsymbol{\theta}^{(k)}, \Lambda^{(k)}])^2]$, where $\mathbf{R}_i = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(k+1)} - E[\tilde{\mathbf{X}}_i\mathbf{a}_i|\boldsymbol{\theta}^{(k)}, \Lambda^{(k)}]$.
- $\boldsymbol{\Sigma}_a^{(k+1)}$ is equal to $\sum_{i=1}^n E[\mathbf{a}_i\mathbf{a}_i'|\boldsymbol{\theta}^{(k)}, \Lambda^{(k)}]/n$.
- $\begin{pmatrix}\boldsymbol{\phi}^{(k+1)}\\ \boldsymbol{\gamma}^{(k+1)}\end{pmatrix}$ solves the partial likelihood score equation using one-step Newton–Raphson iteration

$$0=\sum_{i=1}^{n}\Delta_i\left\{\begin{pmatrix}E[(\tilde{\mathbf{W}}_i(Z_i)'\circ\mathbf{a}_i)|\boldsymbol{\theta}^{(k)},\Lambda^{(k)}]\\ \mathbf{W}_i\end{pmatrix}\right.$$
$$\left.-\frac{\sum_{Z_j\geq Z_i}\begin{pmatrix}E[(\tilde{\mathbf{W}}_j(Z_i)'\circ\mathbf{a}_j)\exp\{\tilde{\mathbf{W}}_j(Z_i)(\boldsymbol{\phi}\circ\mathbf{a}_j)+\mathbf{W}_j(Z_i)\boldsymbol{\gamma}\}|\boldsymbol{\theta}^{(k)},\Lambda^{(k)}]\\ E[\mathbf{W}_j(Z_i)\exp\{\tilde{\mathbf{W}}_j(Z_i)(\boldsymbol{\phi}\circ\mathbf{a}_j)+\mathbf{W}_j(Z_i)\boldsymbol{\gamma}\}|\boldsymbol{\theta}^{(k)},\Lambda^{(k)}]\end{pmatrix}}{\sum_{Z_j\geq Z_i}E[\exp\{\tilde{\mathbf{W}}_j(Z_i)(\boldsymbol{\phi}\circ\mathbf{a}_j)+\mathbf{W}_j(Z_i)\boldsymbol{\gamma}\}|\boldsymbol{\theta}^{(k)},\Lambda^{(k)}]}\right\}.$$

- $\Lambda^{(k+1)}$ is obtained as an empirical function which has jumps only at the observed failure time:

$$\Lambda^{(k+1)}(z)=\sum_{Z_i\leq z}\frac{\Delta_i}{\sum_{Z_j\geq Z_i}E[\exp\{\tilde{\mathbf{W}}_j(Z_i)(\boldsymbol{\phi}^{(k+1)}\circ\mathbf{a}_j)+\mathbf{W}_j(Z_i)\boldsymbol{\gamma}^{(k+1)}\}|\boldsymbol{\theta}^{(k)},\Lambda^{(k)}]}.$$

A.2. Sketched Proof of Consistency

In this section and the following Section A.3, we give the outline of the proofs for both consistency and asymptotic normality. The technical proof is available from the authors. For simplicity, we assume that a as well as both $\tilde{\mathbf{W}}_i(t)$ and $\tilde{\mathbf{X}}_i(t)$ are one-dimensional. The consistency can be completed from the following three steps.

(i) The estimator of $(\hat{\boldsymbol{\theta}},\hat{\Lambda})$ exists.
(ii) We will show for almost everywhere of Ω (the probability space), $\hat{\Lambda}(\tau)$ is bounded for any n. Therefore, for any subsequence, there exists a sub-subsequence such that $(\hat{\boldsymbol{\theta}},\hat{\Lambda})$ converges almost surely to some random variable $(\boldsymbol{\theta}^*,\Lambda^*)$.
(iii) Finally, we will show the limit $(\boldsymbol{\theta}^*,\Lambda^*)$ is the same as the true parameter $(\boldsymbol{\theta}_0,\Lambda_0)$. Since Λ_0 is continuous, such convergence for $\hat{\Lambda}$ can be strengthened to uniform convergence.

Notations. We define the following notations for $i=1,\ldots,n$. The true values for both Λ and $\boldsymbol{\theta}$ are respectively, Λ_0 and $\boldsymbol{\theta}_0=(\sigma_{0y},\sigma_{0a},\boldsymbol{\beta}_0,\phi_0,\gamma_0)$. Define $S_i=\sum_{j=1}^{N_i}\tilde{\mathbf{X}}_i^j(\tilde{\mathbf{X}}_i^j)'/N_i$, $RSS_i=\sum_{j=1}^{N_i}(Y_i^j-\mathbf{X}_i^j\hat{\boldsymbol{\beta}})(\tilde{\mathbf{X}}_i^j)'$ and let

$$p_i=\exp\left\{\frac{(\hat{\phi}\tilde{\mathbf{W}}_i')^2}{S_i/\hat{\sigma}_y^2+1/\hat{\sigma}_a^2}\right\},\qquad \alpha_i=\left\{S_i/\hat{\sigma}_y^2+1/\hat{\sigma}_a^2\right\}^{-1}\left\{\hat{\phi}\tilde{\mathbf{W}}_i'\Delta_i+RSS_i/\hat{\sigma}_y^2\right\},$$
$$\xi_i=\left\{\sqrt{S_i/\hat{\sigma}_y^2+1/\hat{\sigma}_a^2}\right\}^{-1/2},\qquad \psi_i=\exp\{\mathbf{W}_i(Z_i)\hat{\boldsymbol{\gamma}}+\tilde{\mathbf{W}}_i(\hat{\phi}\alpha_i)+(\tilde{\mathbf{W}}_i\phi)^2\xi_i^2/2\}.$$

Corresponding to the true values of the parameters in the above notations, we also define p_i^0,α_i^0,ξ_i^0, and ψ_i^0 where the parameters in the right-hand side of the above notations are the true parameters. Additionally, we denote $G_i(a,\boldsymbol{\theta},\Lambda)$ as

$$\exp\Big\{-\int_0^{Z_i}\exp\Big\{\frac{\phi\tilde{\mathbf{W}}_i a}{\sqrt{S_i/\sigma_y^2+1/\sigma_a^2}}+\phi\tilde{\mathbf{W}}_i\frac{\phi\tilde{\mathbf{W}}_i\Delta_i+\sum_{j=1}^{N_i}(Y_i^j-\mathbf{X}_i^j\boldsymbol{\beta})(\tilde{\mathbf{X}}_i^j)'/\sigma_y^2}{S_i/\sigma_y^2+1/\sigma_a^2}$$
$$+\mathbf{W}_i(t)\boldsymbol{\gamma}\}d\Lambda(t)\Big\}.$$

The proof of (i) can be done by directly checking that if the jump size of $\hat{\Lambda}$ goes to infinity, then $l_n(\hat{\boldsymbol{\theta}},\hat{\Lambda})\to-\infty$.

Proof of Step (ii). For any event of Ω, we assume $\hat{\boldsymbol{\theta}}\to\boldsymbol{\theta}^*=(\sigma_y^*,\sigma_a^*,\boldsymbol{\beta}^*,\phi^*,\boldsymbol{\gamma}^*)$ for some random variable $\boldsymbol{\theta}^*$. We will prove (ii) by arguments of contradiction. Suppose $\hat{\Lambda}(\tau)\to\infty$. The proof can be divided into the following two cases.
(Case 1). $\phi^*=0$. After differentiating $l_n(\boldsymbol{\theta},\Lambda)$ with respect to $\Lambda(Z_i)-\Lambda(Z_i-)$ and by simplification, we obtain

$$\hat{\Lambda}(Z_k)-\hat{\Lambda}(Z_k-)=\Delta_k\div\Bigg\{\sum_{Z_i\geq Z_k}\Bigg[\exp\{\mathbf{W}_i(Z_i)\hat{\boldsymbol{\gamma}}+\hat{\phi}\tilde{\mathbf{W}}_i\alpha_i+(\hat{\phi}\tilde{\mathbf{W}}_i\xi_i)^2/2\}$$
$$\times\frac{\int_a\exp\{-\frac{a^2}{2}-p_i\int_0^{Z_i}e^{\xi_i\hat{\phi}\tilde{\mathbf{W}}_i a+\hat{\phi}\tilde{\mathbf{W}}_i\alpha_i+\mathbf{W}_i(t)\hat{\gamma}}d\hat{\Lambda}(t)\}da}{\int_a\exp\{-\frac{a^2}{2}-\int_0^{Z_i}e^{\xi_i\hat{\phi}\tilde{\mathbf{W}}_i a+(\hat{\phi}\tilde{\mathbf{W}}_i)\alpha_i+\mathbf{W}_i(t)\hat{\gamma}}d\hat{\Lambda}(t)\}da}\Bigg]\Bigg\}.$$

We also construct another empirical function $\bar{\Lambda}$ such that the jump size of $\bar{\Lambda}$ at any Z_k is the same as the previous expression except that $\hat{\boldsymbol{\theta}}$ and $\hat{\Lambda}$ are replaced by $\boldsymbol{\theta}_0$ and Λ_0. It is direct to check that $\bar{\Lambda}$ converges to Λ_0 uniformly. After substituting the expressions for the jump sizes of $\hat{\Lambda}$ and $\bar{\Lambda}$ in $l_n(\hat{\boldsymbol{\theta}},\Lambda)$, which denotes n^{-1} times the observed log-likelihood function, we obtain that

$$0\leq l_n(\hat{\boldsymbol{\theta}},\hat{\Lambda})-l_n(\boldsymbol{\theta}_0,\bar{\Lambda})\leq O_p(1)+\frac{1}{n}\sum_{i=1}^n\log\frac{1}{\sqrt{2\pi}}\int_a e^{-a^2/2}G_i(a,\hat{\boldsymbol{\theta}},\hat{\Lambda})da$$
$$-\frac{1}{n}\sum_{i=1}^n\Delta_i\log\Bigg[\sum_{Z_k\geq Z_i}\frac{\psi_k}{\sum_{Z_k\geq Z_i}\psi_k}\frac{\int_a e^{-a^2/2}G_k(a,\hat{\boldsymbol{\theta}},\hat{\Lambda})^{p_k}da}{\int_a e^{-a^2/2}G_k(a,\hat{\boldsymbol{\theta}},\hat{\Lambda})da}\Bigg].$$

By the convexity of $-\log x$, we have

$$0\leq O_p(1)+\frac{1}{n}\sum_{i=1}^n\Bigg(1-(p_i-1)\psi_i\sum_{Z_k\leq Z_i}\frac{\Delta_k}{\sum_{Z_l\geq Z_k}\psi_l}\Bigg)\log\Bigg[\frac{1}{\sqrt{2\pi}}\int_a e^{-a^2/2}G_i(a,\hat{\boldsymbol{\theta}},\hat{\Lambda})da\Bigg].$$

Since $\sup_{1\leq i\leq n}(p_i-1)$ goes to zero as $\hat{\phi}$ goes to zero and $\int_a e^{-a^2/2}G_i(a,\hat{\boldsymbol{\theta}},\hat{\Lambda})da$ goes to zero when $\hat{\Lambda}(\tau)$ diverges, the right-hand side becomes negative eventually. We thus obtain the contradiction.

(Case 2). $\phi^* \neq 0$. Define $\hat{\xi} = \log \hat{\Lambda}(\tau)$ and re-scale $\hat{\Lambda}$ by the factor $e^{\hat{\xi}}$. We still denote the re-scaled function as $\hat{\Lambda}$ thus $\hat{\Lambda}(\tau) = 1$. By choosing a subsequence (still use subscript n), we suppose $\hat{\Lambda}$ weakly converges to Λ^*. Clearly, $\hat{\xi}$ maximizes the log-likelihood function $l_n(\hat{\boldsymbol{\theta}}, \hat{\Lambda} e^{\xi})$. Hence, $\hat{\xi}$ satisfies the equation

$$\frac{1}{n}\sum_{i=1}^{n} \Delta_i = \frac{1}{n}\sum_{i=1}^{n} e^{\hat{\xi}} \mathscr{W}(Z_i) \frac{\int_a \exp\{-\frac{a^2}{2} - e^{\hat{\xi}+H(Z_i)+\hat{\mu}^2} e^{\hat{\mu} a}\} da}{\int_a \exp\{-\frac{a^2}{2} - e^{\hat{\xi}+H(Z_i)} e^{\hat{\mu} a}\} da}.$$

Here, $\mathscr{W}(Z_i)$, $H(Z_i)$ are bounded functions independent of $\hat{\xi}$ and $\hat{\mu}_i = \hat{\phi}\tilde{\mathbf{W}}_i / \sqrt{\sum_{j=1}^{N_i} (\tilde{\mathbf{X}}_i^j)' \tilde{\mathbf{X}}_i^j / N_i \hat{\sigma}^2 + 1/\hat{\sigma}_a^2}$. As a result,

$$\frac{1}{n}\sum_{i=1}^{n} \Delta_i \geq \frac{1}{n}\sum_{i=1}^{n} I_{Z_i=\tau} e^{\hat{\xi}} \mathscr{W}(\tau) \frac{\int_a \exp\{-\frac{a^2}{2} - e^{\hat{\xi}+H(\tau)+\hat{\mu}_i^2} e^{\hat{\mu}_i a}\} da}{\int_a \exp\{-\frac{a^2}{2} - e^{\hat{\xi}+H(\tau)} e^{\hat{\mu}_i a}\} da}. \tag{A.1}$$

Base on the Laplace approximation (Evans and Swartz, 2000), for large ω, we can approximate $\int_a \exp\{-\frac{a^2}{2} - \omega e^{\hat{\mu} a}\} da$ by

$$\sqrt{\frac{2\pi}{\log \omega}(1 + o(1))} \times \exp\left\{ -\left[\log \omega \left(1 - \frac{\log\log \omega - \log \hat{\mu}^2}{\log \omega} + o\left(\frac{1}{\log \omega}\right)\right)\right]^2 / (2\hat{\mu}^2) \right.$$
$$\left. -\log \omega \left(1 - \frac{\log\log \omega - \log \hat{\mu}^2}{\log \omega} + o\left(\frac{1}{\log \omega}\right)\right) / \hat{\mu}^2 \right\}.$$

Then (A.1) becomes

$$\frac{1}{n}\sum_{i=1}^{n} \Delta_i \geq O(1) e^{[\log(\hat{\xi}+H(\tau)+\hat{\mu}^2)+\log(\hat{\xi}+H(\tau))]/2} \frac{1}{n}\sum_{i=1}^{n} I_{Z_i=\tau} \to \infty$$

as $\hat{\xi}$ tends to infinity. We thus obtain the contradiction.

Proof of Step (iii). From Step (ii), by choosing a subsequence, we can assume $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}^*$ and $\hat{\Lambda}$ converges to Λ^* pointwise. In this step, we will show that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $\Lambda^* = \Lambda_0$. Denote $\mathbf{O}$ as the observed statistics $(\mathbf{Y}, \mathbf{X}, \mathbf{W}, \tilde{\mathbf{X}}, \tilde{\mathbf{W}}, N, Z)$ and denote $\tilde{G}(a, \mathbf{O}; \boldsymbol{\theta}, \Lambda)$ as

$$\left(\frac{1}{\sqrt{2\pi\sigma_y^2}}\right)^N \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left\{ -\frac{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - a)'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - a)}{2\sigma_y^2} - \frac{a^2}{2\sigma_a^2} + \Delta(\phi\tilde{\mathbf{W}}a + \mathbf{W}(Z)\boldsymbol{\gamma}) \right.$$
$$\left. - \int_0^Z e^{\phi\tilde{\mathbf{W}}a + \mathbf{W}(t)\gamma} d\Lambda(t) \right\}.$$

Then $l_n(\hat{\boldsymbol{\theta}}, \hat{\Lambda}) - l_n(\boldsymbol{\theta}_0, \bar{\Lambda}) \geq 0$ deduces that

$$\mathbf{P}_n\left[\Delta \log \frac{d\hat{\Lambda}(Z)}{d\bar{\Lambda}(Z)}\right] + \mathbf{P}_n\left[\log \frac{\int_a \tilde{G}(a, \mathbf{O}; \hat{\boldsymbol{\theta}}, \hat{\Lambda}) da}{\int_a \tilde{G}(a, \mathbf{O}; \boldsymbol{\theta}_0, \bar{\Lambda}) da}\right] \geq 0, \qquad (A.2)$$

where $\mathbf{P}_n$ is the empirical measure and $\mathbf{P}$ is the expectation. On the other hand, from the construction of $\bar{\Lambda}$, $\hat{\Lambda}$ is absolutely continuous with respect to $\bar{\Lambda}$; moreover, $d\hat{\Lambda}/d\bar{\Lambda}$ uniformly converges to $d\Lambda^*/d\Lambda_0$. After taking limits in (A.2), we obtain

$$\mathbf{P}\left[\log\left\{\frac{(d\Lambda^*(Z))^{\Delta} \int_a \tilde{G}(a, \mathbf{O}; \boldsymbol{\theta}^*, \Lambda^*) da}{(d\Lambda_0(Z))^{\Delta} \int_a \tilde{G}(a, \mathbf{O}; \boldsymbol{\theta}_0, \Lambda_0) da}\right\}\right] \geq 0,$$

where $d\Lambda^*(z)/dz = \lambda^*(z)$, $d\Lambda_0(z)/dz = \lambda_0(z)$ are the derivatives with respect to z of $\Lambda^*(z)$ and $\Lambda_0(z)$, respectively. By the positiveness of the Kullback–Leibler information, with probability 1, it holds that

$$\left(\frac{d\Lambda^*(t)}{dt}\right)^{\Delta} \int_a \tilde{G}(a, \mathbf{O}; \boldsymbol{\theta}^*, \Lambda^*) da = \left(\frac{d\Lambda_0(t)}{dt}\right)^{\Delta} \int_a \tilde{G}(a, \mathbf{O}; \boldsymbol{\theta}_0, \Lambda_0) da. \qquad (A.3)$$

Let $\Delta = 0$ and $Z = 0$. After integrating over a and comparing the coefficients of $\mathbf{YY}'$, $\mathbf{YX}'$ and $\mathbf{XX}'$ in the exponential parts (noting that $P(N > k_0) > 0$), we can obtain $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0, \sigma_y^* = \sigma_{0y}, \sigma_a^* = \sigma_{0a}$. Next, we let $\Delta = 0$ and notice that (A.3) becomes

$$E\left[\exp\left\{-\int_0^Z e^{\phi^* \tilde{\mathbf{W}} a + \mathbf{W}(t)\gamma^*} d\Lambda^*(t)\right\}\right] = E\left[\exp\left\{-\int_0^Z e^{\phi_0 \tilde{\mathbf{W}} a + \mathbf{W}(t)\gamma_0} d\Lambda_0(t)\right\}\right],$$

where a has a normal distribution in a normal family where a is a complete statistics. Thus, for any a,

$$\exp\left\{-\int_0^Z e^{\phi^* \tilde{\mathbf{W}} a + \mathbf{W}(t)\gamma^*} d\Lambda^*(t)\right\} = \exp\left\{-\int_0^Z e^{\phi_0 \tilde{\mathbf{W}} a + \mathbf{W}(t)\gamma_0} d\Lambda_0(t)\right\}.$$

We obtain that $\phi^* = \phi_0, \boldsymbol{\gamma}^* = \boldsymbol{\gamma}_0$ and $\Lambda^* = \Lambda_0$.

The proof for the consistency of $(\hat{\boldsymbol{\theta}}, \hat{\Lambda})$ is completed from the steps (i) to (iii).

A.3. Sketched Proof of Asymptotic Normality

The asymptotic properties for the estimators $(\hat{\boldsymbol{\theta}}, \hat{\Lambda})$ follow if we can verify the conditions in van der Vaart and Wellner (1996), Theorem 3.3.1., which is re-stated in Parner (1998), Appendix A. In terms of notations in Parner (1998) Appendix A, we let $\psi = (\boldsymbol{\theta}, \Lambda) \in \psi = \{(\boldsymbol{\theta}, \Lambda) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \|\Lambda - \Lambda_0\|_{L^\infty} \leq \delta\}$ for a fixed constant δ. Define a set $H = \{(\mathbf{h}_1, h_2(t)) : \|\mathbf{h}_1\| \leq 1, \|h_2\|_V \leq 1\}$, where $\|.\|_V$ means the total variation on $[0, \tau]$. We let

$$S_n(\psi)(\mathbf{h}_1, h_2) = \mathbf{P}_n[l_{\boldsymbol{\theta}}\mathbf{h}_1 + l_\Lambda[h_2]], \quad S(\psi)(\mathbf{h}_1, h_2) = \mathbf{P}[l_{\boldsymbol{\theta}}\mathbf{h}_1 + l_\Lambda[h_2]],$$

where $l_{\boldsymbol{\theta}}$ and l_{Λ} are the derivative operators of the log-likelihood function $l(\mathbf{O};\boldsymbol{\theta},\Lambda)$ along the curve $\boldsymbol{\theta}_{\epsilon}=\boldsymbol{\theta}_0+\epsilon\mathbf{h}_1, \Lambda_{\epsilon}(t)=\int_0^t(1+\epsilon h_2(t))d\Lambda_0(t)$. Thus, S_n, S are both maps from $\psi\subset l^{\infty}(H)$ to $l^{\infty}(H)$. By the Donsker property of the class of function

$$\{l_{\boldsymbol{\theta}}\mathbf{h}_1+l_{\Lambda}[h_2]:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|+\|\Lambda-\Lambda_0\|_V<\delta,(\mathbf{h}_1,h_2)\in H\}$$

as well as the smoothness of $l(\mathbf{O};\boldsymbol{\theta},\Lambda)$ in $(\boldsymbol{\theta},\psi)$, the conditions (a), (b), (d), and the first half of (c) in Parner (1998) Appendix A, hold. It remains to prove that ∇S_{ψ_0} is continuously invertible on its range. Using similar arguments to Parner (1998), the invertibility of ∇S_{ψ_0} is equivalent to show that if there exists some $\mathbf{h}_1$ and h_2 such that $l_{\boldsymbol{\theta}}\mathbf{h}_1+l_{\Lambda}[h_2]=0$ at $(\boldsymbol{\theta}_0,\Lambda_0)$, then $\mathbf{h}_1=0$ and $h_2=0$. However, the latter can verified using the same identifiability arguments in the proof of Step (iii).

Finally, from Parner (1998) Appendix A, $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0,\hat{\Lambda}-\Lambda_0)$ converges to a Gaussian element as a random element of $l^{\infty}(H)$. Certainly, $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ has an asymptotic multivariate normal distribution with mean zero.

A.4. Profile Likelihood Based on EM Algorithms

In a general setting, we suppose $\mathbf{Y}^m$ to be missing part and $\mathbf{Y}^o$ to be observable part of complete data $\mathbf{Y}^c=(\mathbf{Y}^m,\mathbf{Y}^o)$. $\mathbf{Y}^c$ has a distribution depending on the parameters of interest, $\boldsymbol{\theta}$, and the nuisance parameter, η. The profile likelihood for any $\boldsymbol{\theta}$ in a neighborhood of $\hat{\boldsymbol{\theta}}$ is defined as

$$pl_n(\boldsymbol{\theta})=arg\,max_{\eta\in\mathscr{S}_n}\frac{1}{n}\sum_{i=1}^{n}\log\int_{\mathbf{Y}_i^m}f_{\mathbf{Y}^c}(\mathbf{Y}_i^m,\mathbf{Y}_i^o;\boldsymbol{\theta},\eta)d\mathbf{Y}_i^m.$$

The PEME algorithms are used to approximate the profile likelihood function and they include three steps.

(E-step). Calculate the conditional expectation of $\sum_{i=1}^{n}l(\mathbf{Y}_i^m,\mathbf{Y}_i^o;\boldsymbol{\theta},\eta)$ given $\{\mathbf{Y}_i^o,i=1,\ldots,n\}$ for the fixed $\boldsymbol{\theta}$ as well as the current value of $\eta=\eta^{(k)}$, where $l(\mathbf{Y}_i^m,\mathbf{Y}_i^o;\boldsymbol{\theta},\eta)$ is the log-likelihood function of $(\mathbf{Y}_i^m,\mathbf{Y}_i^o)$, i.e., $\log f_{\mathbf{Y}^c}(\mathbf{Y}_i^m,\mathbf{Y}_i^o;\boldsymbol{\theta},\eta)$.

(M-step). The updated value of $\eta=\eta^{(k+1)}$ maximizes the above conditional expectation for the fixed $\boldsymbol{\theta}$.

(Evaluation). Repeat the above E-step and M-step till $\eta^{(k)}$ converges then $pl_n(\boldsymbol{\theta})$ can be obtained by computing

$$pl_n(\boldsymbol{\theta})=\frac{1}{n}\sum_{i=1}^{n}\log\int_{\mathbf{Y}_i^m}f_{\mathbf{Y}^c}(\mathbf{Y}_i^m,\mathbf{Y}_i^o;\boldsymbol{\theta},\eta^{(k)})d\mathbf{Y}_i^m$$

either numerically or through the Monte–Carlo approach.

In the following, we verify that each iteration of the PEME algorithms increases the log-likelihood function $\sum_{i=1}^{n}\log f_{\mathbf{Y}^o}(\mathbf{Y}_i^o;\boldsymbol{\theta},\eta)$. At kth iteration, by the algorithm, we have

$$\sum_{i=1}^{n} E[\log f_{\mathbf{Y}^c}(\mathbf{Y}_i^m, \mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k)})|\mathbf{Y}_i^o, \boldsymbol{\theta}, \eta^{(k)}] < \sum_{i=1}^{n} E[\log f_{\mathbf{Y}^c}(\mathbf{Y}_i^m, \mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k+1)})|\mathbf{Y}_i^o, \boldsymbol{\theta}, \eta^{(k)}].$$

Then

$$\begin{aligned}
&\sum_{i=1}^{n} E[\log f_{\mathbf{Y}^m|\mathbf{Y}^o}(\mathbf{Y}_i^m|\mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k)})|\mathbf{Y}_i^o, \boldsymbol{\theta}, \eta^{(k)}] + \sum_{i=1}^{n} \log f_{\mathbf{Y}^o}(\mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k)}) \\
&\quad \leq \sum_{i=1}^{n} E[\log f_{\mathbf{Y}^m|\mathbf{Y}^o}(\mathbf{Y}_i^m|\mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k+1)})|\mathbf{Y}_i^o, \boldsymbol{\theta}, \eta^{(k)}] + \sum_{i=1}^{n} \log f_{\mathbf{Y}^o}(\mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k+1)}) \\
&\quad < \sum_{i=1}^{n} E[\log f_{\mathbf{Y}^m|\mathbf{Y}^o}(\mathbf{Y}_i^m|\mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k)})|\mathbf{Y}_i^o, \boldsymbol{\theta}, \eta^{(k)}] + \sum_{i=1}^{n} \log f_{\mathbf{Y}^o}(\mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k+1)}).
\end{aligned}$$

Hence, $\sum_{i=1}^{n} \log f_{\mathbf{Y}^o}(\mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k)}) < \sum_{i=1}^{n} \log f_{\mathbf{Y}^o}(\mathbf{Y}_i^o; \boldsymbol{\theta}, \eta^{(k+1)})$.

References

P. Diggle, K. Liang, and S. Zeger, *Analysis of Longitudinal Data*, Oxford: London, 1994.

M. Evans and T. Swartz, *Approximating Integrals via Monte Carlo and Deterministic Methods*, Oxford: London, 2000.

J. Hogan and N. Laird, "Mixture models for the joint distribution of repeated measures and event times," *Statistics in Medicine* vol. 16 pp. 239–257, 1997.

P. Hu, A. Tsiatis, and M. Davidian, "Estimating the parameters in the Cox model when covariates are measured with error," *Biometrics* vol. 54 pp. 1407–1419, 1998.

W. Huang, S. Zeger, J. Anthony, and E. Garrett, "Latent variable model for joint analysis of multiple repeated measures and bivariate event times," *Journal of American Statistical Association* vol. 96 pp. 906–914, 2001.

S. Murphy and A. van der Vaart, "On profile likelihood," *Journal of American Statistical Association* vol. 95 pp. 450–485, 2000.

E. Parner, "Asymptotic theory for the correlated gamma-frailty model," *Annals of Statistics* vol. 26 pp. 183–214, 1998.

J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Estimation of regression coefficients when some of the regressors are not always observed," *Journal of the American Statistical Association* vol. 89 pp. 846–866, 1994.

A. Rotnitzky and J.M. Robins, "Semiparametric regression estimation in the presence of dependent censoring," *Biometrika* vol. 82 pp. 805–820, 1995.

A. Rotnitzky, J. M. Robins, and D. Scharfstein, "Semiparametric regression for repeated outcomes with nonignorable nonresponse," *Journal of the American Statistical Association* vol. 93 pp. 1321–1339, 1998.

A. Tsiatis and M. Davidian, "A semiparametric estimator for the proportional hazards models with longitudinal covariates measured with error," *Biometrika* vol. 88 pp. 447–458, 2001.

A. Tsiatis, V. Degruttola, and M. Wulfsohn, "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* vol. 90 pp. 27–37, 1995.

A. van der Vaart and J. Wellner, *Weak Convergence and Empirical Processes*, Springer-Verlag: Heidelberg, 1996.

M. Wu and K. Bailey, ''Estimation and comparison of changes in the presence of informative right censoring: conditional linear model,'' *Biometrics* vol. 45 pp. 939–955, 1989.
M. Wu and R. Carroll, ''Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process,'' *Biometrics* vol. 44 pp. 175–188, 1988.
M. Wulfsohn and A. Tsiatis, ''A joint model for survival and longitudinal data measured with error,'' *Biometrics* vol. 53 pp. 330–339, 1997.
J. Xu and S. Zeger, ''The evaluation of multiple surrogate endpoints,'' *Biometrics* vol. 57 pp. 81–87, 2001a.
J. Xu and S. Zeger, ''Joint analysis of longitudinal data comprising repeated measures and times to events,'' *Applied Statistics* vol. 50 pp. 375–387, 2001b.
H. Zhao and A. Tsiatis, ''Efficient estimation of the distribution of quality adjusted survival time,'' *Biometrics* vol. 55 pp. 1101–1107, 1999.