Published in final edited form as:

Stat Biosci. 2015 May ; 7(1): 19–47. doi:10.1007/s12561-013-9091-z.

Joint Analysis of Survival Time and Longitudinal Categorical Outcomes

Jaeun Choi,

Department of Health Care Policy, Harvard Medical School 180 Longwood Avenue, Boston, MA 02115, USA Tel.: +617-432-0183, Fax: +617-432-0173 choi@hcp.med.harvard.edu

Jianwen Cai,

Department of Biostatistics, University of North Carolina at Chapel Hill McGavran-Greenberg HI, 135 Dauer Drive, CB 7420, Chapel Hill, NC 27599, USA cai@bios.unc.edu

Donglin Zeng, and

Department of Biostatistics, University of North Carolina at Chapel Hill McGavran-Greenberg HI, 135 Dauer Drive, CB 7420, Chapel Hill, NC 27599, USA dzeng@bios.unc.edu

Andrew F. Olshan

Department of Epidemiology, University of North Carolina at Chapel Hill McGavran-Greenberg HI, 135 Dauer Drive, CB 7435, Chapel Hill, NC 27599, USA andyolshan@unc.edu

Abstract

In biomedical or public health research, it is common for both survival time and longitudinal categorical outcomes to be collected for a subject, along with the subject's characteristics or risk factors. Investigators are often interested in finding important variables for predicting both survival time and longitudinal outcomes which could be correlated within the same subject. Existing approaches for such joint analyses deal with continuous longitudinal outcomes. New statistical methods need to be developed for categorical longitudinal outcomes. We propose to simultaneously model the survival time with a stratified Cox proportional hazards model and the longitudinal categorical outcomes with a generalized linear mixed model. Random effects are introduced to account for the dependence between survival time and longitudinal outcomes due to unobserved factors. The Expectation-Maximization (EM) algorithm is used to derive the point estimates for the model parameters, and the observed information matrix is adopted to estimate their asymptotic variances. Asymptotic properties for our proposed maximum likelihood estimators are established using the theory of empirical processes. The method is demonstrated to perform well in finite samples via simulation studies. We illustrate our approach with data from the Carolina Head and Neck Cancer Study (CHANCE) and compare the results based on our simultaneous analysis and the separately conducted analyses using the generalized linear mixed model and the Cox proportional hazards model. Our proposed method identifies more predictors than by separate analyses.

Keywords

EM algorithm; Generalized linear mixed model; Maximum likelihood estimator; Random effect; Simultaneous modeling; Stratified Cox proportional hazards model

1 Introduction

In biomedical or public health research, it is common that both longitudinal outcomes over time and survival endpoint are collected for a subject, along with the subject's characteristics or risk factors. Investigators are often interested in finding important variables which predict both longitudinal outcomes and survival time. Since longitudinal outcomes and survival time are dependent, it is natural to analyze these two outcomes jointly.

Among the existing approaches for longitudinal data and survival time, the selection model and the pattern mixture model have been widely used. The selection model estimates the distribution of survival time given longitudinal data. The selection model with continuous longitudinal data was studied by Tsiatis, De Gruttola, and Wulfsohn (1995), Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Henderson, Diggle and Dobson (2000), Tsiatis and Davidian (2001), Xu and Zeger (2001a,b), Song, Davidian and Tsiatis (2002), Tseng, Hsieh and Wang (2005), Song and Wang (2007) and Ye, Lin and Taylor (2008) among others. The selection model with categorical longitudinal data was considered by Faucett, Schenker and Elashoff (1998), Huang *et al.* (2001), Xu and Zeger (2001a,b), Lin, McCulloch, and Mayne (2002), Chen, Ibrahim, and Lipsitz (2002), Larsen (2004), Yao (2008), and Chakraborty and Das (2010) among others. The pattern mixture model focuses on the trend of longitudinal outcomes conditional on survival time. The pattern mixture model with continuous longitudinal outcomes was studied by Wu and Carroll (1988), Wu and Bailey (1989), Schluchter (1992), Hogan and Laird (1997), Ribaud, Thompson and Allen-Mersh (2000) and more recently by Ding and Wang (2008). Pulkstenis, Ten Have and Landis (1998) considered the pattern mixture model of binary longitudinal outcomes with informative dropout. Albert and Follmann (2000) proposed to model repeated count data subject to informative dropout, and Albert, Follmann, Wang and Suh (2002) and Albert and Follmann (2007) studied binary longitudinal data with informative missingness. However, these methods cannot be applied directly to assess covariate effects on both outcomes. Simultaneous modeling of the longitudinal and survival data is needed for such purpose.

Xu and Zeger (2001b) and Zeng and Cai (2005a) proposed simultaneous models of longitudinal outcome and survival time. In their articles, heterogeneity caused by unobserved factors is represented using subject-specific random effects. Xu and Zeger (2001b) considered both continuous and categorical longitudinal outcome and proposed the Bayesian approach using the MCMC for estimation. Zeng and Cai (2005a) considered continuous longitudinal outcome and adopted the EM algorithm for estimation. In their approach, given random effects, survival time and the repeated measurements of longitudinal outcomes are assumed to follow a Cox proportional hazards model and a Gaussian distribution, respectively. Recently, simultaneous models with varied types of survival events and random effect structures have been studied (Elashoff, Li and Ni 2007, 2008; Liu, Ma and O'Quigley 2008; Rizopoulos, Verbeke and Molenberghs 2008; Rizopoulos, Verbeke, Lesaffre and Vanrenterghem 2008). Bayesian methods were also proposed for inference (Wang and Taylor 2001; Brown and Ibrahim 2003; Dunson and Herring 2005; Chen, Ghosh, Raghunathan, and Sargent 2009; Hu, Li and Li 2009; Huang, Li, Elashoff and Pan 2011). However, among all the aforementioned simultaneous models,

most studies except Xu and Zeger (2001b), Dunson and Herring (2005), Rizopoulos, Verbeke, Lesaffre and Vanrenterghem (2008), and Chen, Ghosh, Raghunathan, and Sargent (2009) are restricted to continuous longitudinal outcomes. For non-continuous longitudinal outcomes, Rizopoulos, Verbeke, Lesaffre and Vanrenterghem (2008) studied binary data with excess zeros through extending the previous study by Rizopoulos, Verbeke and Molenberghs (2008) which assumed an accelerated failure time model and used a copula function for random effects in the continuous longitudinal and survival processes, Dunson and Herring (2005) proposed a general underlying Poisson variable framework for discrete survival and longitudinal outcomes accommodating dependency through an additive gamma frailty model for the Poisson means in Bayesian approach, and Chen, Ghosh, Raghunathan, and Sargent (2009) considered a latent variable-based multivariate regression model with structured variance covariance matrix by assuming probit models for two binary outcomes and a log-normal accelerated failure time model for survival outcome and conducted the Bayesian inference through the Markov Chain Monte Carlo (MCMC) method.

Compared to the studies for continuous longitudinal data and survival time, relatively little work has been done in the simultaneous modeling frame work for categorical longitudinal data and survival time. However, the longitudinal outcomes may not be continuous in some biomedical studies, for example, where the outcomes are disease symptom with categories of mild/moderate/severe, quality of life measurements with dissatisfied/satisfied, or dichotomized test results with categories of positive/negative. With these categorical longitudinal outcomes, the existing theory for continuous longitudinal outcomes cannot be applied directly and the numerical algorithm needs to be modified. Therefore, in this paper, we investigate the simultaneous modeling of survival time and longitudinal categorical outcomes. Survival time is modeled in the Cox proportional hazards with the unspecified baseline hazard rate, and furthermore the hazards model is extended to allow multiple strata. Random effects are introduced into the proposed models to account for the dependence between survival time and longitudinal outcomes due to unobserved factors. We also establish the theoretical justification of the asymptotic properties of the maximum likelihood estimates by employing the theory of empirical process. Therefore, the contributions of this paper to the recent developments in the simultaneous modeling of categorical longitudinal data and survival data since Xu and Zeger (2001b) are the following: (1) We propose an efficient estimation based on the Nonparametric Maximum Likelihood Estimation (NPMLE) with no assumption for baseline hazard rates while Dunson and Herring (2005), Rizopoulos, Verbeke, Lesaffre and Vanrenterghem (2008), and Chen, Ghosh, Raghunathan, and Sargent (2009) considered parametric models for survival outcomes. (2) We implemented our proposed model in an Expectation-Maximization (EM) algorithm while Xu and Zeger (2001b), Dunson and Herring (2005), and Chen, Ghosh, Raghunathan, and Sargent (2009) proposed Bayesian estimation methods. (3) We provided an asymptotic theory for our efficient estimators.

The outline of this paper is as follows. In Section 2, we present a simultaneous modeling for longitudinal categorical outcomes and survival time, 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 and generalization. EM-algorithms are provided in Appendix A and the proofs for asymptotic results are given in Appendix B.

2 Model 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 $b_i, i = 1, \dots, n$, as a vector of subject-specific random effects of dimension d_b and b_i 's are mutually independent and identically distributed from a multivariate normal with mean zero and covariance matrix Σ_b .

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,

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

with $\mu_i(t) = E(Y_i(t) | b_i) = B'(\eta_i(t))$ and $\nu_i(t) = Var(Y_i(t) | b_i) = B''(\eta_i(t)) A(D_i(t; \phi))$, satisfying

$$\eta_i(t) = g(\mu_i(t)) = X_i(t) \beta + \tilde{X}_i(t) b_i$$

and $v_i(t) = v(\mu_i(t))A(D_i(t; \phi))$, where $g(\cdot)$ and $v(\cdot)$ are known link and variance functions respectively, $X_i(t)$ and $\tilde{X}_i(t)$ are the row vectors of the observed covariates for subject i , and β is a column vector of coefficients for $X_i(t)$. The random effect b_i is allowed to differ for different individuals. Additionally, $X_i(t)$ and $\tilde{X}_i(t)$ can be completely different or share some components, and may include dummy variables for different strata.

Given the random effects 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,

$$\lambda_s(t) \exp \left\{ \tilde{Z}_i(t) (\psi \circ b_i) + Z_i(t) \gamma \right\}, \quad (2)$$

where $Z_i(t)$ and $\tilde{Z}_i(t)$ are the row vectors of the observed covariates and may share some components, ψ is a vector of parameters of the coefficients for random effects, γ is a column vector of coefficients for $Z_i(t)$, 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 $Z_i(t)$ does not include dummy variables for strata since baseline hazard rate is stratum-specific. We assume common fixed effects and random effects across strata in both

hazard and longitudinal models. However, the model may allow for possibly different covariate effects for different strata, which can be achieved by including interaction terms of the covariates with the indicator variables for the stratification variable. Subjects in different strata are assumed to be independent. Here, for any vectors a_1 and a_2 of the same dimension, $a_1 \circ a_2$ denotes the component-wise product. In addition, $\tilde{X}_i(t)$ and $\tilde{Z}_i(t)$ have the same dimensions as b_i 's.

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

Let n_i be the number of the observed longitudinal measurements for subject i , and assume that the distributions of n_i and the observation times for longitudinal measurements are independent of the parameters of interest conditional on 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 $(n_i, Y_{ij}, X_{ij}, \tilde{X}_{ij}), j=1, \dots, n_i, i=1, \dots, n$, and $(V_i, \Delta_i, S_i, \{(Z_i(t), \tilde{Z}_i(t)) : t \leq V_i\}), i=1, \dots, n$, 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 independent of T_i and $Y_i(t)$ given the covariates and the random effects, $V_i = \min(T_i, C_i)$, S_i denotes the stratum, and $i = I(T_i < C_i)$.

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

2.2 Inference procedure

For all n subjects, we write $Y = (Y_1^T, \dots, Y_n^T)^T$, $Y_i = (Y_{i1}, \dots, Y_{in_i})^T$, $V = (V_1, \dots, V_n)^T$, and $b = (b_1^T, \dots, b_n^T)^T$. We also denote X , $\tilde{X}$, Z , and $\tilde{Z}$ as block diagonal matrices with the i -th diagonal components, $X_i = (X_{i1}^T, \dots, X_{in_i}^T)^T$, $\tilde{X}_i = (\tilde{X}_{i1}^T, \dots, \tilde{X}_{in_i}^T)^T$, $Z_i = (Z_{i1}^T, \dots, Z_{in_i}^T)^T$, and $\tilde{Z}_i = (\tilde{Z}_{i1}^T, \dots, \tilde{Z}_{in_i}^T)^T$, respectively, and $S = (S_1, \dots, S_n)^T$. Then, the likelihood function of the complete data $(Y, V, b; X, \tilde{X}, Z, \tilde{Z}, S)$ has the form,

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

and the full likelihood function of the observed data $(Y, V; X, \tilde{X}, Z, \tilde{Z}, S)$ for the parameter (θ, Λ) is expressed as

$$L_f(\theta, \Lambda; Y, V) = \int_b L_c(\theta, \Lambda; Y, V, b) db. \quad (3)$$

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 $\Lambda(t)$, $s = 1, \dots, S$, 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 is considered as missing data for $i = 1, \dots, n$. 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 k -th iteration,

(1) E-step—Calculate the conditional expectations of some known functions of b_i , needed in the next M-step, for subject i with $S_i = s$ given observations and the current estimate $(\theta^{(k)}, \Lambda_s^{(k)})$. To do this, denote $q(b_i)$ and $E[q(b_i) | \theta^{(k)}, \Lambda_s^{(k)}]$ as a known function and its

conditional expectation, respectively. By some algebra, $E[q(b_i) | \theta^{(k)}, \Lambda_s^{(k)}]$ can be expressed in terms of a vector of new variables z_G following a multivariate Gaussian distribution with mean zero. The conditional expectation is calculated using the Gauss-Hermite Quadrature numerical approximation with 20 quadrature points.

(2) M-step—After differentiating the conditional expectation of complete data log-likelihood function given observations and the current estimate $(\theta^{(k)}, \Lambda^{(k)})$, the updated estimator $(\theta^{(k+1)}, \Lambda^{(k+1)})$ can be obtained as follows: $(\beta^{(k+1)}, \phi^{(k+1)})$ solves the conditional expectation of complete data log-likelihood score equation using one-step Newton-Raphson iteration;

$$\Sigma_b^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \sum_{s=1}^S E \left[b_i b_i^T | \theta^{(k)}, \Lambda_s^{(k)} \right] I(S_i = s);$$

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

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

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

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

The expressions of the conditional expectation and the conditional score equations calculated in the E- and M-steps for binary and Poisson longitudinal outcomes with survival time are given respectively in Appendices A.1 and A.2.

The observed information matrix is adopted to obtain the variance estimate for $(\hat{\theta}, \hat{\Lambda}(t))$. For the numerical calculation of the observed information matrix, we consider $A_s\{V_i\}$, the jump size of $A_s(t)$ at V_i belonging to stratum s for which $i = 1$, instead of $\lambda_s(V_i)$. That is,

$\Lambda\{\cdot\} = (\Lambda_1^T\{\cdot\}, \dots, \Lambda_S^T\{\cdot\})^T$ with $A_s\{\cdot\} = (A\{T_{s1}\}, \dots, A\{T_{sm_s}\})^T$ for m_s failure times among n_s subjects ($0 \leq m_s \leq n_s$) of the s -th stratum, $s = 1, \dots, S$. Then, by the Louis (1982) formula,

$$I(\theta, \Lambda\{\cdot\}; Y, V) = E_{b|Y,V} [B_c(\theta, \Lambda\{\cdot\}; Y, V, b) | Y, V] - E_{b|Y,V} [U_c(\theta, \Lambda\{\cdot\}; Y, V, b) U_c^T(\theta, \Lambda\{\cdot\}; Y, V, b) | Y, V] + E_{b|Y,V} [U_c(\theta, \Lambda\{\cdot\}; Y, V)] E_{b|Y,V} [U_c^T(\theta, \Lambda\{\cdot\}; Y, V)],$$

where $U_c(\theta, \Lambda\{\cdot\}; Y, V, b)$ and $B_c(\theta, \Lambda\{\cdot\}; Y, V, b)$ are respectively the first derivative vector and the negative of the second derivative matrix of the complete data log-likelihood $l_c(\theta, \Lambda\{\cdot\}; Y, V, b)$ with respect to $(\theta, \Lambda\{\cdot\})$. The variance of $\sqrt{n} \hat{\theta}$ is asymptotically equal to the corresponding sub-matrix of the inverse of the calculated observed information matrix.

The variance of $\hat{\Lambda}(t)$ is obtained using the estimated variances and covariances corresponding to $A\{\cdot\}$ from the inverse of the observed information matrix where $T \geq t$ at the observed failures. In the EM-algorithm for variance estimation, we evaluate these conditional expectations only at the last iteration of the EM procedure for point estimation, where the conditional expectation of U_c is zero.

3 Asymptotic Properties

To study the asymptotic properties of the proposed estimator $(\hat{\theta}, \hat{\Lambda}(t))$ with

$\hat{\theta} = (\hat{\beta}^T, \hat{\phi}^T, \text{Vech}(\hat{\Sigma}_b)^T, \hat{\psi}^T, \hat{\gamma}^T)^T$ and $\hat{\Lambda}(t) = (\hat{\Lambda}_1(t), \dots, \hat{\Lambda}_S(t))^T$, we assume the following conditions below.

(A1) The true parameter $\theta_0 = (\beta_0^T, \phi_0^T, \text{Vech}(\Sigma_{b0})^T, \psi_0^T, \gamma_0^T)^T$ belongs to a known compact set Θ which lies in the interior of the domain for θ .

(A2) The true baseline hazard rate function $\lambda_0(t) = (\lambda_{10}(t), \dots, \lambda_{S0}(t))$ is bounded and positive in $[0, \tau]$.

(A3) For the censoring time C , $P(C \geq \tau | Z, \tilde{Z}) = P(C = \tau | Z, \tilde{Z}) > 0$.

(A4) For the number of observed longitudinal measurements per subject n_i ,

$P(n_i > d_b | X, \tilde{X}) > 0$ with probability one, and $P(n_i = n_0) = 1$ for some integer n_0 .

(A5) Both $X^T X$ and $\tilde{X}^T \tilde{X}$ are full rank with positive probability. Moreover, if there exist constant vectors c_1 and c_2 such that, with positive probability, for any t , $Z(t)c_1 = a_0(t)$ and $\tilde{Z}(t) \circ c_2 = 0$ for a deterministic function $a_0(t)$, then $c_1 = 0$, $c_2 = 0$, and $a_0(t) = 0$.

Assumption (A3) 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 (A4) implies that some proportion of the subjects have at least d_b longitudinal observations, and there exists an integer n_0 such that $P(n_i = n_0) = 1$. Consistency and asymptotic distribution of the proposed estimator are summarized in the following two theorems. The proofs for Theorem 1 and Theorem 2 are given in Appendices B.1 and B.2, respectively, and we will present outlines of the proofs here.

Theorem 1

Under the assumptions (A1)~(A5), as $n \rightarrow \infty$, the maximum likelihood estimator $(\hat{\theta}, \hat{\Lambda}(t))$ is consistent under the product norm of the Euclidean distance and the supremum norm on

$[0, \tau]$. That is, $\|\hat{\theta} - \theta_0\| + \sup_{t \in [0, \tau]} \|\hat{\Lambda}(t) - \Lambda_0(t)\| \rightarrow 0$ as $n \rightarrow \infty$, where

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

Consistency in Theorem 1 can be proved by verifying the following three steps: First, we show that the maximum likelihood estimate $(\hat{\theta}, \hat{\Lambda})$ exists. This can be achieved by showing that the jump size $\Delta_s\{V_i\}$, with $i = 1$, is finite. Second, we show that, with probability one, $\hat{\Lambda}_s(\tau)$, $s = 1, \dots, S$, are bounded as $n \rightarrow \infty$. This can be proved by showing $\log \hat{\Lambda}_s(\tau)$ is bounded. Third, given that the second step is true, by Helly's selection theorem (van der Vaart, 1998), we can choose a subsequence of $\hat{\Lambda}_s(t)$ such that $\hat{\Lambda}_s(t)$ weakly converges to

some right-continuous monotone function $\Lambda_s^*(t)$ with probability one. For any sub-sequence, we can find a further sub-sequence, still denoted as $\hat{\theta}$, such that $\hat{\theta} \rightarrow \theta^*$. Using empirical process formulation and relevant Donsker properties with parameter identifiability, we can show that $\theta^* = \theta_0$ and $\Lambda_s^* = \Lambda_{s0}$, $s=1, \dots, S$. Based on these results, we can conclude that, with probability one, $\hat{\theta}$ converges to θ_0 and $\hat{\Lambda}_s(t)$ converges to $\Lambda_{s0}(t)$ in $[0, \tau]$, $s=1, \dots, S$. Moreover, since $\Lambda_{s0}(t)$ is right-continuous in $[0, \tau]$, the latter can be strengthened to uniform convergence; that is, $\sup_{t \in [0, \tau]} \|\hat{\Lambda}(t) - \Lambda_0(t)\| \rightarrow 0$ almost surely.

Theorem 2

Under the assumptions (A1)~(A5), as $n \rightarrow \infty$, $\sqrt{n} \left((\hat{\theta} - \theta_0)^T, (\hat{\Lambda}(t) - \Lambda_0(t))^T \right)^T$ weakly converges to a Gaussian random element in $R^{d_\theta} \times \ell^\infty[0, \tau] \times \dots \times \ell^\infty[0, \tau]$ and the estimator $\hat{\theta}$ is asymptotically efficient, where d_θ is the dimension of θ and $\ell^\infty[0, \tau]$ is the normed space containing all the bounded functions in $[0, \tau]$.

Once consistency is held, the conditions of Theorem 3.3.1 in van der Vaart and Wellner (1996), which implies the asymptotic normality of Theorem 2, can be verified via the tools of empirical processes. These conditions are restated in Theorem 4 of Parner (1998). The smoothness conditions in Theorem 4 of Parner (1998) can be verified using the regularity of the log-likelihood function in terms of model parameters and the Donsker properties of the score operators. By Theorem 3.3.1 of van der Vaart and Wellner (1996),

$\sqrt{n} (\hat{\theta} - \theta_0, \hat{\Lambda}_s - \Lambda_{s0})$ weakly converges to a Gaussian process, and, by Proposition 3.3.1 in Bickel et al. (1993), $\hat{\theta}$ is an efficient estimator for θ_0 .

When the sample size increases, the number of event times also increases. However, our asymptotic theory is not based on classical asymptotic theory. Instead it is based on modern empirical process theory, where the nuisance parameter is allowed to be infinite dimension. Thus, the asymptotic properties of the proposed estimator are not affected by the increment of the number of event times.

4 Simulation Studies

In this section, we present some results from our simulation studies. Two sets of simulations with different generalized linear mixed models for the longitudinal outcomes are performed. Binary and Poisson data are considered for longitudinal process in the first and second sets of simulations, respectively.

4.1 Binary longitudinal outcomes and survival time

In this first set of simulations, we assume Y_{ij} to be the j -th binary outcome of the i -th subject following

$$P(Y_{ij} = y_{ij} | b_i) = \exp\{y_{ij}\eta_{ij} - \log(1 + \exp\{\eta_{ij}\})\}, \quad y_{ij} = 0, 1, \quad (4)$$

with $\eta_{ij} = X_{ij}\beta + b_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3ij} + b_i$ for $j=1, \dots, n_i$, and consider the hazard of the i -th subject at time t to be

$$h(t|b_i) = \lambda(t) \exp\{\psi b_i + Z_i(t) \gamma\} = \lambda(t) \exp\{\psi b_i + \gamma_1 Z_{1i} + \gamma_2 Z_{2i}\}, \quad (5)$$

where $b_i \sim N(0, \sigma_b^2)$, and, for simplicity, one stratum of hazard for survival time is 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 the hazard and longitudinal models with a fixed effect intercept. There is one additional covariate denoted as X_{3ij} , the time at measurement, included in the longitudinal model. We suppose the longitudinal data are observed for every 0.3 unit of time, and thus X_{3ij} has the value of every 0.3 unit ranging over 0 through 2.4. The longitudinal data are generated from the Bernoulli distribution with the success probability $P(Y_{ij} = 1|b_i)$ given in (4) and the average number of longitudinal observations (n_i) per subject is 3 with the range of 1 to 8. 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 $t_i = -\log(u_i) \times \exp\{-\{\psi b_i + \gamma_1 + \gamma_2 Z_{2i}\}/\lambda\}$. Censoring time is generated from the uniform distribution between 0.4 and 2.4 so that the censoring proportion is around 25~35%. The observed survival time is obtained by the minimum of the generated survival and censoring times. For the comparison of the estimated baseline cumulative hazards over simulations, we consider three time points: 0.9, 1.4, and 1.9, which correspond to the quartiles of the true survival distribution. The three time points are not the only distinct survival times but are selected to report the estimated cumulative hazard function at these points.

We consider different ψ values of -1, 0, and 1 for negative, zero, and positive dependency between longitudinal process and survival time model, respectively. The parameters in the two models are chosen as $\beta_0 = -1$, $\beta_1 = 1$, $\beta_2 = -0.5$, $\beta_3 = -0.5$, $\sigma_b^2 = 0.5$, $\psi = -1/0/1$, $\gamma_1 = -1$, $\gamma_2 = 1$, and $\lambda(t)$. Different sample sizes ($n=200, 400$) are simulated with 1000 replications. The results of the maximum likelihood estimates for θ and the baseline cumulative hazards at the three time points with 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 the 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 the 95% confidence intervals based on the estimated standard error “ESE”. Satterthwaite (1946) method is used for the coverage probability of σ_b^2 .

From Table 1, we can see that even for the smaller sample size ($n=200$), 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 those for ψ . However, the coverage rate of the parameter ψ is improved for larger

sample size. Additional simulations we conducted show that, with sample sizes of 800, the coverage rates for $\psi = -1, 0$ and 1 were 95.5%, 95.9% and 95.9%, respectively. In addition, the simulations show that the variances of the estimators decrease as the sample size (n) increases. We also can see that the estimates are fairly robust and close to the true values for all different ψ values.

4.2 Poisson longitudinal outcomes and survival time

In the second set of simulations, we assume Y_{ij} to follow a Poisson distribution,

$$P(Y_{ij}=y_{ij}|b_i) = \exp\{y_{ij}\eta_{ij} - \exp\{\eta_{ij}\} - \log(y_{ij}!)\},$$

with η_{ij} defined as in Section 4.1. We also consider the same hazards model and simulation setting as those used in Section 4.1 except $\sigma_b^2 = 0.2$. The simulated Poisson longitudinal outcomes range over 0 to 7 with the average 0.5.

Table 2 shows that overall the estimates perform well even for the smaller sample size $n = 200$ with small biases of the estimates except ψ . We conducted additional simulations with sample sizes of 800 and 1000, and the bias for ψ decreases as sample size increases. The estimated standard errors using the observed information matrix are close to the sample standard deviations, and the 95% confidence interval coverage rates are close to 0.95 except for σ_b^2 and $\psi = 0$.

From Table 2, σ_b^2 is seemingly underestimated with higher than the nominal coverage rates, but the coverage rate is improved for larger sample size. This implies that variance of σ_b^2 may not be estimated well with small sample size for Poisson longitudinal distribution. In the mean time, the test for σ_b^2 is conservative with small sample size, but the type I error becomes closer to the nominal level as sample size increases. Profile likelihood may be an alternative estimation approach for σ_b^2 . The 95% confidence interval coverage for $\psi = 0$ also appears to be higher than the nominal level, but the additional simulation with sample sizes of 800 shows that the high coverage rate reduces to the 95% nominal level. Table 2 also shows that the variances of the estimators decrease for larger sample size, and the estimates are fairly robust and close to the true values for all three different ψ values.

In all the simulations, the EM algorithm converged within 60 iterations. The CPU time used for 1000 data sets in the simulation studies was about 6 hours for sample size 200 averaging 20 seconds per each data set and 15 hours for sample size 400 averaging 1 minute per each data set on a computer with a 64-bit operating system.

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 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 2009, 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%. The number of observations per patient ranges from 1 to 3 with average of 1.93 which may look sparse. However, even if the number of longitudinal measurements per subject is sparse, since our estimation is based on pooling information from all the subjects, the actual measurements used for estimation are not sparse. 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 the Head and Neck Cancer Specific symptoms (HNCS) among the 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 variables, 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 variables. Additionally, 2 interactions with race, i.e. race $\times$ the total number of medical conditions reported and race $\times$ 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 also included as a covariate for longitudinal outcomes. A random intercept for the dependence between the QoL satisfaction and the risk of death is included in both models, and assumed to follow a normal distribution with mean zero. In addition to the simultaneous analysis, we also conduct separate analyses fitting the generalized linear mixed model and the Cox proportional hazards model to the longitudinal QoL and survival time respectively and compare the results to those from our proposed simultaneous method.

After fitting the simultaneous models with all the covariates, we use backward variable selection based on the Likelihood Ratio Test (LRT) and find that surgery, chemotherapy, tumor site, age at diagnosis, and all 2 interactions are not statistically significant in both models for HNCS QoL satisfaction and survival time at the significance level 0.05. We

remove these variables and refit the simultaneous models. Then, the LRT shows that race, radiation therapy, the number of persons supported by household income, BMI, and the total number of medical conditions reported are not statistically significant for the risk of death. We further reduce the models by removing them from the hazards model and refit the reduced simultaneous models.

Table 3 gives the results from this final models. From the “Simultaneous” columns, we see that the number of 12 oz. beers consumed per week, household income, tumor stage, and the total number of medical conditions reported 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 associated with HNCS QoL satisfaction and lower risk of death, higher household income is in general associated with HNCS QoL satisfaction and lower risk of death, both patients’ HNCS QoL satisfaction and risk of death are significantly different for patients in different tumor stages, and patients with a greater number of medical conditions reported have lower HNCS QoL satisfaction and higher risk of death. Specifically, for instance, with the log-scaled odds and hazard ratios of 1.060 and -1.076 for HNCS QoL satisfaction and death respectively, patients who consumed 5 to 14 of 12 oz. beers per week appear to have 2.886 times odds for HNCS QoL satisfaction and 0.341 times hazards of death compared to those that consumed 30 or more of 12 oz. beers per week after adjusting for the other covariates in the model. Looking at the number of medical conditions reported, for each additional medial condition reported, the odds ratio of HNCS QoL satisfaction is decreased by 16% and the hazard of death is increased by 29%. In the meantime, race (African-American), radiation therapy, the number of persons supported by household income, and BMI are selected only in the HNCS QoL longitudinal model. 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 likely to be satisfied with longitudinal HNCS QoL while the risk of death is not affected by these factors. Furthermore, we also find that time at survey measurement is statistically significant in the HNCS QoL longitudinal model implying that patients are more satisfied over time. The parameter ψ for the dependence between longitudinal HNCS QoL and survival time is negative and has $p\text{-value}=0.131$. This implies that the longitudinal HNCS QoL and survival time are marginally correlated and some latent factors which increase HNCS QoL satisfaction also decrease the risk of death.

For the purpose of comparison, we conducted separate analyses for longitudinal HNCS QoL and survival time whose final results are given in the last three columns of Table 3. The generalized linear mixed model (GLMM) and the Cox proportional hazards model are used for longitudinal outcomes and survival time respectively. The GLMM also considers individual heterogeneity through subject-specific random effects although it does not incorporate the correlation between longitudinal outcomes and survival time. Comparing the results from the simultaneous and separate analyses of Table 3, we can see our simultaneous analysis additionally indicates the number of persons supported by household income, BMI, and the total number of medical conditions reported in the HNCS QoL longitudinal model ($p\text{-values}=0.025, 0.007, \text{ and } 0.030$, respectively) and the number of 12 oz. beers consumed

per week in the hazard model (p-values=0.045 and 0.005 for ‘None’ and ‘5 to 14’) as significant while they are not statistically significant by separate analyses.

Figure 1 shows the estimated baseline cumulative hazard rates over follow-up time with the 95% confidence interval. Since the baseline cumulative hazard rates are bounded by 0, we first log-transformed the estimated baseline cumulative hazard rates and obtained the 95% lower and upper bounds for the log-scaled estimated baseline cumulative hazards. Then, we re-transformed them into their original scale. The estimated baseline cumulative hazard rates look flat at the very early time within a year, but soon appear to be linearly increasing. Figure 2 shows the Kaplan-Meier estimates (solid line) and the predicted survival probabilities based on the simultaneous analysis (dashed line). These two survival curves are very close to each other which implies our proposed method fits the data well.

6 Concluding Remarks

We have proposed a method for the simultaneous modeling of longitudinal outcomes including both categorical and continuous data with a generalized linear mixed model and survival time with a stratified multiplicative proportional hazards model through random effects. We have also developed a maximum likelihood estimation method for the proposed simultaneous model, and presented asymptotic properties of the proposed estimators. The proposed estimation procedure using EM algorithm has been assessed via simulation studies. 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. The proposed method was applied to data from the CHANCE study.

When the dimension for random effects is high, computational intensity is increased due to high dimensional Gauss-Hermit quadrature (GQ) integration and the convergence could be slow. To handle such situation, alternative numerical methods such as adaptive quadrature or MCMC may be useful.

A stratified Cox PH model is considered for survival data and each stratum has its unspecified baseline hazard. Even though more strata introduce more baseline hazard functions, the number of the parameters which are associated with the jumps of each cumulative hazard functions remains the same as the total number of observed failures. Therefore, the additional strata do not increase computational complexity while using strata gives more flexible and robust structure if we do believe that survival experience between certain groups is very different and are not proportional over time.

In our proposed method, all the information on survival, longitudinal outcomes, and covariates are used. As a result of this, the parameter estimates can be more efficient. The proposed model also generalizes previous work to general longitudinal outcomes. Future work can include relaxing normal assumption for the random effects, considering generalization to mixed types of longitudinal outcomes, and improving computational efficiency.

Acknowledgements

The authors thank the editor, the associate editor, and two referees of the Statistics in Bioscience for their valuable suggestions that considerably improved this article.

Appendix A: EM Algorithms

A.1. EM algorithm – Binary longitudinal data and survival time

(1) E-step

For binary longitudinal outcomes and survival time, we calculate the conditional expectation of $q(b_i)$ for subject i with $S_i=s$ given the observations and the current estimate $(\theta^{(k)}, \Lambda_s^{(k)})$ for some known function $q(\cdot)$. The conditional expectation denoted by $E[q(b_i)|\theta^{(k)}, \Lambda_s^{(k)}]$ can be expressed as the following: Given the current estimate $(\theta^{(k)}, \Lambda_s^{(k)})$,

$$E[q(b_i)|\theta^{(k)}, \Lambda_s^{(k)}] = \frac{\int_{z_G} q(R(z_G)) K(z_G) \exp\{-z_G^T z_G\} dz_G}{\int_{z_G} K(z_G) \exp\{-z_G^T z_G\} dz_G}, \quad (6)$$

where

$$\begin{aligned} R(z_G) &= (\Sigma_b^{(k)})^{\frac{1}{2}} \left[\sqrt{2} z_G + (\Sigma_b^{(k)})^{\frac{1}{2}} \left(\sum_{j=1}^{n_i} y_{ij} \tilde{X}_{ij}^T + \Delta_i (\tilde{Z}_i^T (V_i) \circ \psi^{(k)}) \right) \right], \\ K(z_G) &= \exp \left\{ - \sum_{j=1}^{n_i} \log \left(1 + e^{X_{ij} \beta^{(k)} + \tilde{X}_{ij} R(z_G)} \right) - \int_0^{V_i} e^{\tilde{Z}_i(u) (\psi^{(k)} \circ R(z_G)) + Z_i(u) \gamma^{(k)}} d\Lambda_s^{(k)}(u) \right\}, \end{aligned} \quad (7)$$

$(\Sigma_b^{(k)})^{\frac{1}{2}}$ is a unique non-negative square root of $\Sigma_b^{(k)}$ (i.e. $(\Sigma_b^{(k)})^{\frac{1}{2}} \times (\Sigma_b^{(k)})^{\frac{1}{2}} = \Sigma_b^{(k)}$), and z_G follows a multivariate Gaussian distribution with mean zero.

(2) M-step

Since the parameter φ is set to 1 for logistic distribution, we estimate only β in the longitudinal process. $\beta^{(k+1)}$ solves the conditional expectation of complete data log-likelihood score equation, using one-step Newton-Raphson iteration,

$$\sum_{i=1}^n \sum_{j=1}^{n_i} \left(y_{ij} - \sum_{s=1}^S E \left[\frac{\exp \{ X_{ij} \beta^{(k+1)} + \tilde{X}_{ij} b_i \}}{1 + \exp \{ X_{ij} \beta^{(k+1)} + \tilde{X}_{ij} b_i \}} \middle| \theta^{(k)}, \Lambda_s^{(k)} \right] I(S_i=s) \right) X_{ij}^T = 0.$$

$\Sigma_b^{(k+1)}$, $\psi^{(k+1)}$, $\gamma^{(k+1)}$, and $\Lambda_s^{(k+1)}$ have the same expressions as in Section 2.2.

A.2. EM algorithm – Poisson longitudinal data and survival time

(1) E-step

For Poisson longitudinal outcomes and survival time, given the current estimate $(\theta^{(k)}, \Lambda_s^{(k)})$, the conditional expectation denoted by $E[q(b_i) | \theta^{(k)}, \Lambda_s^{(k)}]$ can be expressed as in (6) with $R(z_G)$ defined as in (7),

$$K(z_G) = \exp \left\{ - \sum_{j=1}^{n_i} e^{X_{ij}\beta^{(k)} + \tilde{X}_{ij}R(z_G)} - \int_0^{V_i} e^{\tilde{Z}_i(u)(\psi^{(k)} \circ R(z_G)) + Z_i(u)\gamma^{(k)}} d\Lambda_s^{(k)}(u) \right\},$$

and z_G follows a multivariate Gaussian distribution with mean zero.

(2) M-step

Since the parameter ϕ is set to 1 for Poisson distribution, we estimate only β in the longitudinal process. $\beta^{(k+1)}$ solves the conditional expectation of complete data log-likelihood score equation, using one-step Newton-Raphson iteration,

$$\sum_{i=1}^n \sum_{j=1}^{n_i} \left(y_{ij} - \sum_{s=1}^S E \left[\exp \{ X_{ij}\beta^{(k+1)} + \tilde{X}_{ij}b_i \} | \theta^{(k)}, \Lambda_s^{(k)} \right] I(S_i=s) \right) X_{ij}^T = 0.$$

$\Sigma_b^{(k+1)}$, $\psi^{(k+1)}$, $\gamma^{(k+1)}$, and $\Lambda_s^{(k+1)}$ have the same expressions as in Section 2.2.

Appendix B: Proofs for Theorems

In Appendices B.1 and B.2, we sketch the proofs for Theorem 1 and Theorem 2. All detailed technical proofs are available from the authors. From (3) in Section 2.2, we can have the observed log-likelihood function, $l_f(\theta, \Lambda; Y, V) = \log\{L_f(\theta, \Lambda; Y, V)\}$. We obtain and use the following modified object function, by replacing $\lambda_s(V_i)$ with $\Lambda_s\{V_i\}$ in $l_f(\theta, \Lambda; Y, V)$ where $\Lambda_s\{V_i\}$ is the jump size of $\Lambda_s(t)$ at the observed time V_i with $i = 1, \dots, n$,

$$\begin{aligned} l_n(\theta, \Lambda) &= \sum_{i=1}^n \log \left[\int_{b_i} \exp \left\{ \sum_{j=1}^{n_i} \left[\frac{Y_{ij}(X_{ij}\beta + \tilde{X}_{ij}b_i) - B(\beta; b_i)}{A(D_i(t_j; \phi))} + C(Y_{ij}; D_i(t_j; \phi)) \right] \right\} \right. \\ &\quad \times \left(\prod_{s=1}^S [\Lambda_s\{V_i\}^{\Delta_i} \exp \{ \Delta_i [\tilde{Z}_i(V_i)(\psi \circ b_i) + Z_i(V_i)\gamma] - \int_0^{V_i} \exp \{ \tilde{Z}_i(u)(\psi \circ b_i) + Z_i(u)\gamma \} d\Lambda_s(u) \}]^{I(S_i=s)} \right) \\ &\quad \times (2\pi)^{-d_b/2} |\Sigma_b|^{-1/2} \exp \left\{ -\frac{1}{2} b_i^T \Sigma_b^{-1} b_i \right\} db_i \Big], \end{aligned} \quad (8)$$

and $(\hat{\theta}, \hat{\Lambda})$ maximizes $l_n(\theta, \Lambda)$ over the space $\{(\theta, \Lambda) : \theta \in \Theta, \Lambda \in \mathbb{W} \times \mathbb{W} \cdots \times \mathbb{W}\}$, where $\mathbb{W}$ consists of all the right-continuous step functions only; that is,

$\Lambda = (\Lambda_1, \dots, \Lambda_S)^T$, $s=1, \dots, S$, $\Lambda_s \in \mathbb{W}$. For the proofs of both Theorem 1 and Theorem 2, the modified object function is used in place of the observed log-likelihood function.

B.1. Proof of consistency – Theorem 1

Consistency can be proved by verifying the following three steps: First, we show the maximum likelihood estimate $(\hat{\theta}, \hat{\Lambda})$ exists. Second, we show that, with probability one, $\hat{\Lambda}_s(\tau)$, $s=1, \dots, S$, are bounded as $n \rightarrow \infty$. Third, if the second step is true, by Helly's selection theorem (van der Vaart 1998), we can choose a subsequence of $\hat{\Lambda}_s(t)$ such that $\hat{\Lambda}_s(t)$ weakly converges to some right-continuous monotone function $\Lambda_s^*(t)$ with probability one. For any sub-sequence, we can find a further sub-sequence, still denoted as $\hat{\theta}$, such that $\hat{\theta} \rightarrow \theta^*$. Thus, in the third step, we show $\theta^* = \theta_0$ and $\Lambda_s^* = \Lambda_{s0}$, $s=1, \dots, S$. Once the three steps are completed, we can conclude that, with probability one, $\hat{\theta}$ converges to θ_0 and $\hat{\Lambda}_s(t)$ converges to $\Lambda_{s0}(t)$ in $[0, \tau]$, $s=1, \dots, S$. Moreover, since $\Lambda_{s0}(t)$ is right-continuous in $[0, \tau]$, the latter can be strengthened to uniform convergence; that is, $\sup_{t \in [0, \tau]} \|\hat{\Lambda}(t) - \Lambda_0(t)\| \rightarrow 0$ a.s. Then, the proof of Theorem 1 will be done.

In the first step, since θ belongs to a compact set Θ by Assumption (A1), it is sufficient to show that $\Lambda_s\{V_i\}$ with $i=1$ is finite. For each subject i with $i=1$, after simple algebra, we have that, from (8),

$$l_n(\theta, \Lambda) \leq \sum_{i=1}^n \log \int_{b_i} \left[\exp \left\{ \sum_{j=1}^{n_i} \left[\frac{Y_{ij}(X_{ij}\beta + \tilde{X}_{ij}b_i)}{A(D_i(t_j; \phi))} + C(Y_{ij}; D_i(t_j; \phi)) \right] \right\} \right. \\ \left. \times \left(\prod_{s=1}^S [(\Lambda_s\{V_i\})^{-\Delta_i} \exp \left\{ -\Delta_i [\tilde{Z}_i(V_i)(\psi \circ b_i) + Z_i(V_i)\gamma]] \right\} \right]^{I(S_i=s)} (2\pi)^{-d_b/2} |\Sigma_b|^{-1/2} \exp \left\{ -\frac{1}{2} b_i^T \Sigma_b^{-1} b_i \right\} \right] db_i.$$

If $\Lambda_s\{V_i\} \rightarrow \infty$ for some i with $i=1$, then $l_n(\theta, \Lambda) \rightarrow -\infty$, which is contradictory to that $l_n(\theta, \Lambda)$ is bounded. Therefore, we conclude that $\Lambda_s\{V_i\}$ must be finite. By the conclusion and assumption (A1), the maximum likelihood estimate $(\hat{\theta}, \hat{\Lambda})$ exists.

In the second step, we define $\hat{\zeta}_s = \log \hat{\Lambda}_s(\tau)$ and rescale $\hat{\Lambda}_s$ by the factor $e^{\hat{\zeta}_s}$. Then, we let $\tilde{\Lambda}_s$ denote the rescaled function; that is, $\tilde{\Lambda}_s(t) = \hat{\Lambda}_s(t) / \hat{\Lambda}_s(\tau) = \hat{\Lambda}_s(t) e^{-\hat{\zeta}_s}$, thus, $\tilde{\Lambda}_s(\tau) = 1$. To prove this second step, it is sufficient to show $\hat{\zeta}_s$ is bounded. After some algebra in (8), we obtain that, for any $\Lambda \in \mathbb{W} \times \mathbb{W} \cdots \times \mathbb{W}$,

$$n^{-1} l_n(\hat{\theta}, \Lambda) = \frac{1}{2} \sum_{i=1}^n \left[\sum_{j=1}^{n_i} \left(\frac{Y_{ij} X_{ij} \hat{\beta}}{A(D_i(t_j; \hat{\phi}))} + C(Y_{ij}; D_i(t_j; \hat{\phi})) \right) + \sum_{s=1}^S \Delta_i(Z_i(V_i) \hat{\gamma}) I(S_i=s) \right. \\ \left. - \frac{1}{2} \log \left\{ (2\pi)^{d_b} |\hat{\Sigma}_b| \right\} + \frac{1}{2} M_i^T M_i - \frac{1}{2} \log |\hat{\Sigma}_b| + \sum_{s=1}^S \Delta_i I(S_i=s) \log \Lambda_s\{V_i\} \right. \\ \left. + \log \int_{b_0} \left[\exp \left\{ -\frac{1}{2} b_0^T b_0 - \sum_{j=1}^{n_i} \frac{B(\hat{\beta}; b_0)}{A(D_i(t_j; \hat{\phi}))} - \sum_{s=1}^S I(S_i=s) \int_0^{V_i} e^{Q_{1i}(t, b_0, \hat{\theta})} d\Lambda_s(t) \right\} \right] db_0 \right],$$

where

$$M_i = \hat{\Sigma}_b^{1/2} \left(\sum_{j=1}^{n_i} \frac{Y_{ij} \tilde{X}_{ij}}{A(D_i(t_j; \hat{\theta}))} + \sum_{s=1}^s I(S_i=s) \Delta_i(\tilde{Z}_i(V_i) \circ \hat{\psi}^T) \right)^T, b_0 = \Sigma_b^{-1/2} b - M_i,$$

and $Q_{1i}(t, b_0, \hat{\theta}) = (\tilde{Z}_i(t) \circ \hat{\psi}^T) \hat{\Sigma}_b^{1/2} b_0 + Z_i(t) \hat{\gamma} + (\tilde{Z}_i(t) \circ \hat{\psi}^T) \Sigma_b^{1/2} M_i$. Thus, since

$0 \leq n^{-1} l_n(\hat{\theta}, \hat{\Lambda}) - n^{-1} l_n(\hat{\theta}, \tilde{\Lambda})$ where $\hat{\Lambda} = e^{\hat{\Sigma}} \circ \tilde{\Lambda}$, it follows that

$$\begin{aligned} 0 &\leq \frac{1}{n} \sum_{i=1}^n \sum_{s=1}^s \Delta_i I(S_i=s) \left(\log e^{\hat{\Sigma}_s} \tilde{\Lambda}_s - \log \tilde{\Lambda}_s \right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \log \int_{b_0} \left[\exp \left\{ -\frac{1}{2} b_0^T b_0 - \sum_{j=1}^{n_i} \frac{B(\hat{\beta}; b_0)}{A(D_i(t_j; \hat{\phi}))} - \sum_{s=1}^s I(S_i=s) e^{\hat{\Sigma}_s} \int_0^{V_i} e^{Q_{1i}(t, b_0, \hat{\theta})} d\tilde{\Lambda}_s(t) \right\} \right] db_0 \\ &\quad - \frac{1}{n} \sum_{i=1}^n \log \int_{b_0} \left[\exp \left\{ -\frac{1}{2} b_0^T b_0 - \sum_{j=1}^{n_i} \frac{B(\hat{\beta}; b_0)}{A(D_i(t_j; \hat{\phi}))} - \sum_{s=1}^s I(S_i=s) \int_0^{V_i} e^{Q_{1i}(t, b_0, \hat{\theta})} d\tilde{\Lambda}_s(t) \right\} \right] db_0. \end{aligned} \quad (9)$$

According to the assumption (A2), there exist some positive constants C_1, C_2 and C_3 such

that $|Q_{1i}(t, b_0, \hat{\theta})| \leq C_1 \|b_0\| + C_2 \|Y_i\| + C_3$. By denoting b_0 as a vector of variables following a standard multivariate normal distribution, from concavity of the logarithm function, in the third term of (9),

$$\begin{aligned} H_i &\triangleq \log \int_{b_0} \left[\exp \left\{ -\frac{1}{2} b_0^T b_0 - \sum_{j=1}^{n_i} \frac{B(\hat{\beta}; b_0)}{A(D_i(t_j; \hat{\phi}))} - \sum_{s=1}^s I(S_i=s) \int_0^{V_i} e^{Q_{1i}(t, b_0, \hat{\theta})} d\tilde{\Lambda}_s(t) \right\} \right] db_0 \\ &\geq (2\pi)^{d_b/2} E_{b_0} \left[-\sum_{j=1}^{n_i} \frac{B(\hat{\beta}; b_0)}{A(D_i(t_j; \hat{\phi}))} - e^{C_1 \|b_0\| + C_2 \|Y_i\| + C_3} \right] = -e^{C_2 \|Y_i\| + C_4} - C_5, \end{aligned}$$

where C_4 and C_5 are positive constants. Since it is easily verified that

$$E_{b_0} \left[\sum_{j=1}^{n_i} \frac{B(\hat{\beta}; b_0)}{A(D_i(t_j; \hat{\phi}))} + e^{C_1 \|b_0\| + C_2 \|Y_i\| + C_3} \right] < \infty, \text{ by the strong law of large numbers}$$

and the assumption (A4), the third term of (9) can be bounded by a constant C_6 . i.e.

$$-\frac{1}{n} \sum_{i=1}^n H_i \leq \frac{1}{n} \sum_{i=1}^n \left(e^{C_2 \|Y_i\| + C_4 + C_5} \right) \triangleq C_6. \text{ Then, (9) becomes}$$

$$\begin{aligned} 0 &\leq \frac{1}{n} \sum_{i=1}^n \sum_{s=1}^s \Delta_i I(S_i=s) \hat{\Sigma}_s \\ &\quad + \frac{1}{n} \sum_{i=1}^n \log \int_{b_0} \left[\exp \left\{ -\frac{1}{2} b_0^T b_0 - \sum_{j=1}^{n_i} \frac{B(\hat{\beta}; b_0)}{A(D_i(t_j; \hat{\phi}))} - \sum_{s=1}^s e^{\hat{\Sigma}_s} \int_0^{V_i} e^{Q_{1i}(t, b_0, \hat{\theta})} d\tilde{\Lambda}_s(t) \right\} \right] db_0 + C_6 \\ &\leq \frac{1}{n} \sum_{i=1}^n \sum_{s=1}^s \Delta_i I(S_i=s) \hat{\Sigma}_s \\ &\quad + \frac{1}{n} \sum_{i=1}^n I(V_i=\tau) \log \int_{b_0} \left[\exp \left\{ -\frac{1}{2} b_0^T b_0 - \sum_{j=1}^{n_i} \frac{B(\hat{\beta}; b_0)}{A(D_i(t_j; \hat{\phi}))} - \sum_{s=1}^s e^{\hat{\Sigma}_s} \int_0^{\tau} e^{Q_{1i}(t, b_0, \hat{\theta})} d\tilde{\Lambda}_s(t) \right\} \right] db_0 + C_7, \end{aligned} \quad (10)$$

where C_7 is a constant. On the other hand, since, for any $\Gamma > 0$ and $x > 0$, $\Gamma \log(1+x/\Gamma)$

$\Gamma x/\Gamma = x$, we have that $e^{-x} (1+x/\Gamma)^{-\Gamma}$. Therefore, with

$$Q_{1i}(t, b_0, \hat{\theta}) \geq -C_1 \|b_0\| - C_2 \|Y_i\| - C_3, \text{ (10) gives that}$$

$$\begin{aligned}
0 &\leq C_7 + \frac{1}{n} \sum_{i=1}^n \Delta_i \left(\sum_{s=1}^S \hat{\zeta}_s \right) + \frac{1}{n} \sum_{i=1}^n I(V_i = \tau) \log \left\{ \Gamma^\Gamma \times \exp \left\{ -\Gamma \sum_{s=1}^S I(S_i = s) \hat{\zeta}_s \right\} \right. \\
&\quad \times \int_{b_0} \left[\exp \left\{ -\frac{1}{2} b_0^T b_0 - \sum_{j=1}^{n_i} \frac{B(\hat{\beta}; b_0)}{A(D_i(t_j; \phi))} + C_1 \Gamma \|b_0\| + C_2 \Gamma \|Y_i\| + C_3 \Gamma \right\} \right] db_0 \left. \right\} \quad (11) \\
&= C_7 + \frac{1}{n} \sum_{i=1}^n \Delta_i \left(\sum_{s=1}^S \hat{\zeta}_s \right) - \frac{\Gamma}{n} \sum_{i=1}^n I(V_i = \tau) \left(\sum_{s=1}^S \hat{\zeta}_s \right) + C_8(\Gamma),
\end{aligned}$$

where $C_8(\Gamma)$ is a deterministic function of Γ . For the s -th stratum, (11) is that

$$0 \leq C_7 + \sum_{i=1}^n \Delta_i I(S_i = s) \hat{\zeta}_s - \frac{\Gamma}{n} \sum_{i=1}^n I(V_i = \tau) I(S_i = s) \hat{\zeta}_s + C_8(\Gamma).$$

By the strong law of large numbers, $\sum_{i=1}^n I(V_i = \tau) I(S_i = s) / n \rightarrow P(V_i = \tau, S_i = s) > 0$. Then, we can choose Γ large enough such that

$\sum_{i=1}^n \Delta_i I(S_i = s) / n \leq (\Gamma / 2n) \sum_{i=1}^n I(V_i = \tau) I(S_i = s)$. Thus, we obtain

$$\begin{aligned}
0 &\leq C_7 + C_8(\Gamma) - \frac{\Gamma}{2n} \sum_{i=1}^n I(V_i = \tau) I(S_i = s) \hat{\zeta}_s. \text{ In other words,} \\
\hat{\zeta}_s &\leq \frac{(C_7 + C_8(\Gamma)) 2n}{\Gamma \sum_{i=1}^n I(V_i = \tau) I(S_i = s)} \rightarrow \frac{(C_7 + C_8(\Gamma)) 2}{\Gamma P(V_i = \tau, S_i = s)}.
\end{aligned}$$

If we denote $B_{s0} = \exp\{2(C_7 + C_8(\Gamma)) / (\Gamma P(V_i = \tau, S_i = s))\}$, we conclude that

$\hat{\Lambda}_s(\tau) \leq B_{s0}$, $s=1, \dots, S$. Note that the above arguments hold for every sample in the probability space except a set with zero probability. Therefore, we have shown that, with probability one, $\hat{\Lambda}_s(\tau)$ is bounded for any sample size n .

In the third step, for convenience, we omit the index i . Then, for the number of observed longitudinal measurements per subject, we use n_N instead of the n_i without subscript i since we denoted sample size as n . Use O to abbreviate the observed statistics

$(Y, X, \tilde{X}, V, \Delta, n_N, s)$ and $\{Z(t), \tilde{Z}(t), 0 \leq t \leq V\}$ for a subject, and define

$$\begin{aligned}
G(b, O; \theta, \Lambda_s) &= \exp \left\{ \sum_{j=1}^{n_N} \left[\frac{Y_j(X_j \beta + \tilde{X}_j b) - B(\beta; b)}{A(D(t_j; \phi))} + C(Y_j; D(t_j; \phi)) \right] \right\} \\
&\quad \times \exp \left\{ \Delta \left[\tilde{Z}(V)(\psi \circ b) + Z(V)\gamma \right] - \int_0^V \exp \left\{ \tilde{Z}(t)(\psi \circ b) + Z(t)\gamma \right\} d\Lambda_s(t) \right\} (2\pi)^{-d_b/2} |\Sigma_b|^{-1/2} \exp \left\{ -\frac{1}{2} b^T \Sigma_b^{-1} b \right\}, \\
Q(\nu, O; \theta, \Lambda_s) &= \frac{\int_b G(b, O; \theta, \Lambda_s) \exp \left\{ \tilde{Z}(\nu)(\psi \circ b) + Z(\nu)\gamma \right\} db}{\int_b G(b, O; \theta, \Lambda_s) db},
\end{aligned}$$

and a class

$\mathcal{F} = \{Q(\nu, O; \theta, \Lambda_s(t)) : \nu \in [0, \tau], \theta \in \Theta, \Lambda_s(t) \in \mathbb{W}, \Lambda_s(0) = 0, \Lambda_s(\tau) \leq B_{s0}\}$, where B_{s0} is the constant given in the second step and $\mathbb{W}$ contains all nondecreasing functions in $[0, \tau]$. Employing empirical process formulation, the class $\mathcal{F}$ can be proved to be as P-Donsker by Theorems 2.5.6 and 2.7.5 in van der Vaart and Wellner (1996). In stratum s , denote m_s as

the number of subjects, and V_{sk} and δ_{sk} as the observed time and censoring indicator for the k -th subject, respectively. By differentiating (8) with respect to $\Lambda_s\{V_{sk}\}$, we obtain

$$\hat{\Lambda}_s\{V_{sk}\} = \frac{\Delta_{sk}}{m_s \mathbf{P}_{m_s} \left\{ I(V_s \geq \nu) Q(\nu, O; \hat{\theta}, \hat{\Lambda}_s) \right\} \big|_{\nu=V_{sk}}}.$$

We also construct $\bar{\Lambda}_s(t)$, another step function with the jump size $\bar{\Lambda}_s\{V_{sk}\}$, given by

$$\bar{\Lambda}_s\{V_{sk}\} = \frac{\Delta_{sk}}{m_s \mathbf{P}_{m_s} \left\{ I(V_s \geq \nu) Q(\nu, O; \theta_0, \Lambda_{s0}) \right\} \big|_{\nu=V_{sk}}}.$$

Through the arguments using empirical process and relevant properties of P-Donsker and Glivenko-Cantelli class, we can prove that $\bar{\Lambda}_s(t)$ uniformly converges to $\Lambda_{s0}(t)$ in $[0, \tau]$.

Next, by the bounded convergence theorem, the fact that $\hat{\theta}$ converges to θ^* and $\hat{\Lambda}_s$ weakly converges to Λ_s^* , and the Arzela-Ascoli theorem, we can prove that $\hat{\Lambda}_s(t)$ uniformly

converges to $\Lambda_s^*(t)$. Then, from $n^{-1}l_n(\hat{\theta}, \hat{\Lambda}) - n^{-1}l_n(\theta_0, \bar{\Lambda}) \geq 0$, using the properties of Glivenko-Cantelli class and Kullback-Leibler information, the following holds, with probability one,

$$(\lambda_s^*(V_s))^{\Delta_s} \int_b G(b, O, \theta^*, \Lambda_s^*) db = (\lambda_{s0}(V_s))^{\Delta_s} \int_b G(b, O, \theta_0, \Lambda_{s0}) db. \quad (12)$$

Our proof will be completed if we can show $\theta^* = \theta_0$ and $\Lambda_s^* = \Lambda_{s0}$ from (12). To show that $\beta^* = \beta_0$, $\varphi^* = \varphi_0$ and $\Sigma_b^* = \Sigma_{b0}$, we let $\delta_s = 0$ and $V_s = 0$ in (12). By the comparison of the coefficients of $Y^T Y$ and Y in the exponential part and the constant term out of the exponential part and assumption (A5), we obtain $\varphi^* = \varphi_0$, $\beta^* = \beta_0$ and $\Sigma_b^* = \Sigma_{b0}$. To show that $\psi^* = \psi_0$, $\gamma^* = \gamma_0$ and $\Lambda_s^* = \Lambda_{s0}$, we let $\delta_s = 0$ in (12). By the similar arguments done for $\beta^* = \beta_0$, $\varphi^* = \varphi_0$ and $\Sigma_b^* = \Sigma_{b0}$, both sides of (12) with $\delta_s = 0$ are expressed as the expected values with respect to the random effects b following a multivariate normal distribution with mean

$\Sigma_{b0}^{1/2} \left(\sum_{j=1}^N Y_j \tilde{X}_j / A(D(t_j; \phi_0)) \right)^T$ and covariance Σ_{b0} . By the fact that for any fixed $\tilde{X}$, treating $\tilde{X}^T Y$ as a parameter in the normal family, b is the complete statistic for $\tilde{X}^T Y$, and the assumptions (A2) and (A5), we have $\psi^* = \psi_0$, $\gamma^* = \gamma_0$ and $\Lambda_s^* = \Lambda_{s0}$. Therefore, the proof of Theorem 1 is completed.

B.2. Proof of asymptotic normality – Theorem 2

Asymptotic distribution for the proposed estimators can be shown if we can verify the conditions of Theorem 3.3.1 in van der Vaart and Wellner (1996). Then, it will be shown that the distribution is Gaussian. For completeness, we use Theorem 4 in Parner (1998) which restated the Theorem 3.3.1 of van der Vaart and Wellner (1996).

Theorem 4 (Parner 1998)

Let U_n and U be random maps and a fixed map, respectively, from ξ to a Banach space such that:

- a. $\sqrt{n} (U_n - U) (\hat{\xi}_n) - \sqrt{n} (U_n - U) (\xi_0) = o_p^* (1 + \sqrt{n} \|\hat{\xi}_n - \xi_0\|)$.
- b. The sequence $\sqrt{n} (U_n - U) (\xi_0)$ converges in distribution to a tight random element Z .
- c. the function $\xi \rightarrow U(\xi)$ is Fréchet differentiable at ξ_0 with a continuously invertible derivative ∇U_{ξ_0} (on its range).
- d. U_{ξ_0} and $\hat{\xi}_n$ satisfies $U_n (\hat{\xi}_n) = o_p^* (n^{-1/2})$ and converges in outer probability to ξ_0 .
Then $\sqrt{n} (\hat{\xi}_n - \xi_0) \Rightarrow \nabla U_{\xi_0}^{-1} Z$.

In our situation, the parameter

$\xi_s = (\theta, \Lambda_s) \in \Xi = \left\{ (\theta, \Lambda_s) : \|\theta - \theta_0\| + \sup_{t \in [0, \tau]} |\Lambda_s(t) - \Lambda_{s0}(t)| \leq \delta, s=1, \dots, S \right\}$ for a fixed small constant δ . We define $\mathcal{H} = \left\{ (\mathbf{h}_1, h_2) : \|\mathbf{h}_1\| \leq 1, \|h_2\|_V \leq 1 \right\}$, where $\|h_2\|_V$ is the total variation of h_2 in $[0, \tau]$ defined as

$$\sup_{0=t_0 \leq t_2 \leq \dots \leq t_k = \tau} \sum_{j=1}^k |h_2(t_j) - h_2(t_{j-1})|,$$

and also define that, for stratum s ,

$$U_{m_s}(\xi_s)(\mathbf{h}_1, h_2) = \mathbf{P}_{m_s} \left\{ l_{\theta}(\theta, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\theta, \Lambda_s)[h_2] \right\}$$

and

$$U_s(\xi_s)(\mathbf{h}_1, h_2) = \mathbf{P} \left\{ l_{\theta}(\theta, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\theta, \Lambda_s)[h_2] \right\},$$

where $l_{\theta}(\theta, \Lambda_s)$ is the first derivative of the log-likelihood function from one single subject belonging to stratum s , denoted by $l(O; \theta, \Lambda_s)$, with respect to θ , and $l_{\Lambda_s}(\theta, \Lambda_s)$ is the derivative of $l(O; \theta, \Lambda_{s\varepsilon})$ at $\varepsilon = 0$, where $\Lambda_{s\varepsilon}(t) = \int_0^t (1 + \varepsilon h_2(u)) d\Lambda_{s0}(u)$. Therefore, both U_{m_s} and U_s map from Ξ to $\ell^{\infty}(\mathcal{H})$, and $\sqrt{m_s} \{U_{m_s}(\xi_s) - U_s(\xi_s)\}$ is an empirical process in the space $\ell^{\infty}(\mathcal{H})$.

Denoting $\mathbf{h}_1 = (\mathbf{h}_1^{\beta T}, \mathbf{h}_1^{\phi T}, \mathbf{h}_1^{bT}, \mathbf{h}_1^{\psi T}, \mathbf{h}_1^{\gamma T})^T$ corresponding to $\theta = (\beta^T, \phi^T, \text{Vech}(\Sigma_b)^T, \psi^T, \gamma^T)^T$, for any $(\mathbf{h}_1, h_2) \in \mathcal{H}$, the class

$$\mathcal{G} = \left\{ l_{\theta}(\theta, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\theta, \Lambda_s)[h_2] - l_{\theta}(\theta_0, \Lambda_{s0})^T \mathbf{h}_1 + l_{\Lambda_s}(\theta_0, \Lambda_{s0})[h_2], \right. \\ \left. \left\| \theta - \theta_0 \right\| + \sup_{t \in [0, \tau]} |\Lambda_s(t) - \Lambda_{s0}(t)| \leq \delta, (\mathbf{h}_1, h_2) \in \mathcal{H} \right\}$$

can be shown as P-Donsker. From the P-Donsker property, it is also implied that

$$\sup_{(\mathbf{h}_1, h_2) \in \mathcal{H}} \mathbf{P} \left[l_{\theta}(\theta, \Lambda_s)^T \mathbf{h}_1 + l_{\Lambda_s}(\theta, \Lambda_s)[h_2] - l_{\theta}(\theta_0, \Lambda_{s0})^T \mathbf{h}_1 + l_{\Lambda_s}(\theta_0, \Lambda_{s0})[h_2] \right]^2 \rightarrow 0$$

as $\|\theta - \theta_0\| + \sup_{t \in [0, \tau]} |\Lambda_s(t) - \Lambda_{s0}(t)| \rightarrow 0$. Then, for the conditions (a)–(d) in Theorem 4 of Parner (1998), we have that

(a) follows from Lemma 3.3.5 (p311) of van der Vaart and Wellner (1996);

(b) holds as a result of P-Donsker property and the convergence is defined in the metric space $\ell^{\infty}(\mathcal{H})$ by the Donsker theorem in van der Vaart and Wellner (1996);

(d) is true because $(\hat{\theta}, \hat{\Lambda}_s)$ maximizes $\mathbf{P}_{m_s} l(O; \theta, \Lambda_s)$, (θ_0, Λ_{s0}) maximizes $\mathbf{P} l(O; \theta, \Lambda_s)$, and $(\hat{\theta}, \hat{\Lambda}_s)$ converges to (θ_0, Λ_{s0}) from Theorem 1;

The first half of condition (c), that the function $\xi \rightarrow U(\xi)$ is *Fréchet* differentiable at ξ_0 , is proved by showing there exists a bounded linear operator for the function.

Thus, it only remains to prove the second half of condition (c), that the derivative ∇U_{ξ_0} is continuously invertible on its range $\ell^{\infty}(\mathcal{H})$. From the proof of the *Fréchet* differentiability of $U(\xi)$ at ξ_0 , we have that for any (θ_1, Λ_{s1}) and (θ_2, Λ_{s2}) in Ξ ,

$$\nabla U_{\xi_0}(\theta_1 - \theta_2, \Lambda_{s1} - \Lambda_{s2})[\mathbf{h}_1, h_2] = (\theta_1 - \theta_2)^T \Omega_1[\mathbf{h}_1, h_2] + \int_0^{\tau} \Omega_2[\mathbf{h}_1, h_2] d(\Lambda_{s1} - \Lambda_{s2})(t), \quad (13)$$

where both Ω_1 and Ω_2 are bounded linear operators on $\mathcal{H}$, and $\Omega = (\Omega_1, \Omega_2)$ maps $\mathcal{H} \subset R^d \times BV[0, \tau]$ to $R^d \times BV[0, \tau]$, where $BV[0, \tau]$ contains all the functions with finite total variation in $[0, \tau]$. The explicit expressions of Ω_1 and Ω_2 can be obtained from P-Donsker property and the derivation of ∇U_{ξ_0} by definition. Thus, ∇U_{ξ_0} is a linear operator from $\ell^{\infty}(\mathcal{H})$ to itself. We note that to prove that ∇U_{ξ_0} is continuously invertible is equivalent to showing that Ω is invertible. Then, by Theorem 4.25 of Rudin (1973), for the proof of invertibility of Ω , it is sufficient to verify that Ω is one to one: if $\Omega[\mathbf{h}_1, h_2] = 0$, then, by choosing $\theta_1 - \theta_2 = \varepsilon^* \mathbf{h}_1$ and $\Lambda_{s1} - \Lambda_{s2} = \varepsilon^* \int h_2 d\Lambda_{s0}$ in (13) for a small constant ε^* , we obtain

$$\nabla U_{\xi_0}(\mathbf{h}_1, \int h_2 d\Lambda_{s0})[\mathbf{h}_1, h_2] = \varepsilon^* \begin{pmatrix} \mathbf{h}_1^T, h_2 \end{pmatrix} \begin{pmatrix} \Omega_1[\mathbf{h}_1, h_2] \\ \Omega_2[\mathbf{h}_1, h_2] \end{pmatrix} = \varepsilon^* \begin{pmatrix} \mathbf{h}_1^T, h_2 \end{pmatrix} \Omega[\mathbf{h}_1, h_2] = 0.$$

Since $\nabla U_{\xi_0}(\mathbf{h}_1, \int h_2 d\Lambda_{s0})[\mathbf{h}_1, h_2]$ is the negative information matrix in the submodel $(\theta_0 + \varepsilon \mathbf{h}_1, \Lambda_{s0} + \varepsilon \int h_2 d\Lambda_{s0})$, the score function along this submodel is $l_{\theta}(\theta_0, \Lambda_{s0})^T \mathbf{h}_1 + l_{\Lambda_s}(\theta_0, \Lambda_{s0})[h_2] = 0$; that is, with probability one, the numerator of the score function

$$\begin{aligned} 0 = & \int_b G(b, O; \theta_0, \Lambda_{s0}) \times \left[\frac{b^T \Sigma_{b0}^{-1} D_b \Sigma_{b0}^{-1} b}{2} - \frac{1}{2} \text{Tr} \left(\Sigma_{b0}^{-1} D_b \right) \right. \\ & + \sum_{j=1}^{n_N} \left\{ - \left(\frac{Y_j (X_j \beta_0 + \tilde{X}_j b) - B(\beta_0; b)}{A(D(t_j; \phi_0))^2} \right) \left(A'(D(t_j; \phi_0)) \right) h_1^\phi + C'(Y_j; D(t_j; \phi_0)) h_1^\phi \right\} \\ & + \sum_{j=1}^{n_N} \left\{ \frac{Y_j X_j}{A(D(t_j; \phi_0))} \mathbf{h}_1^\beta - B'(\beta_0; b) \mathbf{h}_1^\beta \right\} + \Delta_s \left\{ \left(\tilde{Z}(V_s) \circ \mathbf{h}_1^\psi \right)^T b + Z(V_s) \mathbf{h}_1^\gamma \right\} \\ & - \int_0^{V_s} \exp \left\{ \tilde{Z}(t) (\psi_0 \circ b) + Z(t) \gamma_0 \right\} \times \left\{ \left(\tilde{Z}(t) \circ \mathbf{h}_1^\psi \right)^T b + Z(t) \mathbf{h}_1^\gamma \right\} d\Lambda_{s0}(t) \Big] db \\ & + \int_b G(b, O; \theta_0, \Lambda_{s0}) \times \left[\Delta_s h_2(V_s) \right. \\ & \quad \left. - \int_0^{V_s} h_2(t) \exp \left\{ \tilde{Z}(t) (\psi_0 \circ b) + Z(t) \gamma_0 \right\} d\Lambda_{s0}(t) \right] db, \end{aligned} \quad (14)$$

where $A'(D(t_j; \phi_0))$ and $C'(Y_j; D(t_j; \phi_0))$ are the derivatives of $A(D(t_j; \phi))$ and $C(Y_j; D(t_j; \phi))$ with respect to ϕ evaluated at ϕ_0 and $B'(\beta_0; b)$ is the derivative of $B(\beta; b)$ with respect to β evaluated at β_0 . The proof of invertibility of Ω will be completed if we can show $\mathbf{h}_1 = 0$ and $h_2(t) = 0$ from (14).

To show $\mathbf{h}_1 = 0$, particularly we let $\gamma_s = 0$ and $V_s = 0$ in (14). Examining the coefficient for Y and the constant terms without Y and using assumption (A5) and the similar arguments done in Appendix B.1 give $h_1^\phi = 0$, $h_1^\beta = 0$ and $D_b = 0$. On the other hand, letting $\gamma_s = 0$ in (14) with assumptions (A2) and (A5) and the similar arguments done in Appendix B.1 lead to $h_1^\psi = 0$, $h_1^\gamma = 0$ and $h_2(t) = 0$. Hence, the proof of condition (c) is completed.

Since the conditions (a)–(d) have been proved, Theorem 3.3.1 of van der Vaart and Wellner (1996) concludes that $\sqrt{m_s}(\hat{\theta} - \theta_0, \hat{\Lambda}_s - \Lambda_{s0})$ weakly converges to a tight random element in $\ell^\infty(\mathcal{H})$. Then, we have

$$\begin{aligned} & \sqrt{m_s} \nabla U_{\psi_0}(\hat{\theta} - \theta_0, \hat{\Lambda}_s - \Lambda_{s0})[\mathbf{h}_1, h_2] \\ &= \sqrt{m_s}(\mathbf{P}_{m_s} - \mathbf{P}) \left\{ l_{\theta}(\theta_0, \Lambda_{s0})^T \mathbf{h}_1 + l_{\Lambda_s}(\theta_0, \Lambda_{s0})[h_2] \right\} + o_P(1), \end{aligned}$$

where $o_P(1)$ is a random variable which converges to zero in probability in $\ell^\infty(\mathcal{H})$, and, from (13),

$$\begin{aligned} & \sqrt{m_s} \nabla U_{\psi_0}(\hat{\theta} - \theta_0, \hat{\Lambda}_s - \Lambda_{s0})[\mathbf{h}_1, h_2] \\ &= \sqrt{m_s} \left\{ (\hat{\theta} - \theta_0)^T \Omega_1[\mathbf{h}_1, h_2] + \int_0^\tau \Omega_2[\mathbf{h}_1, h_2] d(\hat{\Lambda}_s - \Lambda_{s0})(t) \right\}. \end{aligned}$$

By denoting $(\mathbf{h}_1^*, h_2^*) = \Omega^{-1}(\mathbf{h}_1, h_2)$, we have $(\mathbf{h}_1, h_2) = \Omega(\mathbf{h}_1^*, h_2^*)$, and by replacing $(\mathbf{h}_1, h_2)$ with $(\mathbf{h}_1^*, h_2^*)$ in the above two equations, we obtain

$$\begin{aligned} & \sqrt{m_s} \left\{ \left(\hat{\theta} - \theta_0 \right)^T \mathbf{h}_1 + \int_0^\tau h_2(t) d \left(\hat{\Lambda}_s - \Lambda_{s0} \right) (t) \right\} \\ &= \sqrt{m_s} \left(\mathbf{P}_{m_s} - \mathbf{P} \right) \left\{ l_\theta(\theta_0, \Lambda_{s0})^T \mathbf{h}_1^* + l_{\Lambda_s}(\theta_0, \Lambda_{s0}) [h_2^*] \right\} + o_P(1). \end{aligned} \quad (15)$$

We can see that the first term on the right-hand side in (15) is

$\sqrt{m_s} \{U_{m_s}(\theta_0, \Lambda_{s0}) - U_s(\theta_0, \Lambda_{s0})\}$, which is an empirical process in the space $\ell^\infty(\mathcal{H})$.

Furthermore, it is already shown that $\mathcal{G}$ is P-Donsker. Therefore, $\sqrt{m_s}(\hat{\theta} - \theta_0, \hat{\Lambda}_s - \Lambda_{s0})$ weakly converges to a Gaussian process in $\ell^\infty(\mathcal{H})$.

Choose $h_2 = 0$ in (15) with Proposition 3.3.1 in Bickel *et al.* (1993) concludes that $\hat{\theta}$ is an efficient estimator for θ_0 . Therefore, Theorem 2 is proved.

References

- Albert PS, Follmann DA. Modeling Repeated Count Data Subject to Informative Dropout. *Biometrics*. 2000; 56:667–677. [PubMed: 10985201]
- 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]
- Albert PS, Follmann DA, Wang SA, Suh EB. A Latent Autoregressive Model for Longitudinal Binary Data subject to Informative Missingness. *Biometrics*. 2002; 58:631–642. [PubMed: 12229998]
- Bickel, PJ.; Klaassen, CAJ.; Ritov, Y.; Wellner, JA. Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press; Baltimore: 1993.
- Brown ER, Ibrahim JG. A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data. *Biometrics*. 2003; 59:221–228. [PubMed: 12926706]
- 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]
- 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]
- Chen MH, Ibrahim JG, Lipsitz SR. Bayesian methods for missing covariates in cure rate models. *Lifetime Data Anal*. 2002; 8:117–146. [PubMed: 12048863]
- Ding J, Wang JL. Modeling Longitudinal Data with Nonparametric Multiplicative Random Effects Jointly with Survival Data. *Biometrics*. 2008; 64:546–556. [PubMed: 17888040]
- 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.
- Dunson DB, Herring AH. Bayesian latent variable models for mixed discrete outcomes. *Biostatistics*. 2010; 6:11–25. [PubMed: 15618524]
- 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]
- 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]
- Faucett CL, Schenker N, Elashoff RM. Analysis of Censored Survival Data with Intermittently Observed Time-Dependent Binary Covariates. *J Amer Statist Assoc*. 1998; 93:427–437.
- Faucett CJ, Thomas DC. Simultaneously Modeling Censored Survival Data and Repeatedly Measured Covariates: A Gibbs Sampling Approach. *Stat Med*. 1996; 15:1663–1685. [PubMed: 8858789]
- Gueorguieva RV, Agresti A. A Correlated Probit Model for Joint Modeling of Clustered Binary and Continuous Responses. *J Amer Statist Assoc*. 2001; 96:1102–1112.

- Henderson R, Diggle P, Dobson A. Joint Modeling of Longitudinal Measurements and Event Time Data. *Biometrics*. 2000; 4:465–480.
- Hogan J, Laird N. Mixture Models for the Joint Distribution of Repeated Measures and Event Times. *Stat Med*. 1997; 16:239–257. [PubMed: 9004395]
- 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]
- Huang X, Li G, Elashoff 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]
- Huang W, Zeger S, Anthony J, Garrett E. Latent Variable Model for Joint Analysis of Multiple Repeated Measures and Bivariate Event times. *J Amer Statist Assoc*. 2001; 96:906–14.
- Larsen K. Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics*. 2004; 60:85–92. [PubMed: 15032777]
- Lin HQ, McCulloch CE, Mayne ST. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Stat Med*. 2002; 21:2369–2382. [PubMed: 12210621]
- 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.
- Louis TA. Finding the Observed Information Matrix when Using the EM Algorithm. *J. Royal Statist Soc B*. 1982; 44:226–233.
- Parner E. Asymptotic Theory for the Correlated Gamma-frailty Model. *Ann Stat*. 1998; 26:183–214.
- Pulkstenis EP, Ten Have TR, Landis JR. Model for the Analysis of Binary Longitudinal Pain Data Subject to Informative Dropout through Remedication. *J Amer Statist Assoc*. 1998; 93:438–450.
- Ribaudo HJ, Thompson SG, Allen-Mersh TG. A Joint Analysis of Quality of Life and Survival Using a Random Effect Selection Model. *Stat Med*. 2000; 19:3237–3250. [PubMed: 11113957]
- 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.
- Rizopoulos D, Verbeke G, Molenberghs G. Shared Parameter Models under Random Effects Misspecification. *Biometrika*. 2008; 95:63–74.
- Satterthwaite FW. An Approximate Distribution of Estimates of Variance Components. *Biometrics*. 1946; 2:110–114. [PubMed: 20287815]
- Schluchter MD. Methods for the analysis of informatively censored longitudinal data. *Stat Med*. 1992; 11:1861–1870. [PubMed: 1480878]
- 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]
- 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]
- Tseng YK, Hsieh R, Wang JL. Joint Modelling of Accelerated Failure Time and Longitudinal Data. *Biometrika*. 2005; 92:587–603.
- Tsiatis AA, De Gruttola 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.
- Tsiatis AA, Davidian M. A Semiparametric Estimator for the Proportional Hazards Model with Longitudinal Covariates Measured with Error. *Biometrika*. 2001; 88:447–458.
- van der Vaart, AW. Asymptotic Statistics. Cambridge University Press; 1998.
- van der Vaart, AW.; Wellner, JA. Weak Convergence and Empirical Processes. Springer-Verlag; New York: 1996.
- 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.
- Wu M, Bailey K. Estimation and Comparison of Changes in the Presence of Informative Right Censoring: Conditional Linear Model. *Biometrics*. 1989; 45:939–955. [PubMed: 2486189]
- 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.

- Wulfsohn M, Tsiatis AA. A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics*. 1997; 53:330–39. [PubMed: 9147598]
- Xu J, Zeger S. The Evaluation of Multiple Surrogate Endpoints. *Biometrics*. 2001a; 57:81–87. [PubMed: 11252622]
- Xu J, Zeger S. Joint Analysis of Longitudinal Data Comprising Repeated Measures and Times to Events. *Appl Stat*. 2001b; 50:375–387.
- Yao F. Functional approach of flexibly modelling generalized longitudinal data and survival time. *Journal of Statistical Planning and Inference*. 2008; 138:995–1009.
- 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]
- 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]
- 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.

Fig. 1. Estimated baseline cumulative hazards (solid line) with 95% confidence interval (dashed lines) by the simultaneous analysis of HNCS QoL longitudinal outcome and survival time

Fig. 2. Kaplan-Meier estimates (solid line) and the predicted survival probabilities based on the simultaneous analysis of HNCS QoL longitudinal outcome and survival time (dashed line)

Table 1

Summary of simulation results of maximum likelihood estimation for binary longitudinal outcomes and survival time.

ψ	Par.	TRUE	n=200				n=400			
			Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
-1	β_0	-1	-1.019	.271	.273	.953	-1.002	.189	.192	.960
	β_1	1	1.018	.235	.241	.962	1.000	.169	.169	.951
	β_2	-.5	-.494	.413	.396	.942	-.498	.275	.277	.946
	β_3	-.5	-.469	.223	.221	.947	-.484	.158	.154	.946
	σ_b^2	.5	.518	.216	.270	.966	.519	.164	.190	.955
	ψ	-1	-.967	.471	.584	.923	-.988	.364	.420	.921
	γ_1	-1	-1.002	.242	.254	.961	-1.002	.177	.183	.960
	γ_2	1	1.001	.379	.393	.962	.995	.278	.279	.953
	$\Lambda(.9)$	.9	.922	.228	.227	.959	.909	.167	.157	.943
	$\Lambda(1:4)$	1.4	1.442	.397	.389	.944	1.421	.283	.269	.952
0	$\Lambda(1:9)$	1.9	1.956	.594	.600	.953	1.950	.456	.426	.949
	β_0	-1	-1.016	.281	.276	.957	-1.004	.191	.193	.956
	β_1	1	1.003	.258	.247	.936	.994	.170	.173	.957
	β_2	-.5	-.476	.414	.401	.941	-.485	.275	.281	.959
	β_3	-.5	-.496	.235	.238	.957	-.498	.169	.167	.948
	σ_b^2	.5	.500	.233	.287	.957	.497	.181	.200	.952
	ψ	0	.021	.331	.378	.996	.001	.235	.244	.990
	γ_1	-1	-1.031	.197	.191	.949	-1.014	.133	.131	.953
	γ_2	1	1.036	.314	.315	.953	1.015	.223	.218	.944
	$\Lambda(.9)$	.9	.912	.189	.187	.952	.906	.134	.129	.942
1	$\Lambda(1:4)$	1.4	1.450	.315	.307	.958	1.418	.215	.207	.950
	$\Lambda(1:9)$	1.9	1.990	.485	.464	.948	1.935	.322	.308	.948
	β_0	-1	-1.014	.273	.284	.952	-1.007	.200	.198	.955
	β_1	1	1.017	.252	.251	.956	1.011	.176	.176	.952
	β_2	-.5	-.518	.428	.412	.953	-.512	.287	.287	.950

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

ψ	Par.	TRUE	n=200				n=400			
			Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
β_3		-.5	-.540	.245	.248	.954	-.520	.176	.174	.946
		.5	.543	.250	.300	.947	.524	.179	.209	.966
σ_b^2		1	.956	.488	.609	.898	.992	.366	.450	.930
		-1	-1.000	.264	.255	.945	-.998	.176	.183	.953
γ_1		1	1.009	.381	.395	.961	.990	.283	.280	.953
		.9	.921	.235	.228	.961	.918	.166	.159	.937
$\Lambda(1:4)$		1.4	1.443	.412	.393	.963	1.430	.275	.271	.956
		1.9	1.976	.677	.620	.957	1.946	.416	.424	.959

Table 2

Summary of simulation results of maximum likelihood estimation for Poisson longitudinal outcomes and survival time.

ψ	Par.	TRUE	n=200					n=400				
			Est.	SSD	ESE	CP	Est.	SSD	ESE	CP	Est.	SSD
-1	β_0	-1	-.995	.203	.196	.940	-1.005	.138	.138	.948		
	β_1	1	.998	.178	.171	.942	1.007	.119	.121	.945		
	β_2	-.5	-.510	.273	.264	.946	-.502	.192	.186	.935		
	β_3	-.5	-.489	.150	.151	.950	-.492	.107	.106	.949		
	σ_b^2	.2	.196	.074	.092	.983	.200	.055	.065	.976		
	ψ	-1	-1.038	.716	.771	.947	-1.003	.503	.514	.934		
	γ_1	-1	-1.025	.227	.228	.969	-1.014	.160	.158	.954		
	γ_2	1	1.034	.371	.358	.947	1.009	.248	.248	.959		
	$\Lambda(.9)$	.9	.918	.222	.208	.943	.909	.141	.142	.944		
	$\Lambda(1.4)$	1.4	1.456	.403	.357	.943	1.424	.235	.237	.953		
	$\Lambda(1.9)$	1.9	1.999	.632	.560	.953	1.948	.368	.366	.961		
0	β_0	-1	-1.007	.202	.199	.943	-1.003	.137	.140	.957		
	β_1	1	1.010	.184	.175	.932	.998	.120	.124	.951		
	β_2	-.5	-.513	.283	.268	.937	-.505	.193	.189	.942		
	β_3	-.5	-.500	.164	.161	.951	-.489	.114	.114	.950		
	σ_b^2	.2	.199	.074	.097	.981	.206	.055	.070	.973		
	ψ	0	.006	.571	.610	.993	.011	.391	.387	.978		
	γ_1	-1	-1.039	.188	.194	.958	-1.015	.128	.132	.953		
	γ_2	1	1.021	.326	.319	.948	1.003	.226	.219	.944		
	$\Lambda(.9)$	.9	.917	.191	.188	.952	.909	.129	.130	.950		
	$\Lambda(1.4)$	1.4	1.448	.313	.308	.953	1.432	.206	.210	.950		
	$\Lambda(1.9)$	1.9	2.006	.478	.473	.954	1.966	.312	.316	.950		
1	β_0	-1	-1.014	.195	.202	.952	-1.004	.138	.142	.954		
	β_1	1	1.014	.180	.178	.954	1.008	.126	.125	.951		
	β_2	-.5	-.512	.273	.271	.947	-.511	.190	.191	.945		

ψ	Par.	TRUE	n=200				n=400			
			Est.	SSD	ESE	CP	Est.	SSD	ESE	CP
β_3		-.5	-.514	.174	.172	.943	-.509	.123	.122	.959
		.2	.201	.083	.098	.967	.204	.060	.070	.960
σ_b^2		1	.993	.664	.768	.942	1.008	.477	.512	.942
		-1	-1.030	.230	.224	.952	-1.003	.158	.157	.950
γ_1		1	1.014	.363	.354	.949	1.006	.252	.247	.942
		.9	.925	.22	.207	.949	.91	.144	.142	.948
$\Lambda(.9)$		1.4	1.46	.389	.351	.95	1.435	.246	.237	.942
		1.9	2.018	.639	.554	.957	1.957	.373	.365	.947

Table 3

Analyses results for the HNCS and survival time of the CHANCE study

Parameter		Simultaneous			Separate		
		Est.	ESE	P-value	Est.	ESE	P-value
< HNCS QoL longitudinal model >							
Intercept	β_0	.744	.538	.167	1.190	.390	.002
Race (ref= White)							
– African American	β_1	.564	.229	.014	.511	.256	.047
# of 12 oz. beers consumed per week (ref=30 or more)							
– None	β_2	.636	.269	.018	.622	.300	.038
– less than 1	β_3	.830	.357	.020	.735	.396	.064
– 1 to 4	β_4	1.302	.294	<.001	1.268	.326	<.001
– 5 to 14	β_5	1.060	.251	<.001	1.018	.279	<.001
– 15 to 29	β_6	.601	.289	.037	.547	.327	.095
Household income (ref= level1: 0–10K)							
– level2: 20–30K	β_7	–.271	.231	.241	–.328	.258	.204
– level3: 40–50K	β_8	.297	.255	.245	.250	.282	.376
– level4: 60K	β_9	1.199	.274	<.001	1.045	.286	<.001
Radiation therapy (ref= No)							
– Yes	β_{10}	–1.132	.260	<.001	–1.048	.280	<.001
Tumor stage (ref= I)							
– II	β_{11}	–.416	.300	.166	–.352	.330	.286
– III	β_{12}	–1.335	.284	<.001	–1.198	.314	<.001
– IV	β_{13}	–1.175	.254	<.001	–1.057	.277	<.001
# of persons supported by household income	β_{14}	–.189	.084	.025			
BMI	β_{15}	.041	.015	.007			
Total # of medical conditions reported	β_{16}	–.175	.080	.030			
Time at survey measurement (years)	β_{17}	.241	.066	<.001	.254	.067	<.001
variance of random effects	σ_b^2	.303	.173	.013	1.169	.257	
< Hazards model >							
Random effect coefficient	ψ	–1.427	.946	.131			
# of 12 oz. beers consumed per week (ref=30 or more)							
– None	γ_1	–.772	.386	.045			
– less than 1	γ_2	–.155	.426	.715			
– 1 to 4	γ_3	–.802	.414	.053			
– 5 to 14	γ_4	–1.076	.383	.005			
– 15 to 29	γ_5	–.591	.399	.139			
Household income (ref= level1: 0–10K)							
– level2: 20–30K	γ_6	–.218	.294	.459	–.219	.263	.406

Parameter		Simultaneous			Separate		
		Est.	ESE	P-value	Est.	ESE	P-value
– level3: 40–50K	γ_7	–.941	.371	.011	–.928	.331	.005
– level4: 60K	γ_8	–1.463	.401	<.001	–1.393	.358	<.001
Tumor stage (ref= I)							
– II	γ_9	–.199	.465	.668	–.295	.435	.498
– III	γ_{10}	.235	.433	.588	.136	.389	.727
– IV	γ_{11}	1.059	.360	.003	.914	.295	.002
Total # of medical conditions reported	γ_{12}	.256	.110	.020	.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.