

# Joint Models of Longitudinal Data and Recurrent Events with Informative Terminal Event

Sehee Kim<sup>1,\*</sup>, Donglin Zeng<sup>2</sup>, Lloyd Chambless<sup>2</sup>, and Yi Li<sup>1</sup>

\*shkim@hsph.harvard.edu

<sup>1</sup>Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, U.S.A.

<sup>2</sup>Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, U.S.A.

## Abstract

This article presents semiparametric joint models to analyze longitudinal data with recurrent event (e.g. multiple tumors, repeated hospital admissions) and terminal event such as death. A broad class of transformation models for the cumulative intensity of the recurrent events and the cumulative hazard of the terminal event is considered, which includes the proportional hazards model and the proportional odds model as special cases. We propose to estimate all the parameters using the nonparametric maximum likelihood estimators (NPMLE). We provide the simple and efficient EM algorithms to implement the proposed inference procedure. Asymptotic properties of the estimators are shown to be asymptotically normal and semiparametrically efficient. Finally, we evaluate the performance of the method through extensive simulation studies and a real-data application.

KEY WORDS: Joint models; Longitudinal data; Nonparametric maximum likelihood; Random effects; Recurrent events; Repeated measures; Terminal event; Transformation models.

# 1 Introduction

In many biomedical studies, data are collected from patients who experience the same event at multiple times, such as repeated hospital admissions or medical emergency episodes, recurrent strokes, multiple infection episodes, and tumor recurrences. At the same time, some longitudinal biomarkers are observed either at the occurrence time of the events or at regular clinic visits. In addition, some subjects may experience a terminal event such as death. The work described in this paper arose from the Atherosclerosis Risk in Communities (ARIC) study where there was a growing need for a new model to investigate the association between recurrent coronary heart diseases (CHD) and systolic blood pressures over time whose observations were censored by the patient's death. It was concerned that the usual independent censorship assumption may be inappropriate because we observed some patients died soon after the recurrent CHD event occurred. As longitudinal markers, recurrent events, and death appear to be dependent on and informative of one another, analyzing one or two of these processes but ignoring the dependence from the other processes may lead to bias or result in inefficient inference. Therefore, it is important to jointly model longitudinal markers, recurrent events, and death altogether. In this way, we will be able to make the most efficient use of all data and identify effects of variables after correctly controlling the interplay among these processes.

A number of authors have studied jointly modeling longitudinal outcomes and a terminal event, but not considering recurrent events. Among them, Wulfsohn and Tsiatis (1997), Tsiatis and Davidian (2001, 2004), Hsieh et al. (2006), and Song and Wang (2008) presented joint models for survival endpoint and longitudinal covariates with measurement errors. The same joint modeling approach has also been studied for other purposes where both longitudinal and survival data were outcomes of interest (Xu and Zeger, 2001). Vonesh et al. (2006) addressed the need of jointly modeling for the analysis of repeated measures with informative censoring time. The approach by Henderson et al. (2000) could incorporate longitudinal data with either a recurrent or terminal event, but not both.

For the analysis of longitudinal data with informative observation times, a variety of methods have been proposed. Among them, most commonly used approaches were the marginal models for both longitudinal data and time processes (Lin and Ying, 2001; Lin et al., 2004; Sun et al., 2005, 2007). Under these marginal models, it is challenging to obtain efficient estimators and also impossible to predict future outcomes of an individual given the past information. An alternative approach was suggested by Liang et al. (2009); they studied the joint modeling approach using random effects, requiring specified link functions of the random effects.

Joint models have also been developed for analyzing a recurrent and a terminal events data. Wang et al. (2001), Liu et al. (2004), Rondeau et al. (2007), Ye et al. (2007), and Huang and Liu (2007) adopted a common gamma frailty to account for the dependence of recurrent events on death or informative censoring (drop-out), while Huang and Wang (2004) made no assumption of censoring time and random effect. Most of the existing work required the proportionality and assumed time-independent covariates. Recently, Zeng and Lin (2009) developed transformation models that can deal with non-proportional hazards as well as time-varying covariates.

However, there has been scant literature considering the dependence of repeated measures on both recurrent and terminal events. Both Liu et al. (2008) and Liu and Huang (2009) have proposed to use joint modeling approach to analyze such data under the proportionality assumption. In case where the proportionality does not hold, their joint models may yield biased estimators as demonstrated in examples in Web Table 1. In their inference procedure, the piecewise constant functions were adopted for estimating baseline intensity and hazard functions; however, there was no general rule for selecting the number of knots that led to the best reflection of the underlying baseline functions. Moreover, theoretical properties of the suggested estimators have not been established.

In this paper, we will use general transformation models for modeling both recurrent and terminal events, while accounting for the dependence between these two event processes and

longitudinal outcomes. Our transformation models, including the proportional hazards and the proportional odds models as special cases, can prevent possible model mis-specification errors. Our numerical results demonstrate severe biases in regression parameter estimators when the proportional hazards/odds model is misspecified. The rest of the article is organized as follows. In Section 2, we introduce joint models for longitudinal measurements and recurrent events in the presence of a terminal event. In Section 3, we estimate all the parameters by the nonparametric maximum likelihood estimation (NPMLE) and provide simple and efficient algorithms to implement the proposed inference procedure. The theoretical work that shows the weak convergence and efficiency of the proposed NPMLEs is given in Section 4. Sections 5-6 evaluate the numerical performance of the proposed method through extensive simulation studies and through the application to the Atherosclerosis Risk in Communities (ARIC) study. We conclude with some remarks in Section 7.

## 2 Joint Models

Let  $Y(t)$  denote the longitudinal outcome measured at time  $t$ ,  $N^*(t)$  denote the number of recurrent events occurring by time  $t$ , and  $T$  be the time to a terminal event. We introduce latent random effects to account for the association among these processes. Particularly, let  $b^T = (b_1^T, b_2^T)$  denote the subject-specific random effects following a multivariate normal distribution with mean-zeros and covariance matrix  $\Sigma_b$ ;  $b_1$  is used to explain the correlation between longitudinal outcomes, and  $b_2$  (i.e., frailty term) is used for the correlation in recurrent event times. Let  $Z$  be a vector of external covariates (possibly time-varying), including the unit component. We assume that  $Y(\cdot)$ ,  $N^*(\cdot)$ , and  $T$  are independent, conditional on  $Z$  and  $b$ . We then propose the following joint models combining the conditional longitudinal process  $Y(t|Z; b)$  and the conditional cumulative intensity and hazard functions  $\Lambda_R^*(t|Z; b)$

and  $\Lambda_T^*(t|Z; b)$ , respectively:

$$Y(t|Z; b) = \beta_1^T Z_1(t) + b_1^T \tilde{Z}_1(t) + \epsilon(t), \quad (1)$$

$$\Lambda_R^*(t|Z; b) = G_R\left(\int_0^t \exp\left\{\beta_2^T Z_2(s) + b_2^T \tilde{Z}_2(s)\right\} d\Lambda_R(s)\right), \quad (2)$$

$$\Lambda_T^*(t|Z; b) = G_T\left(\int_0^t \exp\left\{\beta_3^T Z_3(s) + (b \circ \phi)^T \tilde{Z}_3(s)\right\} d\Lambda_T(s)\right), \quad (3)$$

where  $\beta^T = (\beta_1^T, \beta_2^T, \beta_3^T)$  is a vector of unknown regression parameters,  $\Lambda_R(\cdot)$  and  $\Lambda_T(\cdot)$  are unspecified increasing, baseline cumulative intensity and hazard functions,  $\phi^T = (\phi_1^T, \phi_2^T)$  is a set of unknown constants, and  $b \circ \phi$  denotes the component-wise product. Both  $Z_i(t)$  and  $\tilde{Z}_i(t)$  ( $i = 1, 2, 3$ ) are some subsets of  $Z$ , but  $Z_2(t)$  and  $Z_3(t)$  do not have the unit component. This allows each of three outcomes  $\{Y(\cdot), N^*(\cdot), T\}$  to depend on different sets of predictors. Additionally,  $\epsilon(t)$  is a Gaussian white noise process with mean-zero and variance  $\sigma_e^2$ . Both  $G_R$  and  $G_T$  are continuously differentiable and strictly increasing transformation functions to be specified in the analysis. For example,  $G_R(x)$  and  $G_T(x)$  can take a form of the logarithmic transformation,

$$\begin{cases} \log(1 + \gamma_L x) / \gamma_L, & \gamma_L > 0 \\ x, & \gamma_L = 0, \end{cases} \quad (4)$$

or a form of the Box-Cox transformation,

$$\begin{cases} \{(1 + x)^{\gamma_{BC}} - 1\} / \gamma_{BC}, & \gamma_{BC} > 0 \\ \log(1 + x), & \gamma_{BC} = 0. \end{cases} \quad (5)$$

According to the choice of  $\gamma_L$  and  $\gamma_{BC}$  in both classes of transformations, the transformation model can represent the proportional hazards model  $x$  (when  $\gamma_L=0, \gamma_{BC}=1$ ) or the proportional odds model  $\log(x + 1)$  (when  $\gamma_L=1, \gamma_{BC}=0$ ). In fact, these types of transformations arise from a Cox model with missing covariate (Kosorok et al., 2004). Thus, the parameters

can still be interpreted similarly as in the Cox model but conditional on some unobserved missing covariate. The greatest importance of the transformation models is that they lead to correct estimation of predictors and survival prediction, even if neither the Cox model nor the proportional odds model does hold.

Note that models (1) and (2) characterize the dependence between  $Y(\cdot)$  and  $N^*(\cdot)$  via the covariance structure between  $b_1$  and  $b_2$ . For example, when both  $b_1$  and  $b_2$  are scalar, their correlation (say  $\rho$ ) quantifies the degree of such dependence not explained by the observed covariates. Model (3) shows that the terminal event depends on  $Y(\cdot)$  through the shared random effects  $b_1$ , and depends on  $N^*(\cdot)$  through the shared frailty  $b_2$ , respectively; moreover, such dependences can be characterized by  $\phi_1$  and  $\phi_2$ . For example,  $\phi_1 = 0$  implies that i) the dependence between the longitudinal process and the terminal event can be fully explained by the observed covariates, and ii) censoring in the longitudinal process occurs at random.

Let  $C$  be the non-informative censoring time assumed to be independent of  $\{Y(\cdot), N^*(\cdot), T, b\}$  given  $Z$ , and let  $X = \min(T, C)$  denote the observed terminal event time. The observed data for the  $i^{th}$  subject with  $m_i$  repeated measurements are  $\{Y_i(t_{ik}), N_i(t), X_i, \Delta_i, Z(t); t_{ik} \leq X_i, t \leq X_i, i = 1, \dots, n, k = 1, \dots, m_i\}$ , where  $N_i(t) = N_i^*(t \wedge X_i)$ ,  $\Delta_i = I(T_i \leq C_i)$  with  $I(\cdot)$  being the indicator function. Under models (1)-(3), the log-likelihood function of the observed data is given by

$$\begin{aligned} & \sum_{i=1}^n \log \int_b \prod_{k=1}^{m_i} \left[ \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp \left\{ \frac{-(Y_i(t_{ik}) - \beta_1^T Z_{1i}(t_{ik}) - b_1^T \tilde{Z}_{1i}(t_{ik}))^2}{2\sigma_e^2} \right\} \right] \\ & \times \prod_t \left[ \lambda_R(t) e^{\beta_2^T Z_{2i}(t) + b_2^T \tilde{Z}_{2i}(t)} G'_R \left( \int_0^t e^{\beta_2^T Z_{2i}(s) + b_2^T \tilde{Z}_{2i}(s)} d\Lambda_R(s) \right) \right]^{R_i(t)\Delta N_i^*(t)} \\ & \times \left[ \lambda_T(X_i) e^{\beta_3^T Z_{3i}(X_i) + (b\circ\phi)^T \tilde{Z}_{3i}(X_i)} G'_T \left( \int_0^{X_i} e^{\beta_3^T Z_{3i}(t) + (b\circ\phi)^T \tilde{Z}_{3i}(t)} d\Lambda_T(t) \right) \right]^{\Delta_i} \\ & \times \exp \left\{ -G_R \left( \int_0^{X_i} e^{\beta_2^T Z_{2i}(t) + b_2^T \tilde{Z}_{2i}(t)} d\Lambda_R(t) \right) - G_T \left( \int_0^{X_i} e^{\beta_3^T Z_{3i}(t) + (b\circ\phi)^T \tilde{Z}_{3i}(t)} d\Lambda_T(t) \right) \right\} \\ & \times f(b; \Sigma_b) db, \end{aligned}$$

where  $R_i(t) = I(X_i \geq t)$  is the indicator for the risk set,  $\Delta N_i^*(t)$  denotes the jump size of the

underlying recurrent event at time  $t$ ,  $f(b; \Sigma_b)$  denotes a multivariate normal density function of  $b$ , and  $\lambda_R(t) = \Lambda'_R(t)$  and  $\lambda_T(t) = \Lambda'_T(t)$  are the derivatives of  $\Lambda_R$  and  $\Lambda_T$ , respectively. Note in (1) and (2) that the observation times of longitudinal outcomes do not need to be the same as the recurrent event times. Instead, the longitudinal measures may be observed at some scheduled visits or at the times when the recurrent events occur.

### 3 Inference Procedure

#### 3.1 Nonparametric Maximum Likelihood Estimation

We propose to use the nonparametric maximum likelihood estimation (NPMLE) for estimating parameters  $(\beta, \phi, \sigma_e^2, \Sigma_b)$  and infinite-dimensional parameters  $\Lambda_R(t)$  and  $\Lambda_T(t)$ . In the log-likelihood, we assume the cumulative intensity function  $\Lambda_R(t)$  and the cumulative hazard function  $\Lambda_T(t)$  to be step functions with the jumps at the observed event times, and we replace the intensity function  $\lambda_R(t)$  and the hazard function  $\lambda_T(t)$  with the jump size of  $\Lambda_R$  and  $\Lambda_T$  at time  $t$ , denoted by  $\Lambda_R\{t\}$  and  $\Lambda_T\{t\}$ , respectively. The modified log-likelihood  $l_n(\beta, \phi, \sigma_e^2, \Sigma_b, \Lambda_R, \Lambda_T)$  is given by

$$\begin{aligned}
& \sum_{i=1}^n \log \int_b \prod_{k=1}^{m_i} \left[ \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp \left\{ \frac{-(Y_i(t_{ik}) - \beta_1^T Z_{1i}(t_{ik}) - b_1^T \tilde{Z}_{1i}(t_{ik}))^2}{2\sigma_e^2} \right\} \right] \\
& \times \prod_t \left[ \Lambda_R\{t\} q_{2i}(t) G'_R \left( \sum_{j: w_{jr} \leq t} q_{2i}(w_{jr}) \Lambda_R\{w_{jr}\} \right) \right]^{R_i(t) \Delta N_i^*(t)} \\
& \times \exp \left\{ -G_R \left( \sum_{j: w_{jr} \leq X_i} q_{2i}(w_{jr}) \Lambda_R\{w_{jr}\} \right) - G_T \left( \sum_{j: w_{jt} \leq X_i} q_{3i}(w_{jt}) \Lambda_T\{w_{jt}\} \right) \right\} \\
& \times \left[ \Lambda_T\{X_i\} q_{3i}(X_i) G'_T \left( \sum_{j: w_{jt} \leq X_i} q_{3i}(w_{jt}) \Lambda_T\{w_{jt}\} \right) \right]^{\Delta_i} \times f(b; \Sigma_b) db, \quad (6)
\end{aligned}$$

where  $q_{2i}(t) = \exp\{\beta_2^T Z_{2i}(t) + b_2^T \tilde{Z}_{2i}(t)\}$ ,  $q_{3i}(t) = \exp\{\beta_3^T Z_{3i}(t) + (b \circ \phi)^T \tilde{Z}_{3i}(t)\}$ , and  $w_{jr}$  and  $w_{jt}$  are the  $j^{\text{th}}$  ordered observed recurrent and terminal event times, respectively. Hence the

likelihood can be expressed as a function of a finite number of parameters, which include  $(\beta, \phi, \sigma_e^2, \Sigma_b)$  and the jump sizes of  $\Lambda_R$  and  $\Lambda_T$ .

### 3.2 EM Algorithm

To obtain the NPMLEs and their variance estimators, we use the expectation-maximization (EM) algorithm (Dempster et al., 1977), treating  $b$  as missing data. In the E-step, we compute conditional expectations of the log-likelihood for the complete data, given the observed data and current parameter estimates. Particularly, using a numerical approximation method such as the Gaussian quadrature, we can evaluate the integration of certain functions of  $b$ , say  $g(b)$ . We denote such expectation by  $\hat{E}[g(b) | Y_i(t), N_i(t), X_i, \Delta_i, Z(t)]$ , hereafter abbreviated as  $\hat{E}[g(b)]$ . In the M-step, we maximize the conditional expectation of the complete data log-likelihood function given the observed data. Specifically, the closed-form of the maximizers exist for  $(\beta_1, \sigma_e^2, \Sigma_b)$  as follows:

$$\begin{aligned}\hat{\beta}_1 &= (Z_1^T Z_1)^{-1} Z_1^T (Y - \hat{E}[\tilde{Z}_1 b_1]), \\ \hat{\sigma}_e^2 &= \hat{E}[(Y - Z_1 \beta_1 - \tilde{Z}_1 b_1)^T (Y - Z_1 \beta_1 - \tilde{Z}_1 b_1)] / \sum_{i=1}^n m_i, \\ \hat{\Sigma}_b &= \hat{E}[b b^T],\end{aligned}$$

where  $Y$  denotes the vector of longitudinal measurements at the observed times, and  $Z_1$  and  $\tilde{Z}_1$  denote matrices with each row equal to the observed covariates  $Z_1(t)^T$  and  $\tilde{Z}_1(t)^T$  at the same times, respectively. For the rest of parameters  $(\beta_2, \beta_3, \phi, \Lambda_R\{\cdot\}, \Lambda_T\{\cdot\})$ , the quasi-Newton algorithm is used to update the parameter estimates at each M-step.

When covariates of the recurrent and terminal events  $(Z_2, \tilde{Z}_2, Z_3, \tilde{Z}_3)$  are time-independent, we propose to use recursive formulae, provided in Web Appendix 1, which reduce the number of parameters to be maximized to a very small set of parameters. Basic ideas of the recursive formulae can be described as follows. In the forward recursive formula, since  $\Lambda_R(t)$



and  $\Lambda_T(t)$  can be calculated from the jumps which are observed before time  $t$ , only ( $\lambda_{1R}$ ,  $\lambda_{1T}$ ) are involved in the quasi-Newton iteration, where  $\lambda_{1R}$  and  $\lambda_{1T}$  are the jump sizes at the first observed event times of the recurrent and terminal events, respectively. This forward recursive formula is applicable to time-varying covariates with a slight modification. In the backward recursive formula, similarly,  $\Lambda_R(t)$  and  $\Lambda_T(t)$  can be expressed as a function of the jumps which are observed after time  $t$  and the sum of all jumps. Thus, the backward recursive formula requires to maximize only the last jump sizes and the sums of all observed jump sizes of the recurrent and terminal events. [The R code is available at <http://xxx>.]

To estimate the covariance matrix of the NPMLEs, we compute the observed information matrix via the Louis formula (Louis, 1982) as given in Web Appendix 1. Then, the inverse of the observation information is the estimator of the covariance matrix of the NPMLEs.

In implementation of the proposed EM algorithm, the choice of reasonable starting values becomes more important as model complexity grows. In a simple random intercept model, we use 0 for the initial values of  $\beta$  and  $\phi$ , 1 for variances,  $1/m_r$  for  $\Lambda_R\{\cdot\}$ , and  $1/m_t$  for  $\Lambda_T\{\cdot\}$ , where  $m_r$  and  $m_t$  are the numbers of the observed recurrent and terminal events, respectively. However, with the same initial setting in the random intercept and slope models, we experienced 1% of convergence failure in our simulation studies; the rest of them converged within 50 iterations, on average. We think this was because of computational instability brought to the terminal event component by the additional variation (the random slope). As soon as we used reasonable initial values for these terminal jumps (the inverse of the size of risk set for  $\Lambda_T\{\cdot\}$ , i.e.,  $1/\sum_{i=1}^n I(X_i \geq w_{jt})$ ), numerical convergence problems were not encountered in any simulated datasets or real data examples we analyzed. The EM algorithm we construct here is more reliable than imagined because the quasi-Newton iteration involves only a small number of parameters due to the proposed recursive formulae in Web Appendix 1.

## 4 Asymptotic Properties

Let  $\theta$  be the vector of  $(\beta, \phi, \sigma_e^2, \text{Vec}(\Sigma_b))$  and let  $(\theta_0, \Lambda_{0R}(t), \Lambda_{0T}(t))$  be the true parameter values of  $(\theta, \Lambda_R(t), \Lambda_T(t))$ , where  $\text{Vec}(\Sigma_b)$  denotes the vector consisting of the upper triangular elements of  $\Sigma_b$ . We then establish the asymptotic properties of the NPMLEs under the following conditions:

(A1) The parameter value  $\theta_0$  belongs to the interior of a compact set  $\Theta$  within the domain of  $\theta$ . Additionally,  $\Lambda'_{0R}(t) > 0$  and  $\Lambda'_{0T}(t) > 0$ , for all  $t \in [0, \tau]$ , where  $\tau$  is the duration of the study.

(A2) With probability 1,  $Z(\cdot)$  is left-continuous with uniformly bounded left and right derivatives in  $[0, \tau]$ .

(A3) For some constant  $\delta_0$ ,  $P(C \geq \tau | Z) > \delta_0 > 0$  with probability 1.

(A4)  $E[N^*(\tau)] < \infty$  with probability 1.

(A5) For some positive constant  $M_0$  and  $\|c\| = 1$ ,  $M_0^{-1} < \sigma_{0e}^2 < M_0$  and  $M_0^{-1} < c^T \Sigma_{0b} c < M_0$ .

(A6) The transformation functions  $G_R(\cdot)$  and  $G_T(\cdot)$  are four-times differentiable with  $G_R(0) = G_T(0) = 0$ ,  $G'_R(0) > 0$ , and  $G'_T(0) > 0$ . In addition, there exist positive constants  $\mu_0$  and  $\kappa_0$  such that for any integer  $m \geq 0$  and for any sequence  $0 < x_1 < \dots < x_m \leq y$ ,

$$\prod_{j=1}^m \{(1 + x_j) G'_R(x_j)\} \exp\{-G_R(y)\} \leq \mu_0^m (1 + y)^{-\kappa_0}$$

and

$$(1 + x) G'_T(x) \exp\{-G_T(x)\} \leq \mu_0 (1 + x)^{-\kappa_0}.$$

Furthermore, there exists a constant  $\rho_0 > 0$  such that

$$\sup_x \left\{ \frac{|G''_R(x)| + |G_R^{(3)}(x)| + |G_R^{(4)}(x)|}{G'_R(x) (1 + x)^{\rho_0}} \right\} + \sup_x \left\{ \frac{|G''_T(x)| + |G_T^{(3)}(x)| + |G_T^{(4)}(x)|}{G'_T(x) (1 + x)^{\rho_0}} \right\} < \infty,$$

where  $G_R^{(3)}$ ,  $G_R^{(4)}$ ,  $G_T^{(3)}$ , and  $G_T^{(4)}$  are the third and fourth derivatives.

(A7) For some  $t \in [0, \tau]$ , if there exist a deterministic function  $c(t)$  and  $v$  such that  $c(t) +$

$v^T Z(t) = 0$  with probability 1, then  $c(t) = 0$  and  $v = 0$ .

(A8) For some  $t \in [0, \tau]$ ,  $\tilde{Z}_i^T(t)\tilde{Z}_i(t)$  ( $i = 1, 2$ ) has full rank with some positive probability.

(A9) Let  $K$  be the number of repeated measures and let  $d_b$  be the dimension of  $b_1$ . With probability one,  $P(K > d_b) > 0$ .

Conditions (A1)-(A3) are the standard assumptions for survival analysis. Conditions (A4)-(A5) are necessary to prove the existence of the NPMLEs. It can be easily verified that Condition (A6) holds for all transformations commonly used, including the classes of Box-Cox and logarithmic transformations described in Section 2. Conditions (A7)-(A8) entail the linear independence of design matrices of covariates for the fixed and random effects. Condition (A9) prescribes that some subjects have at least  $d_b$  repeated measures.

Under the above conditions, the following theorems show the consistency and asymptotic normality of  $(\hat{\theta}, \hat{\Lambda}_R, \hat{\Lambda}_T)$  and the asymptotic efficiency of  $\hat{\theta}$ .

**Theorem 1** *Under Conditions (A1)-(A9), almost surely,*

$$|\hat{\theta} - \theta_0| \rightarrow 0, \quad \sup_{t \in [0, \tau]} |\hat{\Lambda}_R(t) - \Lambda_{0R}(t)| \rightarrow 0, \quad \sup_{t \in [0, \tau]} |\hat{\Lambda}_T(t) - \Lambda_{0T}(t)| \rightarrow 0.$$

**Theorem 2** *Under Conditions (A1)-(A9),  $\sqrt{n}(\hat{\theta} - \theta_0, \hat{\Lambda}_R - \Lambda_{0R}, \hat{\Lambda}_T - \Lambda_{0T})$  weakly converges to a zero-mean Gaussian process in  $R^{d_\theta} \times BV[0, \tau] \times BV[0, \tau]$ , where  $d_\theta$  is the dimension of  $\theta$  and  $BV[0, \tau]$  denotes the space of all functions with bounded variations in  $[0, \tau]$ . Furthermore, the asymptotic covariance matrix of  $\sqrt{n}(\hat{\theta} - \theta_0)$  achieves the semiparametric efficiency bound for  $\theta_0$ .*

Furthermore, in Web Appendix 2, we show that the inverse of the observed information matrix is a consistent estimator of the asymptotic covariance matrix of the NPMLEs. This result allows us to make inference for any functional of  $(\theta, \Lambda_R, \Lambda_T)$ . To prove Theorems 1-2, we apply the general asymptotic theory of Zeng and Lin (2007). The desired asymptotic properties of the NPMLEs are established followed by the arguments in Appendix B of

Zeng and Lin (2007) if we can verify that their regularity conditions hold for our joint model setting. Checking the regularity conditions, however, is challenging in our cases. The detailed proofs are provided in Web Appendix 2.

## 5 Simulation Studies

In this section, we examined the performance of the proposed methods through extensive simulation studies. We considered a dichotomous covariate of  $Z_1$  taking the value of 0 or 1 with the equal probability of 0.5 and a continuous covariate of  $Z_2$  randomly sampled from the uniform distribution on  $[-1, 1]$ . We generated data for the longitudinal outcomes from  $Y(t | Z_1, Z_2; b_1) = 0.7 + Z_1 + 0.5Z_2 + b_1 + \epsilon(t)$ , where  $\epsilon(t) \sim N(0, \sigma_\epsilon^2)$  with  $\sigma_\epsilon^2 = 1$ , the recurrent event process from the cumulative intensity  $\Lambda_R^*(t | Z_1, Z_2; b_2) = G_R(e^{Z_1 + 0.5Z_2 + b_2} \Lambda_R(t))$ , where  $\Lambda_R(t) = \nu_1 t$ , and the terminal event time from the cumulative hazard  $\Lambda_T^*(t | Z_1, Z_2; b_1, b_2) = G_T(e^{Z_1 + 0.5Z_2 + b_1\phi_1 + b_2\phi_2} \Lambda_T(t))$ , where  $\Lambda_T(t) = \nu_2 t^2$ .

For each subject, the correlation within repeated measures was reflected by  $b_1 \sim N(0, \sigma_1^2)$ , and the correlation within recurrent event times was reflected by another random effect  $b_2 \sim N(0, \sigma_2^2)$ . In addition, their dependence was given by  $\rho$ , which was the correlation between  $(b_1, b_2)$ . Particularly, we chose  $\sigma_1^2 = \sigma_2^2 = \rho = 0.5$ . We considered two cases of  $\phi = (0.5, 0.2)$  and  $(0, 0.2)$ , where we simulated some positive dependence (i.e.  $\phi_1 = 0.5$ ) or no dependence (i.e.  $\phi_1 = 0$ ) explained by  $b_1$  in the latter.

The non-informative censoring time  $C_i$  was randomly sampled from the uniform distribution on  $[1, 5]$ , and  $(\nu_1, \nu_2)$  was chosen according to the considered transformation models in order to achieve the desired total number of recurrent event times of 2~3 and the desired censoring rate of 35%, on average. We set longitudinal observation times to be fixed intervals so that a subject had about six longitudinal measurements, on average.

The results presented in Table 1 and Table 2 are based on 1000 replications for  $n=200$  and  $n=400$ . Tables 1-2 include the difference between the estimate and true parameter

(Bias), the sample standard deviation of the parameter estimators (SE), and the average of the standard error estimators (SEE), and the coverage probability of 95% confidence intervals (CP). The confidence intervals for  $\Lambda_R(\cdot)$  and  $\Lambda_T(\cdot)$  are constructed based on the log transformation, and those for  $\rho$  are based on the Fisher transformation. In addition, we use the Satterthwaite approximation to compute the confidence intervals of  $\sigma_e^2$ ,  $\sigma_1^2$ , and  $\sigma_2^2$ .

Table 1 shows that the NPMLEs under  $G_R(x) = G_T(x) = x$  are noticeably unbiased, the standard error estimators calculated via the Louis formula well reflect the true variations of the proposed estimators, and the coverage probabilities lie in a reasonable range, even with a moderate sample size 200. As the sample size increases to 400, the estimators have smaller bias, the variations of the parameter estimators become smaller, and the coverage probabilities are more accurate overall. The simulation results shown in Table 2 are similar to Table 1, indicating that the proposed method seems to work well for the transformation  $G_R(x) = x$  and  $G_T(x) = \log(1 + x)$  as well. We also studied other combinations of transformations such as  $\{G_R(x), G_T(x)\} = \{\log(1 + x), x\}$  and  $\{G_R(x), G_T(x)\} = \{\log(1 + x), \log(1 + x)\}$ , and the results are similar and hence omitted here to save space.

To further investigate the performance of the proposed method, we also considered more practical scenarios for the correlation and covariate structures of the longitudinal process; including 1) random intercept and slope, and 2) time-dependent covariate  $Z_2(t) = Z_2t$ . Simulation results reported in Web Tables 2-4 reveal that the good asymptotic properties in the proposed method held even for more complicated correlation structure than just the compound symmetry as well as even for time-dependent covariate processes.

## 6 Application

We applied the proposed method to the data from the Atherosclerosis Risk in Communities (ARIC) study. The cohort study was designed to investigate the trends in rates of hospitalized myocardial infarction (MI) and coronary heart diseases (CHD) in men and women aged

45-64 years from four US communities; Minneapolis suburbs (Minnesota), Forsyth County (North Carolina), Washington County (Maryland), and Jackson County (Mississippi). It is well known that some risk factors for coronary heart diseases differ considerably by race, therefore, our research focused on a total of 870 white patients living in Forsyth County, who were diagnosed with hypertension at the first examination in 1987-89.

The existing studies (Chambless et al., 2003; Wattanakit et al., 2005) found that systolic blood pressure (SBP) was an important risk factor for both incidence and recurrence of CHD in the ARIC data. We also observed from the preliminary analysis that patients who had experienced more recurrent CHD events were likely to be in a higher risk of death. Thus, the primary objective of this analysis was to characterize these relationships between SBP changes over time, recurrent CHD events, and death together, to assess the effects of baseline covariates on these three outcomes, and to utilize the final models for the accurate prediction of risk of recurrent CHD events and death. To model such a complicated system, we propose a joint transformation model for the main outcomes consisting of three components: (a) longitudinal SBP measures, (b) recurrent CHD events, and (c) death.

Beginning with the first screen examination (baseline) in 1987-89, longitudinal measures were collected at approximately three-year intervals, in 1990-92, 1993-95, and 1996-98. The recurrent event of interest was the multiple occurrences of CHD events including MI, which were classified based on Mortality and Morbidity Classification Committee (MMCC) reviews or computer algorithm if MMCC reviews were not required. Follow-up process for the recurrent CHD events and death continued until 2005 through reviewing death certificates and hospital discharge records and investigating out-of-hospital deaths, while the follow-up for longitudinal measures ended with each patient's last examination (up to 1998). The median follow-up time was 16.6 years with the largest follow-up time being 19 years, and 28% of patients died during the study period. 138, 28, and 8 patients experienced one, two, and more than two CHD events, respectively.

In our joint model, we included the baseline covariates: age, sex, total-cholesterol, and

indicators for hypertension lowering medication use, diabetes (with fasting glucose  $\geq 126$  mg/dL) and ever smoker along with visiting time in years. Among them, total-cholesterol was standardized at mean 0 and standard deviation 1, and age variable was centered at the mean age of 54 and divided by 10 to represent a decade. In addition, subject-specific random intercepts  $b_1$  and  $b_2$  were included in the joint model to cope with correlations within and between three outcomes, while quantifying the associations between processes.

We applied the logarithmic and the Box-Cox transformations in Section 2, and the Akaike information criterion (AIC) was used to determine the best transformation model. Figure 1 shows the surfaces of the log-likelihood functions by the combinations of  $G_R(x)$  and  $G_T(x)$  transformations, where the largest log-likelihood value corresponded to the proportional intensity model for the recurrent CHD and the proportional odds model for death, i.e.,  $G_R(x) = x$  and  $G_T(x) = \log(1+x)$ , based on both logarithmic and Box-Cox transformations.

Table 3 summarizes the estimation results under the selected best model. The analysis results using our joint model are reported under the Full Model, while those from the model ignoring the dependent censoring by death are reported under the Reduced Model. In the Full Model, age at entry was significant for all three outcomes; elder patients had a higher SBP level ( $\hat{\beta}_{11} = 0.338$ ,  $p < .001$ ), a higher intensity rate of CHD occurrences ( $\hat{\beta}_{31} = 0.837$ ,  $p = .002$ ), and a higher risk of death ( $\hat{\beta}_{31} = 2.187$ ,  $p < .001$ ). On average, SBP level increased over time ( $\hat{\beta}_{17} = 0.014$ ,  $p = .001$ ), and patients who took hypertension medications tended to have lower SBP levels ( $\hat{\beta}_{14} = -0.546$ ,  $p < .001$ ). The risk of CHD was significantly different by sex; men were likely to be in a higher risk than women ( $\hat{\beta}_{22} = 0.871$ ,  $p = .001$ ). Both smoking and diabetes were associated with the elevation in the risk for CHD and death. Examples of the quantitative interpretations related to applied transformations are as follows: the estimate of diabetes effect  $\hat{\beta}_{26} = 1.843$  can be interpreted as the log relative risk of CHD (the proportional intensity model), and the estimate  $\hat{\beta}_{36} = 2.671$  can be interpreted as the log odds ratio of death (the proportional odds model) for diabetes vs. no diabetes.

For the model association, the significant correlation between  $b_1$  and  $b_2$  ( $\hat{\rho} = 0.241$ ,

$p = .007$ ) suggested that there seemed to be some positive association between SBP levels and recurrence due to unobserved random factors, even after adjusting for the commonly observed covariates. The non-significant  $\hat{\phi}_1$  indicated that the observed covariates in the fitted model appeared to fully explain dependence between the longitudinal SBP pattern and death. In contrast, the highly significant  $\hat{\phi}_2$  appeared to support the positive correlation between recurrence of CHD and death, and comparing the magnitude of  $\hat{\sigma}_2 = 1.76$  relative to  $\hat{\phi}_2 = 1.91$  suggested that the strength of this association varied from patient to patient as much as the common factor across patients. Note that, if we ignore this dependent censoring by death, then the regression parameter estimators in the recurrent event become biased towards the null, in which smaller standard error estimators would be observed as in the Reduced Model (Table 3), while those in the longitudinal process remain very similar to the results from the Full Model. To sum up, these results coincided with our initial expectations that patients with higher SBP would be exposed to a higher risk of CHD and that patients who get admitted to hospital more frequently with CHD would be at even higher risk of death. These findings can easily flow to other interesting application points: 1) to predict the survival distribution after the incidence of CHD at a fixed time  $s$ , and 2) to estimate the expected SBP levels over time after the incidence of CHD at a fixed time  $s$ . To answer the question 1), the conditional survival distribution can be calculated as

$$\begin{aligned}
& P[T > t \mid Z, \Delta N^*(s) = N^*(s) = 1, T > s] \\
&= \frac{\int_b e^{-G_T \left( \int_0^t e^{\beta_3^T Z_3(u) + (b \circ \phi)^T \tilde{Z}_3(u)} d\Lambda_T(u) \right)} f(\Delta N^*(s) = N^*(s) = 1, T > s \mid Z, b) f(b; \Sigma_b) db}{\int_b f(\Delta N^*(s) = N^*(s) = 1, T > s \mid Z, b) f(b; \Sigma_b) db}.
\end{aligned}$$

Also, for the question 2), the conditional expectation of longitudinal SBPs is given by

$$\begin{aligned}
& E[Y(t) \mid Z, \Delta N^*(s) = N^*(s) = 1, T > t] \\
&= \frac{\int_b \left\{ \beta_1^T Z_1(t) + b_1^T \tilde{Z}_1(t) \right\} f(\Delta N^*(s) = N^*(s) = 1, T > t \mid Z, b) f(b; \Sigma_b) db}{\int_b f(\Delta N^*(s) = N^*(s) = 1, T > t \mid Z, b) f(b; \Sigma_b) db}, \quad \text{for } t > s.
\end{aligned}$$



For illustration purposes, the predicted survival distribution for a female patient who has one CHD event at study year 5 and has the average baseline measures of age and total-cholesterol level, no hypertension medication, never smoking, and no diabetes is displayed in Figure 2, along with the 95% pointwise confidence intervals. The confidence intervals are obtained by applying the functional delta method and evaluating at the NPMLEs.

## 7 Concluding Remarks

We have presented joint transformation models for repeated measures and recurrent event times with an informative terminal event. We have provided an efficient EM algorithm to compute the maximum likelihood estimators of the model parameters. The nonparametric maximum likelihood estimators are shown to be consistent, asymptotically normal, and asymptotically efficient. The proposed approach has been applied to the ARIC data, and the resulting joint models can be used in predicting a patient's future survival rate and longitudinal measures given his/her past history, for example.

To obtain the variance estimates of  $(\hat{\theta}, \hat{\Lambda}_R(t), \hat{\Lambda}_T(t))$ , we have used the inverse of the observed information matrix evaluated at the NPMLEs. The validity of using the Fisher information has been justified in many other models including frailty models (Parner, 1998) and transformation models (Kosorok et al., 2004). The main reason is that the nonparametric parameters can be estimated at the same rate  $(\sqrt{n})$  as the parametric components, which implying that the nonparametric baseline functions can be treated equally as usual parameters. Even if this approach yields consistent variance estimators, in practice inverting such a large dimensional matrix may be intimidating with the large number of observations. This limitation can be overcome by using a profile likelihood variance estimation (Murphy and van der Vaart, 2000) rather than a full likelihood approach if the parameter of interest is only  $\theta$ . Particularly, let  $p\ell_n(\theta)$  be the profile log-likelihood function for  $\theta$ , expressed as  $p\ell_n(\theta) = \sup_{\{\Lambda_R, \Lambda_T\}} \sum_{i=1}^n \ell_i(\theta, \Lambda_R, \Lambda_T) = \sum_{i=1}^n \ell_i(\theta, \Lambda_{R\theta}^*, \Lambda_{T\theta}^*)$ , where  $\ell_i(\theta, \Lambda_R, \Lambda_T)$  is the ob-

served log-likelihood function for the  $i^{\text{th}}$  subject. Then the negative second-order numerical difference of  $p\ell_n(\theta)$  at  $\theta = \hat{\theta}$  can approximate the inverse of the asymptotic variance of  $\hat{\theta}$  (Zeng and Cai, 2005). That is,

$$-n^{-1}\{p\ell_n(\hat{\theta} + r_n c) - 2p\ell_n(\hat{\theta}) + p\ell_n(\hat{\theta} - r_n c)\}/r_n^2$$

can estimate  $c^T I_{\theta}^* c$  for any vector  $c$  such that  $\|c\|=1$  and any constant  $r_n = O(n^{-1/2})$ , where  $I_{\theta}^*$  is the efficient information for  $\theta$ . In this approach, the computation of  $\Lambda_{R\theta}^*$  and  $\Lambda_{T\theta}^*$  which maximize the observed log-likelihood function for a fixed  $\theta$  in the neighborhood of  $\hat{\theta}$  (i.e.,  $\theta = \hat{\theta} \pm r_n c$ ) can be conducted by utilizing the EM algorithm in Section 3.2, but holding  $\theta$  fixed all the time in both the E-step and M-step.

In lack of modeling-checking technics for frailty models or joint models, if the proportional hazards or odds models are misused, regression coefficients can be significantly biased as shown in our simulation studies (Web Table 1). We expect that the transformation models can be a useful tool in finding the model with better fit to data and less biased. The transformation parameter can be easily tested using the profile likelihood ratio test. Let  $p\tilde{\ell}_n(\gamma)$  be the profile log-likelihood function for  $\gamma$ , expressed as

$$p\tilde{\ell}_n(\gamma) = \sup_{\{\theta, \Lambda_R, \Lambda_T\}} \sum_{i=1}^n \ell_i(\theta(\gamma), \Lambda_R(t; \gamma), \Lambda_T(t; \gamma)) = \sum_{i=1}^n \ell_i(\hat{\theta}(\gamma), \hat{\Lambda}_R(t; \gamma), \hat{\Lambda}_T(t; \gamma)), \quad (7)$$

where  $\{\hat{\theta}(\gamma), \hat{\Lambda}_R(t; \gamma), \hat{\Lambda}_T(t; \gamma)\}$  are the NPMLEs for a fixed  $\gamma$ , treating  $G_R$  and  $G_T$  are deterministic functions. Maximizing (7) with respect to  $\gamma$  yields the profile likelihood estimator  $\tilde{\gamma}$ . Substituting this estimate into (7) yields the profile likelihood  $p\tilde{\ell}_n(\tilde{\gamma})$ . On the other hand, under the null hypothesis  $H_0 : \gamma = \gamma_0$  for a constant  $\gamma_0$ , the profile likelihood is  $p\tilde{\ell}_n(\gamma_0) = \sum_{i=1}^n \ell_i(\hat{\theta}(\gamma_0), \hat{\Lambda}_R(t; \gamma_0), \hat{\Lambda}_T(t; \gamma_0))$ . The profile likelihood ratio statistic is then constructed as

$$T_n = -2[p\tilde{\ell}_n(\gamma_0) - p\tilde{\ell}_n(\tilde{\gamma})].$$

The test statistic  $T_n$  can be approximated by a chi-square distribution with 1 degrees of freedom (Murphy and van der Vaart, 2000). For the boundary parameter  $\gamma_0 = 0$ , we expect that the test statistic follows asymptotically a mixture of two chi-square distributions with 0 and 1 degrees of freedom, respectively.

The proposed maximum likelihood estimation for the joint models requires distributional assumptions of random effects, while the estimating equation approach proposed by Sun, Sun, and Liu (2007) is free of these assumptions. As for future research, when modeling the terminal event may not be a major concern, it may be interesting to compare our method to their marginal approach in terms of computation complexity and efficiency gain in estimating regression parameters. Our model assumes that longitudinal measures are linearly related to all covariates considered. Where it is believed that the longitudinal measures are nonlinearly related to some predictors, we can increase the flexibility of our joint models by including some nonparametric functions of those predictors additively in the longitudinal components. Another promising extension of our joint model would be to the context of generalized linear mixed models (GLM) for analyses of discrete longitudinal outcomes. This is rather straightforward in that random effects in GLMs are commonly assumed to follow a mean-zero normal distribution as in our model. Therefore, the same inference procedures by EM algorithm would also be applicable in GLMs. The only difference in computation would be a lack of closed-form estimators in the M-step of the longitudinal process; instead, a numerical method such as iteratively re-weighted least squares may be involved. We can also extend our model to multiple types of recurrent and/or terminal events.

## 8 Supplementary Materials

Web Tables and Appendixes referenced in Sections 1, 3-5 and 7 are available under the separate report *Supplementary\_Materials.pdf*.

## References

- Chambless, L., Folsom, A., Sharrett, A., Sorlie, P., Couper, D., Szklo, M., and Nieto, F. (2003). *Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study*. *Journal of clinical epidemiology* **56**, 880–890.
- Dempster, A., Laird, N., and Rubin, D. (1977). *Maximum likelihood from incomplete data via the EM algorithm*. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Henderson, R., Diggle, P., and Dobson, A. (2000). *Joint modelling of longitudinal measurements and recurrent events*. *Biostatistics* **1**, 465–480.
- Hsieh, F., Tseng, Y., and Wang, J. (2006). *Joint modeling of survival and longitudinal data: Likelihood approach revisited*. *Biometrics* **62**, 1037–1043.
- Huang, C. and Wang, M. (2004). *Joint modeling and estimation for recurrent event processes and failure time data*. *Journal of the American Statistical Association* **99**, 1153–1165.
- Huang, X. and Liu, L. (2007). *A joint frailty model for survival and gap times between recurrent events*. *Biometrics* **63**, 389–397.
- Kosorok, M., Lee, B., and Fine, J. (2004). *Robust inference for univariate proportional hazards frailty regression models*. *The Annals of Statistics* **32**, 1448–1491.
- Liang, Y., Lu, W., and Ying, Z. (2009). *Joint modeling and analysis of longitudinal data with informative observation times*. *Biometrics* **65**, 377–384.
- Lin, D. and Ying, Z. (2001). *Semiparametric and nonparametric regression analysis of longitudinal data*. *Journal of the American Statistical Association* **96**, 103–126.
- Lin, H., Scharfstein, D., and Rosenheck, R. (2004). *Analysis of longitudinal data with irregular, outcome-dependent follow-up*. *Journal of the Royal Statistical Society, Series B* **66**, 791–813.

- Liu, L. and Huang, X. (2009). Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. Journal of the Royal Statistical Society, Series C* **58**, 65–81.
- Liu, L., Huang, X., and O’Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. Biometrics* **64**, 950–958.
- Liu, L., Wolfe, R., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. Biometrics* **60**, 747–756.
- Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society, Series B* **44**, 226–233.
- Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. Journal of the American Statistical Association* **95**, 449–485.
- Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. The Annals of Statistics* **26**, 183–214.
- Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V., and Soubeyran, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. Biostatistics* **8**, 708–721.
- Song, X. and Wang, C. (2008). Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. Biometrics* **64**, 557–566.
- Sun, J., Park, D., Sun, L., and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. Journal of the American Statistical Association* **100**, 882–889.

- Sun, J., Sun, L., and Liu, D. (2007). *Regression analysis of longitudinal data in the presence of informative observation and censoring times*. Journal of the American Statistical Association **102**, 1397–1406.
- Tsiatis, A. and Davidian, M. (2001). *A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error*. Biometrika **88**, 447–458.
- Tsiatis, A. and Davidian, M. (2004). *Joint modeling of longitudinal and time-to-event data: an overview*. Statistica Sinica **14**, 809–834.
- Vonesh, E., Greene, T., and Schluchter, M. (2006). *Shared parameter models for the joint analysis of longitudinal data and event times*. Statistics in Medicine **25**, 143–163.
- Wang, M., Qin, J., and Chiang, C. (2001). *Analyzing recurrent event data with informative censoring*. Journal of the American Statistical Association **96**, 1057–1065.
- Wattanakit, K., Folsom, A., Chambless, L., and Nieto, F. (2005). *Risk factors for cardiovascular event recurrence in the Atherosclerosis Risk in Communities (ARIC) study*. American Heart Journal **149**, 606–612.
- Wulfsohn, M. and Tsiatis, A. (1997). *A joint model for survival and longitudinal data measured with error*. Biometrics **53**, 330–339.
- Xu, J. and Zeger, S. (2001). *Joint analysis of longitudinal data comprising repeated measures and times to events*. Applied Statistics **50**, 375–387.
- Ye, Y., Kalbfleisch, J., and Schaubel, D. (2007). *Semiparametric analysis of correlated recurrent and terminal events*. Biometrics **63**, 78–87.
- Zeng, D. and Cai, J. (2005). *Simultaneous modelling of survival and longitudinal data with an application to repeated quality of life measures*. Lifetime Data Analysis **11**, 151–174.
- Zeng, D. and Lin, D. Y. (2009). *Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events*. Biometrics **65**, 746–752.

Table 1: Simulation results for  $G_R(x) = G_T(x) = x$ . Bias and SE are the bias and the standard deviation estimates, SEE is the average of the standard error estimator, and CP is the coverage probability of 95% confidence intervals.  $\tau$  denotes the study duration.

<i>Parameter</i>	<i>True</i>	<i>N = 200</i>				<i>N = 400</i>			
		<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>	<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>
$\phi = (0.5, 0.2)$									
$\beta_1$	0.7	-.011	.083	.081	.934	-.010	.059	.058	.944
	1.0	-.012	.123	.118	.935	-.012	.086	.083	.946
	0.5	-.008	.102	.108	.947	-.007	.072	.072	.953
$\sigma_e^2$	1.0	.002	.041	.041	.958	-.000	.031	.029	.940
$\beta_2$	1.0	-.012	.138	.138	.951	-.021	.094	.097	.955
	0.5	-.007	.118	.119	.952	-.009	.084	.083	.949
$\Lambda_R(\tau/4)$	1.0	-.014	.111	.112	.948	-.012	.077	.079	.955
$\Lambda_R(\tau/2)$	2.0	-.033	.213	.215	.942	-.026	.149	.152	.946
$\beta_3$	1.0	.019	.199	.204	.962	-.003	.137	.141	.957
	0.5	.012	.175	.171	.954	-.000	.118	.118	.954
$\phi$	0.5	.028	.214	.211	.960	.005	.147	.143	.944
	0.2	-.016	.219	.219	.960	-.010	.153	.149	.945
$\Lambda_T(\tau/4)$	0.1	-.003	.022	.022	.962	-.002	.016	.016	.953
$\Lambda_T(\tau/2)$	0.4	-.009	.069	.069	.953	-.003	.048	.049	.960
$\sigma_1^2$	0.5	-.013	.068	.067	.951	-.006	.049	.048	.951
$\sigma_2^2$	0.5	.013	.095	.092	.946	.013	.063	.064	.947
$\rho$	0.5	-.006	.091	.095	.955	-.005	.066	.066	.961
$\phi = (0, 0.2)$									
$\beta_1$	0.7	-.009	.082	.082	.939	-.009	.057	.058	.950
	1.0	-.014	.116	.118	.947	-.016	.084	.083	.936
	0.5	-.001	.101	.103	.947	-.005	.072	.072	.952
$\sigma_e^2$	1.0	.001	.042	.041	.952	-.000	.029	.029	.949
$\beta_2$	1.0	-.016	.139	.139	.940	-.017	.099	.097	.932
	0.5	-.004	.121	.118	.947	-.008	.082	.083	.949
$\Lambda_R(\tau/4)$	1.0	-.012	.108	.113	.955	-.015	.082	.079	.938
$\Lambda_R(\tau/2)$	2.0	-.028	.209	.216	.949	-.031	.159	.152	.942
$\beta_3$	1.0	.015	.196	.191	.950	.003	.135	.133	.947
	0.5	.009	.163	.160	.946	.003	.107	.110	.960
$\phi$	0.0	.001	.204	.200	.954	.007	.139	.135	.953
	0.2	-.008	.234	.221	.944	-.018	.157	.148	.947
$\Lambda_T(\tau/4)$	0.1	-.002	.022	.022	.958	-.001	.015	.015	.941
$\Lambda_T(\tau/2)$	0.4	-.001	.068	.067	.960	-.001	.047	.047	.944
$\sigma_1^2$	0.5	-.007	.069	.068	.954	-.004	.046	.047	.960
$\sigma_2^2$	0.5	.010	.092	.093	.946	.010	.063	.064	.948
$\rho$	0.5	-.002	.096	.094	.947	-.003	.066	.065	.941

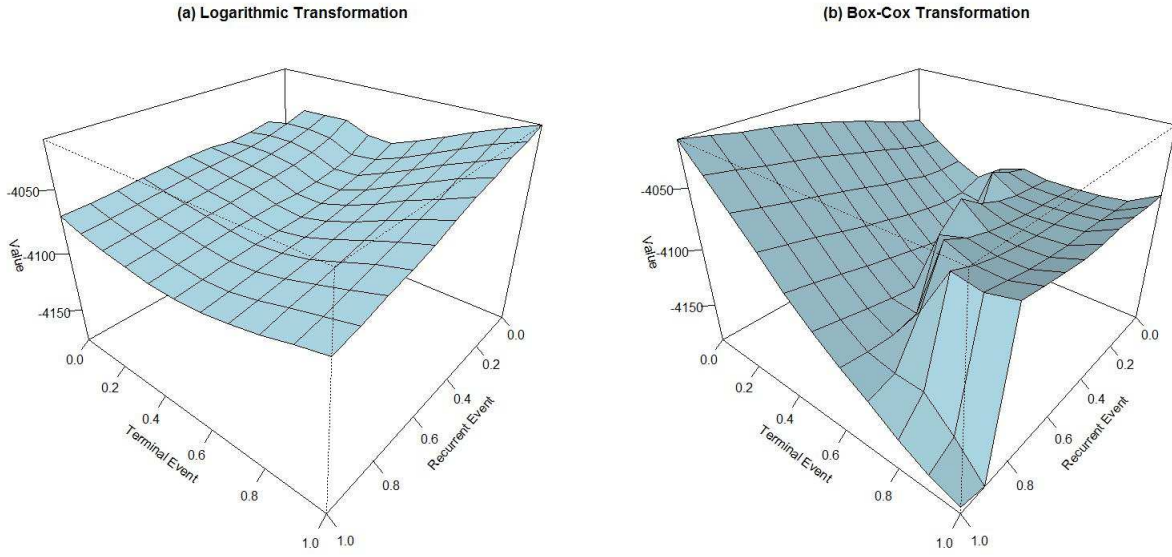


Figure 1: Log-likelihood surfaces under the (a) logarithmic transformations and the (b) Box-Cox transformations for the ARIC study. The x-axis and y-axis correspond to the transformation parameters ( $\gamma_L$  in (a) and  $\gamma_{BC}$  in (b)) for recurrent event and terminal event, respectively.



Table 2: Simulation results for  $G_R(x) = x$  and  $G_T(x) = \log(1+x)$ . Bias and SE are the bias and the standard deviation estimates, SEE is the average of the standard error estimator, and CP is the coverage probability of 95% confidence intervals.  $\tau$  denotes the study duration.

<i>Parameter</i>	<i>True</i>	<i>N = 200</i>				<i>N = 400</i>			
		<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>	<i>Bias</i>	<i>SE</i>	<i>SEE</i>	<i>CP</i>
$\phi = (0.5, 0.2)$									
$\beta_1$	0.7	-.009	.084	.082	.935	-.008	.058	.058	.949
	1.0	-.003	.120	.119	.956	-.006	.085	.084	.939
	0.5	-.007	.104	.103	.943	-.002	.074	.073	.940
$\sigma_e^2$	1.0	.002	.041	.041	.955	.001	.030	.029	.946
	$\beta_2$	1.0	-.003	.157	.156	.957	-.006	.107	.108
$\Lambda_R(\tau/4)$	0.6	-.009	.135	.132	.950	-.001	.091	.092	.949
	$\Lambda_R(\tau/2)$	1.2	-.012	.156	.154	.953	-.016	.109	.107
$\beta_3$	1.0	.010	.282	.284	.952	.001	.198	.198	.954
	0.5	.004	.228	.241	.966	.005	.168	.169	.951
$\phi$	0.5	-.009	.349	.367	.969	.007	.245	.246	.957
	0.2	-.012	.436	.444	.969	-.011	.301	.299	.949
$\Lambda_T(\tau/4)$	0.1	.004	.036	.037	.965	-.002	.026	.026	.966
$\Lambda_T(\tau/2)$	0.4	-.000	.136	.131	.953	.000	.093	.092	.948
$\sigma_1^2$	0.5	-.009	.070	.069	.956	-.004	.049	.049	.949
$\sigma_2^2$	0.5	.015	.145	.112	.945	.009	.078	.078	.941
$\rho$	0.5	-.002	.108	.108	.961	-.004	.073	.075	.953
$\phi = (0, 0.2)$									
$\beta_1$	0.7	-.008	.082	.082	.950	-.008	.057	.058	.954
	1.0	-.001	.120	.119	.953	-.005	.084	.084	.948
	0.5	.001	.106	.103	.945	-.006	.071	.072	.958
$\sigma_e^2$	1.0	.001	.041	.041	.947	.001	.028	.029	.953
	$\beta_2$	1.0	-.011	.151	.152	.956	-.005	.111	.108
$\Lambda_R(\tau/4)$	0.5	.005	.134	.130	.944	-.002	.092	.092	.944
	0.15	-.001	.081	.079	.944	-.007	.056	.055	.943
$\Lambda_R(\tau/2)$	0.6	-.005	.151	.151	.947	-.019	.106	.106	.949
$\beta_3$	1.0	.013	.276	.276	.951	.023	.193	.192	.947
	0.5	.009	.234	.235	.950	.003	.163	.164	.947
$\phi$	0.0	.014	.367	.357	.963	.008	.246	.241	.956
	0.2	-.029	.459	.440	.956	-.045	.297	.293	.942
$\Lambda_T(\tau/4)$	0.1	-.004	.036	.036	.955	-.004	.025	.025	.949
$\Lambda_T(\tau/2)$	0.4	.004	.126	.128	.959	-.010	.085	.087	.952
$\sigma_1^2$	0.5	-.006	.070	.068	.951	-.005	.049	.048	.955
$\sigma_2^2$	0.5	.007	.112	.110	.943	.011	.077	.077	.946
$\rho$	0.5	-.007	.108	.107	.959	-.004	.073	.074	.959

Table 3: Analysis results for the ARIC study. The Fisher transformation is used for testing  $\rho$ , while 50:50 mixture of  $\chi^2$  distributions is used for testing variances.

<i>Effect</i>	<i>Estimate</i>	<i>Std.Err</i>	<i>p-value</i>	<i>Estimate</i>	<i>Std.Err</i>	<i>p-value</i>
	<u>Full Model</u>			<u>Reduced Model</u>		
	<i>Longitudinal measures of SBP</i>					
Intercept	1.026	.062	< .001	1.028	.061	< .001
Age	0.338	.044	< .001	0.336	.044	< .001
Male	-0.043	.053	.413	-0.039	.053	.461
Total-cholesterol ( <i>mg/dL</i> )	-0.007	.023	.766	-0.007	.023	.772
Hypertension medication	-0.546	.054	< .001	-0.550	.054	< .001
Ever smoker	-0.046	.054	.394	-0.056	.052	.288
Diabetes	0.088	.070	.207	0.089	.070	.200
Visit Year	0.014	.004	.001	0.013	.004	.001
$\sigma_e^2$	0.500	.016	< .001	0.500	.016	< .001
	<i>Recurrent CHD event</i>					
Age	0.837	.264	.002	0.506	.150	.001
Male	0.871	.220	< .001	0.795	.187	< .001
Total-cholesterol ( <i>mg/dL</i> )	0.195	.109	.073	0.184	.071	.009
Hypertension medication	0.191	.234	.414	0.108	.181	.552
Ever smoker	0.862	.309	.005	0.352	.182	.053
Diabetes	1.843	.253	< .001	1.394	.179	< .001
	<i>Terminal event</i>					
Age	2.187	.405	< .001			
Male	0.611	.364	.093			
Total-cholesterol ( <i>mg/dL</i> )	0.043	.231	.854			
Hypertension medication	0.336	.393	.392			
Ever smoker	2.336	.653	< .001			
Diabetes	2.671	.374	< .001			
$\phi_1$	-0.460	.471	.329			
$\phi_2$	1.910	.317	< .001			
	<i>Variance components for random effect</i>					
$\sigma_1^2$	0.331	.024	< .001	0.332	.024	< .001
$\sigma_2^2$	3.100	.883	< .001	0.809	.223	< .001
$\rho$	0.241	.086	.007	0.348	.107	.003

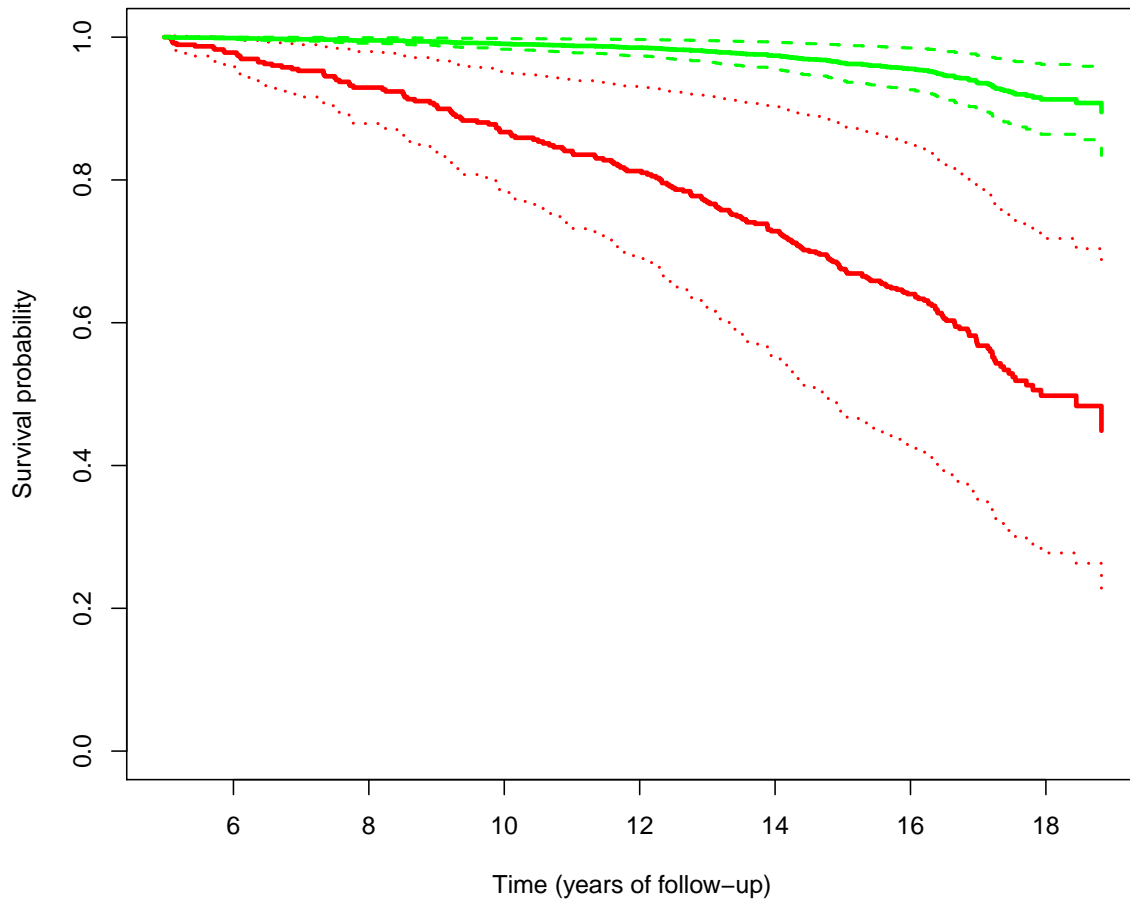


Figure 2: Predicted survival probabilities for a subject who had no CHD event (upper 3 lines) and for a subject who had 1 CHD event (lower 3 lines) at the 5<sup>th</sup> year of study. The solid curves are point estimates, and the dotted curves are the 95% point-wise confidence intervals.