A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects

Xin Huang,
Amgen Inc., 1120 Veterans Boulevard, Mail Stop ASF3-3, South San Francisco, CA 94080, USA

Gang Li,
Department of Biostatistics, School of Public Health, University of California at Los Angeles, Los Angeles, CA 90095, USA

Robert M. Elashoff, and
Department of Biomathematics, University of California at Los Angeles, Los Angeles, CA 90095, USA

Jianxin Pan
School of Mathematics, The University of Manchester, Manchester, UK

Xin Huang: xin@amgen.com; Gang Li: vli@ucla.edu; Robert M. Elashoff: relashof@biomath.medsch.ucla.edu; Jianxin Pan: jianxin.pan@manchester.ac.uk

Abstract

This article studies a general joint model for longitudinal measurements and competing risks survival data. The model consists of a linear mixed effects sub-model for the longitudinal outcome, a proportional cause-specific hazards frailty sub-model for the competing risks survival data, and a regression sub-model for the variance–covariance matrix of the multivariate latent random effects based on a modified Cholesky decomposition. The model provides a useful approach to adjust for non-ignorable missing data due to dropout for the longitudinal outcome, enables analysis of the survival outcome with informative censoring and intermittently measured time-dependent covariates, as well as joint analysis of the longitudinal and survival outcomes. Unlike previously studied joint models, our model allows for heterogeneous random covariance matrices. It also offers a framework to assess the homogeneous covariance assumption of existing joint models. A Bayesian MCMC procedure is developed for parameter estimation and inference. Its performances and frequentist properties are investigated using simulations. A real data example is used to illustrate the usefulness of the approach.

Keywords

Cause-specific hazard; Bayesian analysis; Cholesky decomposition; Mixed effects model; MCMC; Modeling covariance matrices

---

1 Introduction

Joint modeling of longitudinal and survival data has received a great deal of attention in the past decades in many studies in which both a longitudinal outcome during follow-up and the occurrence of some key events are recorded. In the statistical literature, joint models have been proposed to adjust inferences on longitudinal measurements in the presence of non-ignorable missing values due to dropout (Schluchter 1992; DeGruttola and Tu 1994; Little 1995; Hogan and Laird 1997; Henderson et al. 2000; Elashoff et al. 2007, 2008; Hu et al. 2009); to solve difficulties in Cox proportional hazards model arising from time-dependent covariates which are possibly missing at some event times or subject to substantial measurement error (Faucett and Thomas 1996; Wulfsohn and Tsiatis 1997; Faucett et al. 1998; Wang and Taylor 2001; Xu and Zeger 2001; Song et al. 2002; Brown and Ibrahim 2003; Tseng et al. 2005; Ye et al. 2008); and to assess covariates effects on both endpoints simultaneously (Henderson et al. 2000; Zeng and Cai 2005; Elashoff et al. 2007, 2008; Liu et al. 2008).

All the aforementioned joint models assume that the random effects covariance matrix is the same for all subjects. However, examining whether this matrix is the same for all subjects (homogeneous) or whether it differs depending on subject-specific characteristics (heterogeneous) is often neglected in the modeling. Furthermore, ignoring the heterogeneity can result in biased estimates of the fixed and random effects for the longitudinal outcome (Heagerty and Kurland 2001; Daniels and Zhao 2003). Accounting for heterogeneity in covariance matrices has been discussed by serval authors in the field of generalized linear regression models (Chiu et al. 1996), non-linear mixed models (Davidian and Giltinan 1995), and linear mixed models (Pourahmadi and Daniels 2002; Lin et al. 1997; Zhang and Weiss 2000; Daniels and Zhao 2003). Nonetheless, no work has been done on modeling the entire random effects covariance matrix for the joint models.

In this paper, we propose an approach that allows heterogeneous random effects covariance matrix within the framework of joint analysis of longitudinal measurements and competing risks failure time data. In our joint model, a linear mixed effects sub-model is used to characterize the longitudinal measurements, a cause-specific hazards frailty sub-model for the competing risks survival data (Prentice and Breslow 1978), together with a regression sub-model for the joint multivariate random effects covariance matrix which links the first two sub-models. Specifically, we first use a modified Cholesky decomposition to decompose the covariance matrix into a lower-triangular matrix and a diagonal matrix, and then model these matrix entries using regression models (Pourahmadi 1999; Daniels and Zhao 2003). By jointly modeling the random effects covariance matrices, our model is distinct from previously studied joint models (e.g. Hu et al. 2009) that consist of only two sub-models for the longitudinal and survival outcomes respectively. Our model has several advantages. First of all, unlike existing joint models that assume a homogeneous covariance matrix, our model allows for heterogeneous covariance matrices. Secondly, as discussed in the remark of Sect. 2, our model includes homogeneous models as special cases. Thirdly, the covariance model enables dimension reduction. With different choices of regression covariates, it provides a flexible means to model the heterogeneity and reduce the number of variance–covariance parameters to be estimated. Forthly, the resulting estimated covariance matrices of the multivariate random effects are guaranteed to be positive definite, which is not always the case for other existing joint models. Finally, our model provides a useful framework to assess the homogeneous covariance assumption of existing joint models which is otherwise untestable. Likelihood-based inference for our model is rather challenging with high-dimensional random effects. We develop a Bayesian MCMC algorithm to fit the joint model. Gibbs sampling technique, together with Metropolis-Hastings sampling and adaptive rejection sampling (ARS) methods, is used to draw random samples from the full

conditional distributions of parameters. With the Bayesian approach, prior information can be incorporated in a natural way. If no prior information is available, we recommend noninformative priors for parameters to allow data to dominate the determination of the posterior distributions.

This paper is organized as follows: in Sect. 2 we describe the joint model formulation. In Sect. 3 we develop the Bayesian estimation and inference methods. In Sect. 4, a real data application is illustrated using the data from the Scleroderma clinical trial (Tashkin et al. 2006). In Sect. 5, the performance of our method is examined by simulation studies. Some concluding remarks are provided in Sect. 6. Details of the MCMC algorithm are deferred to the Appendix.

2 Joint model

Our joint model consists of three components: a linear mixed effect model for the longitudinal measurements, a cause-specific hazards model for the competing risks survival data, and a regression model for the variance–covariance matrix of the multivariate latent random effects based on a modified Cholesky decomposition.

2.1 Longitudinal sub-model

Suppose there are m subjects in the study. For the ith subject at time t, the longitudinal outcome $Y_i(t)$ follows a linear mixed effects model:

$$Y_i(t)=X_i^{(1)}(t)^T\beta+Z_i(t)^T U_i+\varepsilon_i(t)$$

(1)

where $X_i^{(1)}(t)$ and $Z_i(t)$ are vectors of covariates associated with the fixed effects β ($p \times 1$) and the random effects U_i ($q \times 1$) respectively. We assume that the measurement error $\varepsilon_i(t)$, which is distributed as $N(0, \sigma^2)$, is independent of U_i and $\varepsilon_i(t_1) \perp \varepsilon_i(t_2)$ for any $t_1 \neq t_2$.

2.2 Cause-specific hazards sub-model

During follow-up, each subject may experience one of g distinct competing causes of failure or may be right censored. Let $C_i = (T_i, D_i)$ be the competing risks survival data on subject i, where T_i is the failure or censoring time, and D_i assumes a value from 0, 1, …, g, with $D_i = 0$ indicating a noninformative censored event and $D_i = k$ indicating the kth failure type, $k = 1$, …, g. Dependent (or informative) censoring is treated as one of the g types of failures. The cause-specific hazards sub-model for the competing risks survival data is specified as follows:

$$\lambda_k(t;X_i^{(2)}(t),\upsilon_i,\gamma_k,v_k)=\lim_{h\to 0}\frac{P\left[t\le T_i<t+h,D_i=k|T_i\ge t,X_i^{(2)}(t),\upsilon_i,\gamma_k,v_k\right]}{h}$$
$$=\lambda_{0k}(t)\exp\left\{X_i^{(2)}(t)^T\gamma_k+v_k\upsilon_i\right\}.$$

(2)

The function $\lambda_k(t;X_i^{(2)},\upsilon_i,v_k,\gamma_k)$ is the instantaneous failure rate from cause k at time t given the vector of covariates $X_i^{(2)}(t)$ and the latent unknown factor υ_i, in the presence of all other failure types. The regression coefficient v_k represents the effect of the latent variable υ_i with v_1 set to 1 to ensure identifiability. The parameter γ_k represents the effects of the observed covariates $X_i^{(2)}(t)$ on cause k. We further assume that the kth baseline hazard is a step function, $\lambda_{0k}(t)=\lambda_{0k}^{(s)}$, for $t_k^{(s-1)}<t \le t_k^{(s)}$, where $0<t_k^{(1)}<\cdots<t_k^{(S^k)}<\infty$ is a partition of $(0, \infty)$ and S^k indicates the number of steps for the kth baseline hazard.

2.3 Variance–covariance regression sub-model

We model the association between the longitudinal and survival sub-models by assuming that U_i and v_i jointly have a multivariate normal distribution:

$$W_i = \begin{pmatrix} U_i \\ v_i \end{pmatrix} \sim N_{q+1}\left(\begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \sum_i = \begin{pmatrix} \Sigma_{U_i} & \Sigma_{Uv_i} \\ \Sigma_{Uv_i}^T & \sigma_{v_i}^2 \end{pmatrix} \right)$$

(3)

Similar to Pourahmadi (1999), we model the covariance matrices Σ_i through a modified Cholesky decomposition $M_i \sum_i M_i^T = H_i$, where H_i is a diagonal matrix with positive entries and M_i is the lower triangular matrix with $-\varphi_{i, jl}$ as its $(j, l)^{\text{th}}$ entry. This decomposition has a clear statistical interpretation: the below-diagonal entries of M_i are the negatives of generalized autoregressive parameters (GARP), $\varphi_{i, jl}$, in the autoregressive model

$$W_{ij} = \sum_{l=1}^{j-1} \varphi_{i, jl} W_{jl} + e_{ij}, \quad j = 1, \ldots, q+1.$$

(4)

The diagonal entries of H_i are the innovation variances (IV) $h_{ij}^2 = \text{var}(e_{ij})$ and we have $\text{cov}(e_{ij}, e_{jk}) = 0$ if $j \neq k$ ($1 \leq j, k \leq q + 1$ and $i = 1, \ldots, m$). The GARPs and the logarithms of the IVs are modeled with linear and log link functions:

$$\begin{cases} \varphi_{i, jl} = a_{i, jl}^T \eta_1 & \text{for } i = 1, \ldots, m \\ \log h_{ij}^2 = b_{ij}^T \eta_2 & j = 1, \ldots, q+1, \quad l = 1, \ldots, j-1 \end{cases}$$

(5)

where $a_{i, jl}$ and b_{ij} are covariates, and η_1 and η_2 are low-dimensional parameter vectors. For example, $a_{i, jl}$ and b_{ij} may contain group indicators, implying that the random effects covariances are heterogeneous. The homogeneous random effects assumption in existing joint models becomes a testable assumption within our model framework. Furthermore, the resulting estimated covariance matrix is guaranteed to be positive definite. The latent association between the longitudinal measurements and survival outcomes can be assessed by testing the hypothesis $\Sigma_{Uv_i} = \mathbf{0}$. Finally, we assume that conditional on all the covariates and random effects, the longitudinal measurements and the competing risks survival data are independent.

Remark: Choice of design vectors for GARP/IV parameters—As we mentioned earlier, the choice of covariate vectors $a_{i, jl}$ and $b_{i, j}$ are flexible. For example, a 3-dimensional random effects variance–covariance matrix has six parameters. We can model the homogeneous unstructured covariance matrix by setting $a_{i, jl} = b_{ij} = 1$ for all $j = 1, \ldots, 3$, $l = 1, \ldots, j - 1$. If we assume the design vectors contain subject-dependent covariate, say, a group indicator (G), the unstructured heterogeneous covariance matrix can be modeled with $a_{i, jl} = b_{ij} = (1, G_i)$ for all $j = 1, \ldots, 3, l = 1, \ldots, j - 1$; that is,

$$\begin{cases} a_i = (a_{i, jl}) = (1, G_i, 1, G_i, 1, G_i)^T; \eta_1 = (\eta_{11}^{Int}, \eta_{11}^G, \eta_{12}^{Int}, \eta_{12}^G, \eta_{13}^{Int}, \eta_{13}^G)^T \\ b_i = (b_{ij}) = (1, G_i, 1, G_i, 1, G_i)^T; \eta_2 = (\eta_{21}^{Int}, \eta_{21}^G, \eta_{22}^{Int}, \eta_{22}^G, \eta_{23}^{Int}, \eta_{23}^G)^T. \end{cases}$$

(6)

When there are high-dimensional random effects with limited data, one can impose a restricted covariance structure and assume some of the GARP are the same to reduce the number of parameters.

3 Estimation and inference

The standard maximum likelihood method involves integrating out latent variables from the log-likelihood function which is difficult when dealing with high-dimensional variables. We develop a Bayesian estimation procedure and a Markov chain Monte Carlo (MCMC) method for estimation and inference.

3.1 Likelihood

Suppose the longitudinal outcome $Y_i(t)$ is observed at time points t_{ij} for $j = 1, \ldots, n_i$, and denote $Y_i = (Y_{i1}, \ldots, Y_{in_i})$. Let $\Omega = \{\beta, \sigma^2, \gamma, v, \lambda_0, \eta_1, \eta_2\}$, where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_g)$, $v = (v_2, \ldots, v_g)$ and $\lambda_0 = (\lambda_{01}^{(1)}, \lambda_{01}^{(2)}, \ldots, \lambda_{0g}^{(S^g)})$. It is convenient to work directly with the joint distribution of the observed data (Y, C) and the unobservable random effects W, conditional on Ω, which facilitates the MCMC implementation. The conditional joint density of (Y, C) and W is:

$$
\begin{aligned}
p(Y, C, W|\Omega) &= \prod_{i=1}^{m} p(Y_i|W_i, \Omega) p(C_i|W_i, \Omega) p(W_i|\Omega) \\
&\propto \prod_{i=1}^{m} (2\pi\sigma^2)^{-\frac{n_i}{2}} \exp\left\{ -\frac{(Y_i - X_i^{(1)}\beta - Z_i U_i)^T (Y_i - X_i^{(1)}\beta - Z_i U_i)}{2\sigma^2} \right\} \\
&\quad \times \prod_{k=1}^{g} \left((\lambda_k(T_i))^{I(D_i=k)} \exp\{-H_k(T_i)\} \right) \\
&\quad \times \exp\left\{ -\frac{1}{2} \sum_{j=1}^{q+1} \left[b_{ij}^T \eta_2 + \left(W_{ij} - \sum_{l=1}^{j-1} a_{ijl}^T \eta_1 W_{il} \right)^2 \times \exp\left(-b_{ij}^T \eta_2 \right) \right] \right\}
\end{aligned}
$$

(7)

where

$$
\lambda_k(T_i) = \lambda_{0k}(T_i) \exp\left\{ X_i^{(2)}(T_i)\gamma_k + v_k v_i \right\}
$$

(8)

and

$$
H_k(T_i) = \int_0^{T_i} \lambda_{ok}(t) \exp\left(X_i^{(2)}(t)\gamma_k + v_k v_i \right) dt.
$$

(9)

Under the piecewise constant hazard assumption,

$$
H_k(T_i) = \exp(v_k v_i) \sum_{s=1}^{S^k} I\left(T_i > t_k^{(s-1)} \right) \lambda_{0k}^{(s)} \int_{t_k^{(s-1)}}^{min(T_i, t_k^{(s)})} \exp\left(X_i^{(2)}(t)\gamma_k \right) dt.
$$

(10)

3.2 Priors and MCMC sampling procedure

We assume independent priors for Ω. We use Normal priors for the parameters β, γ, v, η_1 and η_2, leading to conjugate posteriors for β and some components of the η_1. We use an

inverse Gamma prior for the measurement error variance σ^2 and a gamma prior for each step of the kth baseline hazard function λ_{0k} by which conjugate posterior distributions are easy to obtain.

Markov Chain Monte Carlo (MCMC) methods are used for posterior sampling. It involves sampling directly from the full conditional distribution, Metropolis–Hastings (MH) sampling (Hastings 1970; Chib and Greenberg 1995) and adaptive rejection sampling (ARS) (Gilks and Wild 1992). Since the full conditional distributions of the parameters β, σ^2, and $\lambda_{0k}^{(s)}$, ($s = 1, \ldots, S^g$, $k = 1, \ldots, g$) are standard distributions, drawing random variates from their full conditional distributions is straightforward. For the rest of the parameters and random effects (U_i, v_i), we either use a Metropolis–Hastings step with the normal approximation to the full conditional distribution as the candidate distribution or apply the ARS technique. The technical details on the sampling distributions are given in the Appendix.

The initial values of the parameters for sampling are obtained by modeling the longitudinal data and survival data separately by a linear mixed model and a Cox proportional hazards model. The initial value for $\lambda_{0k}^{(s)}$ ($s = 1, \ldots, S^k$, $k = 1, \ldots, g$) can be obtained by drawing a random variate from the gamma full conditional distribution described in the Appendix. We estimate the parameters by their posterior medians. Approximate 95% probability intervals are based on 2.5th percentile and 97.5th percentile. Standard errors are obtained from the standard deviations of the posterior samples. The convergence of the Gibbs sampler is monitored by examining time series plots of the parameters over iteration and the Gelman and Rubin (1992) approach of using multiple chains.

4 Application

We analyze the data from a scleroderma lung study (SLS) (Tashkin et al. 2006) with our proposed joint model. The study enrolled 158 patients with scleroderma-related interstitial lung disease, randomized to receive either CYC (79 patients) or identical appearing placebo (79 patients) for 12 months. An additional year of follow-up was performed to determine if CYC effects persisted after treatment. The primary outcome is forced vital capacity (FVC, % predicted), measured at 3-month intervals from the baseline. We are interested in evaluating if oral cyclophosphamide (CYC) can either improve the %FVC level of a patient or decrease the risk of treatment failure or death.

Since the full dose of CYC is not reached until month 6, our analysis is based on 6–24 months %FVC scores which includes 140 subjects. We observe 14 treatment failures or deaths, 32 informative and 5 noninformative dropouts. A dropout is non-informative if there is no evidence showing that the dropout is related to the disease or the treatment, and informative otherwise. Since the informative dropout is related to the patient's disease condition, it not only causes non-ignorable missing data in %FVC, but also is an informatively censored event for treatment failure or death.

We consider two baseline factors in our joint model when assessing the CYC treatment effects: baseline %FVC (FVC_0), and lung fibrosis (FIB_0). It is suggested by clinicians that the beneficial effects of CYC on pulmonary function continue to increase after stopping treatment at 12 months and eventually begin to wane after 18 months. Therefore, we fit the following linear spline mixed effects model with change point at month 18 for longitudinal measurements %FVC:

$$\%\text{FVC}_{ij}=\beta_0+\beta_1\text{FVC}_{0i}+\beta_2\text{FIB}_{0i}+\beta_3\text{CYC}_i+\beta_4\text{Time}_{ij}+\beta_5(\text{Time}_{ij}-18)_+$$
$$+\beta_6\text{FVC}_{0i}\times\text{CYC}_i+\beta_7\text{FIB}_{0i}\times\text{CYC}_i+\beta_8\text{Time}_{ij}\times\text{CYC}_i$$
$$+\beta_9(\text{Time}_{ij}-18)_+\times\text{CYC}_i+Z_{ij}U_i+\varepsilon_{ij}$$

(11)

where U_i is the subject-specific random effect and the ε_{ij} is the mutually independent measurement error.

We consider multiple choices for random effects covariates Z_i and select the model based on the Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002). The DIC has the advantage of being easy to compute using output from a Gibbs sampler and has a similar form as the Akaika Information Criterion (AIC): a goodness-of-fit term measured by deviance evaluated at the posterior mean of the parameters, and a penalty term defined by twice of the effective number of parameters. The effective number of parameters is computed as the mean deviance minus the deviance evaluated at the posterior mean. That is,

$$\text{DIC}=\text{dev}(\overline{\Omega})+2p_D$$

(12)

where $\overline{\Omega}$ is the posterior mean of parameter Ω, $p_D=\overline{\text{dev}}-\text{dev}(\overline{\Omega})$ and $\overline{\text{dev}}$ is the posterior mean of the deviance (the average of the deviances calculated using the estimated parameters at each step of the MCMC sampler). Based on the form of the DIC, it is obvious that the smaller the DIC value, the better the model proposed. We note that there are several versions of DIC for missing data models (Celeux et al. 2006; Chen 2006). Here we use the DIC constructed from the conditional distribution while treating both Ω and W as parameters because it is easy to compute. We conduct a small simulation to evaluate the DIC which selects 147 times out of 200 datasets and the effective dimension is always positive.

A cause-specific competing risks sub-model is applied to model disease-related dropout (risk 1) and treatment failure or death (risk 2):

$$\lambda_1(t)=\lambda_{01}(t)\exp(\gamma_{11}\text{FVC}_{0i}+\gamma_{12}\text{FIB}_{0i}+\gamma_{13}\text{CYC}_i+\gamma_{14}\text{FVC}_{0i}\times\text{CYC}_i+\gamma_{15}\text{FIB}_{0i}\times\text{CYC}_i+v_i)$$

(13)

and

$$\lambda_2(t)=\lambda_{02}(t)\exp(\gamma_{21}\text{FVC}_{0i}+\gamma_{22}\text{FIB}_{0i}+\gamma_{23}\text{CYC}_i+\gamma_{24}\text{FVC}_{0i}\times\text{CYC}_i+\gamma_{25}\text{FIB}_{0i}\times\text{CYC}_i+v_2v_i).$$

(14)

The latent variables from both sub-models are assumed to have a multivariate normal distribution with mean zero and variance–covariance matrices

$$\sum_i=\left(\begin{array}{cc}\Sigma_{U_i} & \Sigma_{Uv_i} \\ \Sigma_{Uv_i}^T & \sigma_{v_i}^2\end{array}\right).$$

We test the homogeneous random effects covariance matrix assumption by considering subject-dependent covariates for a_{ijl} and b_{ij}. Specifically, we choose $a_{ijl}=b_{ij}=(1,\text{CYC}_i)$, which allows heterogeneous covariance matrices for different treatment groups, and test the null hypothesis by examining if the 95% credible interval of CYC effects contains zero for all the GARP and IV parameters.

A 3-step baseline hazard function, with the time points defining the steps being equally split percentiles of the observed event times, is utilized for the informatively censored events and the event of treatment failure or death. Sensitivity analyses with 4- and 5-step baseline hazard functions are conducted and show no significant difference. We apply independent noninformative prior distributions for all the parameters which all assumed to have relatively large variances. The corresponding priors for the parameters are $\beta_0 \sim N(70, 10^3)$ and $\beta_l \sim N(0, 10^3)$ for $l = 1, \ldots, 9$; $\sigma^2 \sim IG(10^{-3}, 10^{-3})$; $\gamma_{kr} \sim N(0, 10^3)$ for $k = 1, 2$ and $r = 1, \ldots, 5$; $\lambda_{0k}^{(s)} \sim \Gamma(10^{-3}, 10^{-3})$ for $s = 1, \ldots, S^k$ and $S^1 = S^2 = 3$; $v_2 \sim N(0, 10^5)$; and each element of η_1 and $\eta_2 \sim N(0, 10^5)$.

Table 1 summarizes the covariance matrix parameters of different models, each was based on 30,000 iterations of MCMC sampling chains following a 15,000-iteration "burn-in" period. Since we include baseline %FVC as a fixed effect covariate, we do not consider random intercept to avoid possible confounding effects. We consider a one-random-slope (before 18 months) model, a structured two-random-slope model assuming the entries of last row in matrix M_i from the decomposition are the same, and an unstructured two-random-slope model. The structured random effects covariance matrix model might be useful when dealing with high-dimensional random effects model but with limited data. For the last element of the innovation variance parameter, we do not include the CYC effects due to the convergence issue. It is clear that none of the 95% credible intervals for CYC exclude zero. Therefore, we don't have sufficient evidence to reject the homogeneous random effects covariances assumption. All the effective numbers of parameters (p_D) are positive which is not an indication of possibly poor fit between the models and the data (Spiegelhalter et al. 2002). The conditional DIC we use tends to produce increasing p_Ds for increasing model complexity as suggested by Celeux et al. (2006). The two-random-slope model with unstructured covariance matrix has the smallest DIC, indicating that it might provide the best fit for the SLS data. Combining the earlier results, we chose the homogenous two-random-slope model with unstructured covariance matrix as our final model and its covariance parameters and DIC values are listed in the last column of Table 1.

The results of the selected two-random-slope model are summarized in Table 2. For comparison purposes, we perform separate analysis of the two endpoints, which is done by fitting a linear mixed model with two random slopes (11) for %FVC and a cause-specific hazards frailty model for the competing risks failure time data (13), (14), respectively. The two methods produce similar point estimates and credible intervals for baseline %FVC, lung fibrosis and their interactions with CYC, but give different results on the interactions of CYC and time trends. With the joint model, the significance of the interactions between CYC and time trends indicates that the developing trends of %FVC in the two groups are different. The %FVC declines for the placebo group ($\beta_4 = -0.12$) but increases for the CYC group ($\beta_4 + \beta_8 = 0.14$) in the first 18 months. After 18 months the %FVC declines for the CYC group ($\beta_4 + \beta_5 + \beta_8 + \beta_9 = -0.45$) since the CYC effects decrease gradually after the treatment stops, while a positive slope is found for the placebo group ($\beta_4 + \beta_5 = 0.15$). However, none of the time trends is significantly different from zero. The difference might be explained by the significant covariances Σ_{U1v} and Σ_{U2v} between the random slopes in the longitudinal model and the latent variable of the survival model, which indicates dependence between the longitudinal measurement %FVC and the survival process. We also observe significantly positive coefficient v_2 which shows that there is a latent association between the two competing risks. The negative sign of Σ_{U1v} and positive sign of Σ_{U2v} together with positive v_2 indicate that in the first 18 months, there tends to be a lower risk for both treatment failure or death and informatively censored events due to dropout for patients with higher than average increasing rate of %FVC over time; after 18 months, the trend is reversed due to the negative association between the two slopes. The consequence of such

informative dropout process results in biased estimates in time trends and attenuated slope changes comparing the CYC group with the placebo group for the separate analysis. The results are confirmed by the simulation study in the later section. The overall effects of treatment CYC on %FVC scores are evaluated by testing the null hypothesis H_0: $\beta_3 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$ which yields a p-value 0.01 for the joint model and 0.03 for the separate model.

When modeling the competing risks survival data, the two methods produce similar point estimates and CIs for most parameters and identify the same set of significant effects. The joint model is able to identify the relationship (v_2) between the two competing risks much better than the separate model since the separate model does not rely on the additional information from the longitudinal endpoints. We note that, in our second simulation study in the next section, the estimate for v_2 is not reliable under the current sample size and event rates even for the joint model. Hence we would not overinterpret the quantity in this application. However, the simulation also suggests that the bias of v_2 does not seem to affect the estimation of other parameters in the joint model. No significant overall effects of CYC are identified for the time to treatment failure or death by testing the null hypothesis H_{02}: $\gamma_{23} = \gamma_{24} = \gamma_{25} = 0$ because of the relatively short follow-up period.

5 Simulation studies

We carry out two simulation studies to assess the performance of our method. In the first simulation, the data are generated with heterogeneous covariance matrices and we want to show how the parameter estimates and standard errors would be affected if we ignore the heterogeneity. The longitudinal measurements are simulated from the following model:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 X_{2i} + U_i t_{ij} + \varepsilon_{ij} \tag{15}$$

where $t_{ij} = 0, 0.15, 0.3, \ldots, 3$, is the scheduled visit time and $X_{2i} \sim$ Bernoulli(0.5) is a group indicator. The measurement error $\varepsilon_{ij} \sim N(0, 5)$. We simulate two competing risks failure times with the following cause-specific hazards:

$$\lambda_1(t; X_{1i}, X_{2i}, v_i, \gamma_1) = \lambda_{01}(t) \exp\{\gamma_{11} X_{1i} + \gamma_{12} X_{2i} + v_i\}$$
$$\lambda_2(t; X_{1i}, X_{2i}, v_i, \gamma_2, v_2) = \lambda_{02}(t) \exp\{\gamma_{21} X_{1i} + \gamma_{22} X_{2i} + v_2 v_i\} \tag{16}$$

where $X_1 \sim N(2, 1.0)$, and X_2 are shared with the longitudinal model. We use constant baseline hazards of 0.12 and 0.25 for risk 1 and risk 2 respectively to generate the event time data. The random effects are generated from the multivariate normal distribution with covariance matrices Σ_i which are decomposed into the GARPs and IVs modeled with covariates $a_{ijl} = b_{ij} = (1, X_{2i})$. In other words, the covariance matrices are different in the two groups: strong positive correlation in one group and strong negative correlation in the other. The parameters are given in Table 3. With this setup, the rate of risk 1 is approximately 0.40, the rate of risk 2 is 0.38 and censoring rate is 0.22. Longitudinal responses are missing after the observed or censored event times. The average number of total longitudinal observations is 11.6 per subject. We use $N(0, 10^5)$ priors for each component of β, γ, v, η_1 and η_2, $IG(10^{-3}, 10^{-3})$ for σ^2, and $\Gamma(10^{-3}, 10^{-3})$ for λ_0. The simulation is based on 200 Monte Carlo samples with sample size of 200 and 500. The MCMC sampling in all simulation studies is run using 5,000 iterations, and the estimation results are based on the last 2,500 iterations.

We analyze the simulated data with a joint model that models the covariance matrices with subject-specific covariates (heterogeneous) and a joint model with subject-independent

covariates (homogeneous). We also compare the results with the homogeneous model proposed by Hu et al. (2009). Tables 3 and 4 report the biases, estimated standard errors (the median of estimated standard error), and coverage rates of the 95% credible intervals. The parameters η_1 and η_2 are transformed back to variance–covariance parameters in the table. It is seen that the heterogeneous joint model gives almost unbiased estimates for all the parameters. Our method for the homogeneous model performed similarly to Hu et al. (2009). Both homogeneous joint analyses lead to large bias in some of the parameter estimates including γ_{12}, γ_{22} and v_2, which indicates that we may obtain biased parameter estimates for the survival endpoint when combining the information of the longitudinal outcome if the correlation of the two endpoints is incorrectly modeled. Therefore, ignoring the heterogeneity can result in biased estimates and invalid inference.

We conduct the second simulation by generating data with structures similar to the SLS. The longitudinal measurements and the competing risks event times are simulated from model (11–14) with $Z_{ij} = ([\text{Time}_{ij}, \text{Time}18_{ij}])$, where the covariates are generated from distributions close to what we observe in the real data. All the parameters for the joint model are set to the estimated values from the joint analysis for SLS in Table 2. Weibull distributions are used as the true baseline hazard function which produce similar risk rates and censoring rate to those in SLS. The results of the joint model and the separate analysis are compared in Table 5 using 200 simulated datasets with a sample size $m = 140$. MCMC sampling is run in 10,000 iterations, and the estimation results were based on the last 5,000 iterations. It is shown that the joint model produces good point estimates and coverage rates for most of the parameters in the longitudinal sub-model except for the time trend after 18 months (β_5) and the corresponding variance (Σ_{U22}). The separate analysis gives biased estimates for both time trends and their corresponding variances. These biases do not decrease even for a large sample size of 500 (simulation results are not reported here) since the biases are the consequences of the informative dropout process, which cannot be accounted for by the linear mixed effects alone. In contrast, the biases in the joint model are much improved with increased sample size. The random effects coefficient v_2 and frailty σ_v^2 and their standard errors are poorly estimated by the separate competing risks models. The joint model gives biased estimate for v_2 as well, which suggests that with a small sample size of 140 and low event rates (10% for risk 1 and 23% for risk 2), even the joint analysis may not provide good estimates for the frailty at the survival endpoint.

6 Discussion

For simplicity, we assume in our model that the measurement errors are mutually independent and normally distributed with constant variance. This assumption can be weakened and our method can be modified to handle correlated normal random errors. Our model also assumes that the longitudinal sub-model and survival sub-model are independent conditional on the observed data and latent variables. This may not be satisfied in a real study such as the SLS, in which one of the risks in the survival endpoint, treatment failure or death, is partly determined by the longitudinal outcome %FVC. We did some sensitivity analyses and found that our model is robust for mild violation of the independence assumption.

Our model can be extended to clustered data. Frequently, clustered data arises from multi-site clinical trials or from studies across families, in which each site or family can be viewed as a cluster. The cluster effect can be conveniently incorporated as a random effect or as design vectors for the GARP/IV parameters to take into account the heterogeneity across the cluster. Similarly, our method can be extended to recurrent event data where each subject may repeatedly experience a certain phenomenon. In addition, within our joint model framework, the linear mixed sub-model can be extended to the generalized linear mixed

effects model (Diggle et al. 2002) to handle non-normally distributed data, such as binomial or Poisson outcomes. Due to the complexity of the likelihood function in both GLMMs and joint models, only a few papers discussed such a generalized joint model framework (Molenberghs et al. 1997; Faucett et al. 1998; Larsen 2005; Yao 2008). Although in our joint model the posterior sampling distributions for the fixed and random effects in the longitudinal sub-model need to be changed, the parameters in the survival sub-model and the joint variance–covariances parameters can be sampled with our described algorithm. One of the possible approaches to sample the parameters in the GLMM sub-model is to update the fixed and random effects by constructing a normal proposal distribution with mean and variance from a single iteration of weighted least squares based on the previous value (Gamerman 1997).

We finally note that the modified Cholesky decomposition can provide an unconstrained and statistically meaningful reparameterization of a covariance matrix, but at the expense of imposing an order among the underlying random variables. Despite of this shortcoming, it has been used effectively in various applications including multivariate quality control, multivariate time series, finance and random effects models (Pourahmadi 2007).

References

Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics. 2003; 59:221–228. [PubMed: 12926706]

Celeux G, Forbes F, Robert CP, Titterington DM. Deviance information criteria for missing data models. Bayesian Anal. 2006; 4:651–674.

Chen MH. Comments on article by celeux et al. Bayesian Anal. 2006; 4:677–680.

Chib S, Greenberg E. Understanding the metropolis-hastings algorithm. Am Stat. 1995; 49:327–335.

Chiu TYM, Leonard T, Tsui KW. The matrix-logarithmic covariance model. J Am Stat Assoc. 1996; 91:198–210.

Daniels MJ, Pourahmadi M. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. Biometrika. 2002; 89:553–566.

Daniels MJ, Zhao YD. Modelling the random effects covariance matrix in longitudinal data. Stat Med. 2003; 22:1631–1647. [PubMed: 12720301]

Davidian, M.; Giltinan, DM. Nonlinear models for repeated measurement data. Chapman and Hall; New York: 1995.

De Gruttola V, Tu XM. Modeling progression of CD4-lymphocyte count and its relationship to survival time. Biometrics. 1994; 50:1003–1014. [PubMed: 7786983]

Diggle, P.; Hergerty, P.; Liang, KY.; Zeger, S. Analysis of longitudinal data. Oxford University Press; Oxford: 2002.

Elashoff R, 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 R, 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, Thomas DC. Simultaneously modeling of censored survival data and repeated measured covariates: a gibbs sampling approach. Stat Med. 1996; 16:1663–1685. [PubMed: 8858789]

Faucett CL, Schenker N, Elashoff RM. Analysis of censored survival data with intermittently observed time-dependent binary covariates. J Am Stat Assoc. 1998; 93:427–437.

Gamerman D. Sampling from the posterior distribution in generalized linear mixed models. Stat Comput. 1997; 7:57–68.

Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences (with discussion). Stat Sci. 1992; 7:457–511.

Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. Appl Stat. 1992; 41:337–348.

Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970; 57:97–109.

Heagerty PJ, Kurland BF. Misspecified maximum likelihood estimate and generalised linear mixed models. Biometrika. 2001; 88:973–986.

Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. Biostatistics. 2000; 4:465–480. [PubMed: 12933568]

Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. Stat Med. 1997; 16:259–272. [PubMed: 9004396]

Hu WH, Li G, Li N. A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. Stat Med. 2009; 29:1601–1619. [PubMed: 19308919]

Larsen K. The Cox proportional hazards model with a continuous latent variable measured by binary indicators. Biometrics. 2005; 61:1049–1055. [PubMed: 16401278]

Lin X, Raz J, Harlow SD. Linear mixed models with heterogeneous within-cluster variances. Biometrics. 1997; 53:910–923. [PubMed: 9290222]

Little RJA. Modeling the drop out mechanism in repeated measures studies. J Am Stat Assoc. 1995; 90:1112–1121.

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; 27:5679–5691. [PubMed: 18693300]

Molenberghs G, Kenward MG, Lesaffre E. The analysis of longitudinal ordinal data with nonrandom drop-out. Biometrika. 1997; 84:33–44.

Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterization. Biometrika. 1999; 86:677–690.

Pourahmadi M. Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters. Biometrika. 2007; 94:1006–1013.

Pourahmadi M, Daniels MJ. Dynamic conditional linear mixed models for longitudinal data. Biometrics. 2002; 58:225–231. [PubMed: 11890319]

Prentice RL, Breslow NE. Retrospective studies and failure time models. Biometrika. 1978; 65:153–158.

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]

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). J R Stat Soc Ser B. 2002; 64:583–639.

Tashkin DP, Elashoff RM, et al. Cyclophosphamide versus placebo in scleroderma lung disease. New Engl J Med. 2006; 354:2655–2666. [PubMed: 16790698]

Tseng YK, Hsieh F, Wang JL. Joint modelling of accelerated failure time and longitudinal data. Biometrika. 2005; 92:587–603.

Wang Y, Taylor JMG. Joint modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. J Am Stat Assoc. 2001; 96:895–905.

Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics. 1997; 53:330–339. [PubMed: 9147598]

Xu J, Zeger SL. Joint analysis of longitudinal data comprising repeated measures and times to events. Appl Stat. 2001; 50:375–387.

Yao F. Functional approach of flexibly modelling generalized longitudinal data and survival time. J Stat Plan 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. 2005; 11:151–174. [PubMed: 15940822]

Zhang F, Weiss RE. Diagnosing explainable heterogeneity of variance in random effects models. Can J Stat. 2000; 28:3–18.

Appendix

This section provides details for the full conditional distributions of the parameters used in the Gibbs sampling algorithm. We use $p(.)$ and $p(.|.)$ to denote marginal and conditional densities, respectively. We denote the prior distribution by $p_0(.)$. Based on the modified Cholesky decomposition, the random effects v_i can be written as $v_i = \sum_{l=1}^{q} a_{iql}^T \eta_1 U_{il} + e_{i,q+1}$ where $e_{i,q+1} \sim N(0, \exp(b_{i,q+1}^T \eta_2))$. Instead of sampling v_i directly, we sample $e_{i,\,q+1}$, leading to a faster convergence rate.

1. Sample β from

$$p(\beta|\cdot) \propto N\left(\left(\sum_{i=1}^{m} X_i^{(1)T} X_i^{(1)}\right)^{-1} \left(\sum_{i=1}^{n} X_i^{(1)T}(Y_i - Z_i U_i)\right), \left(\frac{\sum_{i=1}^{n} X_i^{(1)T} X_i^{(1)}}{\sigma^2}\right)^{-1}\right) p_0(\beta).$$

2. Sample σ^2 from

$$p(\sigma^2|\cdot) \sim IG\left(\frac{\sum_{i=1}^{m} n_i}{2} - 1, \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(Y_{ij} - \beta^T X_i^{(1)}(t_{ij}) - U_i^T Z(t_{ij}))^2\right) \times p_0(\sigma^2).$$

3. Sample the random effects U_i from

$$p(U_i|\cdot) \propto N(\mu_{U_i|Y_i}, \textstyle\sum_{U_i|Y_i}) \times \prod_{k=1}^{g} \exp\left\{\left(\sum_{l=1}^{q} a_{iql}^T \eta_1 U_{il} + e_{i,q+1}\right) \times v_k I(D_i = K) - H_k(T_i)\right\},$$

where $\sum_{U_i|Y_i} = \left(\frac{Z_i^T Z_i}{\sigma^2} + \sum_{u_i}^{-1}\right)^{-1}$, $\mu_{U_i|Y_i} = \sum_{U_i|Y_i}\left[\frac{Z_i^T(Y_i - X_i\beta)}{\sigma^2}\right]$ and $\sum_{u_i}^{-1} = M_i^{*T} H_i^{*-1} M_i^*$, M_i^* is a $q \times q$ matrix consisting of the first q columns and rows of M_i, H_i^* is a $q \times q$ matrix consisting of the first q columns and rows of H_i. We use the one-step Metropolis–Hastings algorithm to obtain the update in the sampling sequence with the normal density from the longitudinal data as the proposal density. The random effects U_i is obtained by first sampling a random variable form the conditional density based on the longitudinal data and then using the conditional likelihood contribution from the survival data to determine the acceptance of the new draw.

4. Sample η_1 from

$$p(\eta_1|\cdot) \propto N\left(\left(\sum_{i=1}^{m} Q_i^T H_i^{*-1} Q_i\right)^{-1}\left(\sum_{i=1}^{m} Q_i^T H_i^{*-1} U_i\right), \left(\sum_{i=1}^{m} Q_i^T H_i^{*-1} Q_i\right)^{-1}\right)$$
$$\times \prod_{k=1}^{g} \exp\left\{\left(\sum_{l=1}^{q} a_{iql}^T \eta_1 U_{il} + e_{i,q+1}\right) v_k I(D_i = k) - H_k(T_i)\right\} p_0(\eta_1),$$

where Q_i is a $q \times q_1$ matrix with first row $Q_{i1} = 0$ and jth row $Q_{ij} = \sum_{l=1}^{j-1} a_{ijl}^T U_{il}$ for $j = 2, ..., q$. We sample η_1 in two steps: sample the entries only involves U_i from the normal conditional density, sample the entries involves U_i and v_i with ARS.

5. Sample η_2 from

$$p(\eta_2|\cdot) \propto \exp\left[-\frac{1}{2}\sum_{i=1}^{m}\left(\sum_{j=1}^{q}\{b_{ij}^T\eta_2 + \left(U_{ij} - \sum_{l=1}^{j-1}a_{ijl}^T\eta_1 U_{il}\right)^2\exp(-b_{ij}^T\eta_2)\} + b_{i,q+1}^T\eta_2 + e_{i,q+1}^2\exp(-b_{i,q+1}^T\eta_2)\right)\right]p_0(\eta_2).$$

We use a Metropolis–Hastings step with a normal approximation to the full conditional as the candidate distribution. For details, see Daniels and Pourahmadi (2002).

6. Sample γ_{kr}, $k = 1, \ldots, g$, $r = 1, \ldots, R$ from

$$p(\gamma_{kr}|\cdot) \propto \exp\left[\gamma_{kr}\sum_{i=1}^{m}I(D_i=k)X_{ir}^{(2)}(T_i) - \sum_{i=1}^{m}H_k(T_i)\right]p_0(\gamma_k).$$

We use a Metropolis–Hastings step within the single component sampler to update the values of these parameters. For each of these parameters, we propose a normal density as the proposal density, which has the current value of the parameter as its mean and its standard deviation is set equal to four times the standard error of a maximum partial likelihood estimate from a standard Cox model (Wang and Taylor 2001).

7. Sample v_k with ARS from

$$p(v_k|\cdot) \propto \exp\left[\sum_{i=1}^{m}I(D_i=k)v_k\left(\sum_{l=1}^{q}a_{iql}^T\eta_1 U_{il}+e_{i,q+1}\right) - \sum_{i=1}^{m}\int_0^{T_i}\lambda_{0k}\exp\left(\gamma_k^TX_i^{(2)}+v_k\left(\sum_{l=1}^{q}a_{iql}^T\eta_1 U_{il}+e_{i,q+1}\right)\right)dt\right]p_0(v_k).$$

8. Sample $e_{i,q+1}$ ($i = 1, \ldots, m$) from

$$p(e_{i,q+1}|\cdot) \propto N(0,\exp(b_{i,q+1}^T\eta_2)) \times \prod_{k=1}^{g}\exp[e_{i,q+1}v_kI(D_i=k) - H_k(T_i)].$$

The sample is obtained by first sampling a candidate from the normal densities as its assumption and then using the conditional likelihood contribution from the survival data to determine the acceptance of the new draw.

9. Sample each piece of λ_{0k} ($k = 1, \ldots, g$) from

$$p(\lambda_{0k}^{(s)}|\cdot) \propto \Gamma(\alpha_k^s,\beta_k^s)p_0(\lambda_{0k}^{(s)}),$$

where $\alpha_k^s=\sum_{i=1}^{m}I(D_i=k, t_k^{(s-1)}<T_i \leq t_k^{(s)})+1$ indicates the number of events occurring in the time interval $(t_k^{(s-1)}, t_k^{(s)}]$, and
$\beta_k^s=\sum_{i=1}^{m}I(T_i>t_k^{(s-1)})\int_{t_k^{(s-1)}}^{\min(T_i,t_k^{(s)})}\exp(\gamma_k^TX_i^{(2)}+v_kv_i)dt$, for $s = 1, \ldots, S^k$.

Table 1

Random effects covariance matrix parameters for four different models

	[Time$_{ij}$] Est. (95% CI)	[Time$_{ij}$, Time18$_{ij}$]^a Est. (95% CI)	[Time$_{ij}$, Time18$_i$]^b Est. (95% CI)	[Time$_{ij}$, Time18$_i$]^c Est. (95% CI)
Generalized autoregressive parameters				
η_{11} Intercept	−0.27 (−1.02, 0.18)	−1.24 (−2.08, −0.38)	−1.01 (−1.81, −0.20)	−1.17 (−1.74, −0.58)
η_{11} CYC	−0.71 (−1.65, 0.08)	−0.67 (−1.82, 0.93)	−0.26 (−1.45, 0.92)	
η_{12} Intercept		0.23 (0.03, 0.71)	−0.27 (−1.23, 0.29)	−0.51 (−1.23, −0.01)
η_{12} CYC		0.44 (−1.47, 2.00)	−1.48 (−2.64, 0.15)	
η_{13} Intercept			0.36 (0.10, 0.96)	0.34 (0.08, 0.85)
η_{13} CYC			−0.65 (−1.49, 0.13)	
Innovation variances				
η_{21} Intercept	−1.49 (−1.87, −1.09)	−1.29 (−1.67, −0.90)	−1.32 (−1.69, −0.91)	−1.31 (−1.59, −1.03)
η_{21} CYC	−0.06 (−0.63, 0.49)	0.01 (−0.58, 0.59)	−0.01 (−0.59, 0.57)	
η_{22} Intercept	−4.72 (−7.21, −2.19)	0.0.23 (−0.53, 0.84)	0.24 (−0.48, 0.84)	−0.10 (−0.68, 0.38)
η_{22} CYC		−0.66 (−3.73, 0.45)	−0.68 (−1.95, 0.37)	
η_{23} Intercept		−5.11 (−7.55, −2.08)	−4.58 (−7.38, −1.79)	−4.62 (−7.26, −1.60)
Model fit				
DIC	5661.09	5565.38	5549.32	5540.80
p_D	151.77	212.84	218.26	212.98

^{a}Structured heterogeneous two-random-slope model;

^{b}Unstructured heterogeneous two-random-slope model;

^{c}Unstructured homogeneous two-random-slope model

Table 2

Analysis of SLS data using the unstructured homogeneous two-random-slope model

	Joint analysis Estimate (95% CI)	Separate analysis Estimate (95% CI)
Longitudinal outcome %FVC		
Int (β_0)	65.33 (64.72, 67.87)	65.94 (64.41, 67.47)
FVC$_0$ (β_1)	0.89 (0.80, 0.99)[†]	0.89 (0.79, 0.99)[†]
FIB$_0$ (β_2)	−1.85 (−2.94, −0.79)[†]	−1.86 (−2.94, −0.78)[†]
CYC (β_3)	−0.98 (−3.18, 1.26)	−0.76 (−2.94, 1.42)
Time (β_4)	−0.12 (−0.29, 0.06)	−0.05 (−0.23, 0.13)
Time18 (β_5)	0.27 (−0.17, 0.72)	0.11 (−0.34, 0.56)
FVC$_0 \times$ CYC (β_6)	0.14 (0.00, 0.28)[†]	0.14 (0.00, 0.28)[†]
FIB$_0 \times$ CYC (β_7)	1.74 (0.13, 3.27)[†]	1.78 (0.23, 3.33)[†]
Time $\times$ CYC (β_8)	0.26 (0.01, 0.50)[†]	0.21 (−0.04, 0.46)
Time18 $\times$ CYC (β_9)	−0.72 (−1.33, −0.08)[†]	−0.64 (−1.29, 0.00)
σ^2	21.55 (19.23, 24.25)	21.28 (18.80, 23.76)
Σ_{U11}	0.27 (0.20, 0.36)	0.25 (0.18, 0.32)
Σ_{U12}	−0.31 (−0.53, −0.14)	−0.31 (−0.49, −0.12)
Σ_{U22}	1.29 (0.70, 2.14)	1.34 (0.67, 2.00)
p-value for H_0: $\beta_3 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$	0.01	0.03
Cause-specific hazards (time to informative dropout)		
FVC$_0$ (γ_{11})	−0.06 (−0.12, −0.01)[†]	−0.06 (−0.12, −0.00)[†]
FIB$_0$ (γ_{12})	0.22 (−0.27, 0.78)	0.20 (−0.35, 0.75)
CYC (γ_{13})	0.23 (−0.60, 1.12)	0.40 (−0.46, 1.26)
FVC$_0 \times$ CYC (γ_{14})	0.10 (0.03, 0.18)[†]	0.09 (0.03, 0.15)[†]
FIB$_0 \times$ CYC (γ_{15})	0.13 (−0.60, 0.83)	0.07 (−0.64, 0.76)
p-value for H_{01}: $\gamma_{13} = \gamma_{14} = \gamma_{15} = 0$	0.08	0.07
Cause-specific hazards (time to treatment failure or death)		
FVC$_0$ (γ_{21})	0.02 (−0.07, 0.09)	0.03 (−0.07, 0.13)
FIB$_0$ (γ_{22})	0.29 (−0.62, 1.19)	0.28 (−0.80, 1.36)
CYC (γ_{23})	−1.33 (−3.44, 0.21)	−1.14 (−3.26, 0.98)
FVC$_0 \times$ CYC (γ_{24})	−0.07 (−0.21, 0.06)	−0.08 (−0.24, 0.08)
FIB$_0 \times$ CYC (γ_{25})	−0.58 (−2.31, 0.91)	−0.88 (−2.78, 1.02)
p-value for H_{02}: $\gamma_{23} = \gamma_{24} = \gamma_{25} = 0$	0.39	0.48
Random effects for survival endpoint		
v_2	3.04 (1.27, 7.65)	−0.31 (−79.80, 81.16)
σ_v^2	0.38 (0.07, 1.42)	0.04 (0.00, 0.40)
Covariance of U_i and v_i		
Σ_{U1v}	−0.25 (−0.51, −0.09)	–

	Joint analysis Estimate (95% CI)	Separate analysis Estimate (95% CI)
Σ_{U2v}	0.60 (0.21, 1.33)	–

[†] p-Value < 0.05

Table 3

Comparison of simulated bias, standard error (SE) and coverage probability (CP) between two homogeneous (incorrect) models and a heterogeneous (correct) model (sample size = 200)

m	Parameter	True	Homogeneous[a] Bias	SE	CP	Homogeneous Bias	SE	CP	Heterogeneous Bias	SE	CP
200	Longitudinal										
	β_0	10	0.023	0.055	0.935	0.017	0.054	0.925	0.007	0.057	0.914
	β_1	1.5	−0.029	0.069	0.950	−0.019	0.065	0.960	−0.010	0.073	0.921
	β_2	−1	0.029	0.173	0.885	0.036	0.215	0.860	−0.029	0.117	0.911
	σ^2	1	−0.002	0.30	0.965	0.001	0.029	0.965	0.002	0.033	0.938
	σ^2_{u1}	2.5	–	–	–	–	–	–	−0.001	0.525	0.942
	σ^2_{u0}	0.5	–	–	–	–	–	–	0.041	0.182	0.935
	Survival										
	γ_{11}	0.8	−0.26	0.137	0.885	0.011	0.158	0.920	−0.006	0.139	0.925
	γ_{12}	−1	**0.133**	0.289	0.840	**0.047**	0.416	0.880	−0.024	0.302	0.932
	γ_{21}	0.5	−0.066	0.163	0.910	−0.066	0.195	0.875	−0.026	0.150	0.928
	γ_{22}	−1	**0.378**	0.347	0.755	**0.350**	0.426	0.810	−0.007	0.359	0.932
	ν_2	1.5	**−0.281**	1.049	0.715	**−0.352**	1.891	0.715	−0.116	0.863	0.912
	$\sigma^2_{\nu 1}$	1	–	–	–	–	–	–	0.026	0.706	0.825
	$\sigma^2_{\nu 0}$	0.5	–	–	–	–	–	–	−0.048	0.301	0.801
	Joint covariances										
	σ_{w1}	1.5	–	–	–	–	–	–	**−0.216**	0.785	0.805
	σ_{w0}	−0.4	–	–	–	–	–	–	0.080	0.247	0.787

[a]Homogeneous model from Hu et al. (2009)

The bold numbers represent relatively large biases

Table 4

Comparison of simulated bias, standard error (SE) and coverage probability (CP) between a homogeneous (incorrect) model and a heterogeneous (correct) model (sample size = 500)

m	Parameter	True	Homogeneous[a]			Homogeneous			Heterogeneous		
			Bias	SE	CP	Bias	SE	CP	Bias	SE	CP
500	Longitudinal										
	β_0	10	0.024	0.035	0.895	0.024	0.035	0.893	0.001	0.037	0.916
	β_1	1.5	−0.032	0.044	0.875	−0.031	0.043	0.902	−0.002	0.045	0.948
	β_2	−1	0.055	0.113	0.835	0.053	0.127	0.907	0.008	0.087	0.924
	σ^2	1	−0.001	0.019	0.960	−0.001	0.019	0.937	0.001	0.019	0.948
	σ^2_{u1}	2.5	–	–	–	–	–	–	0.007	0.346	0.920
	σ^2_{u0}	0.5	–	–	–	–	–	–	0.017	0.105	0.913
	Survival										
	γ_{11}	0.8	−0.004	0.085	0.890	−0.017	0.091	0.917	−0.005	0.087	0.920
	γ_{12}	−1	**0.153**	0.177	0.815	**0.162**	0.219	0.815	−0.016	0.171	0.937
	γ_{21}	0.5	−0.001	0.099	0.900	−0.033	0.101	0.922	−0.012	0.093	0.920
	γ_{22}	−1	**0.336**	0.211	0.605	**0.346**	0.216	0.649	−0.021	0.207	0.920
	ν_2	1.5	**0.102**	0.525	0.825	**0.152**	1.516	0.809	−0.075	0.843	0.937
	σ^2_{v1}	1	–	–	–	–	–	–	−0.034	0.354	0.906
	σ^2_{v0}	0.5	–	–	–	–	–	–	−0.030	0.182	0.899
	Joint covariances										
	σ_{w1}	1.5	–	–	–	–	–	–	−0.064	0.398	0.923
	σ_{w0}	−0.4	–	–	–	–	–	–	0.009	0.145	0.882

[a] Homogeneous model from Hu et al. (2009)

The bold numbers represent relatively large biases

Comparison of simulated bias, standard error (SE) and coverage probability (CP) between joint and separate analyses (sample size = 140)

Parameter	True	Joint			Separate		
		Bias	**SE**	**CP**	**Bias**	**SE**	**CP**
Longitudinal							
Fixed effects							
β_0	66.33	0.004	0.765	0.975	0.459	0.768	0.925
β_1	0.89	−0.002	0.051	0.955	0.002	0.051	0.940
β_2	−1.85	−0.028	0.597	0.925	−0.030	0.592	0.940
β_3	−0.98	−0.049	1.031	0.955	0.210	1.019	0.960
β_4	−0.12	−0.002	0.099	0.935	**0.082**	0.096	0.815
β_5	0.27	**−0.111**	0.278	0.900	**−0.299**	0.239	0.775
β_6	0.14	−0.002	0.073	0.950	−0.003	0.072	0.935
β_7	1.74	0.026	0.839	0.945	0.032	0.832	0.935
β_8	0.26	−0.006	0.130	0.940	**−0.036**	0.127	0.945
β_9	−0.72	0.040	0.337	0.935	**0.101**	0.327	0.920
Random effects							
σ^2	21.55	0.049	1.357	0.935	0.041	1.310	0.940
Σ_{U11}	0.27	−0.001	0.038	0.965	**−0.022**	0.035	0.855
Σ_{U12}	−0.31	0.060	0.095	0.860	0.066	0.083	0.860
Σ_{U22}	1.29	**−0.284**	0.401	0.790	**0.235**	0.345	0.790
Competing risks							
Fixed effects							
γ_{11}	−0.06	−0.004	0.030	0.950	0.005	0.025	0.940
γ_{12}	0.22	0.077	0.320	0.925	0.017	0.270	0.930
γ_{13}	0.23	0.011	0.518	0.920	0.038	0.448	0.935
γ_{14}	0.10	0.005	0.042	0.950	−0.010	0.031	0.965
γ_{15}	0.13	0.063	0.434	0.920	−0.064	0.345	0.955
γ_{21}	0.02	−0.009	0.041	0.945	−0.008	0.029	0.945
γ_{22}	0.29	0.055	0.473	0.925	−0.061	0.356	0.925

Parameter	True	Joint			Separate		
		Bias	SE	CP	Bias	SE	CP
γ_{23}	−1.33	−0.052	1.107	0.925	−0.098	1.049	0.915
γ_{24}	−0.07	0.003	0.075	0.960	0.010	0.059	0.995
γ_{25}	−0.58	−0.010	0.891	0.940	0.095	0.805	0.965
Random effects							
ν_2	3.04	**−0.456**	1.043	0.925	**−0.414**	3.212	1.000
σ_ν^2	0.38	0.076	0.728	0.995	**−0.333**	0.058	0.285
Joint covariances							
Σ_{U1v}	−0.25	−0.100	0.128	0.970	–	–	–
Σ_{U2v}	0.60	−0.022	0.418	0.945	–	–	–

The bold numbers represent relatively large biases