A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data

Wenhua Hu[1,*], **Gang Li**[2], and **Ning Li**[3]

[1]Bristol-Myers Squibb, 5 Research Pkwy, Wallingford, CT 06450, U.S.A.

[2]Department of Biostatistics, School of Public Health, University of California at Los Angeles, CA, U.S.A.

[3]Department of Epidemiology and Biostatistics, College of Public Health and Health Professions, University of Florida, FL, U.S.A.

SUMMARY

In this paper, we develop a Bayesian method for joint analysis of longitudinal measurements and competing risks failure time data. The model allows one to analyze the longitudinal outcome with nonignorable missing data induced by multiple types of events, to analyze survival data with dependent censoring for the key event, and to draw inferences on multiple endpoints simultaneously. Compared with the likelihood approach, the Bayesian method has several advantages. It is computationally more tractable for high-dimensional random effects. It is also convenient to draw inference. Moreover, it provides a means to incorporate prior information that may help to improve estimation accuracy. An illustration is given using a clinical trial data of scleroderma lung disease. The performance of our method is evaluated by simulation studies.

Keywords

joint modeling; competing risks; longitudinal data; Bayesian approach

1. INTRODUCTION

In many studies, a response variable (e.g. biomarker) is repeatedly measured during the followup and the occurrence of some key event is monitored for each subject. Often the key event is dependently censored by some events such as disease-related dropout. An example is the Scleroderma Lung Study (SLS) [1]. The objective of this study is to evaluate the effectiveness of oral cyclophosphamide (CYC) versus placebo in the treatment of active, symptomatic lung disease due to scleroderma. The primary lung function outcome is forced vital capacity (FVC, as % predicted) measured longitudinally. The time to treatment failure or death is also recorded. In addition to independent censoring such as moving to another place, there is also dependent censoring by disease-related or treatment-related dropouts, such as those due to worsening disease, or serious adverse events (AEs).

When modeled separately, random effects models [2] and Cox proportional hazards models [3] are often used to evaluate treatment effect on the repeated measures and survival data, respectively. However, joint modeling of both outcomes is often necessary. First of all, for the longitudinal analysis, missing data of the longitudinal response induced by the events

*Correspondence to: Wenhua Hu, Bristol-Myers Squibb, 5 Research Pkwy, Wallingford, CT 06450, U.S.A. wenhua.hu@bms.com.

might be related to the response itself, making the missing mechanism nonignorable. The separate model without modeling the missing mechanism, such as a linear mixed effects model may give biased inference. For instance, in the SLS study, patients with lower (worse) %FVC may tend to experience failure or drop out from the study earlier than healthier individuals. This leads to fewer measurements of low %FVC, which might result in overestimating the overall %FVC level for the period after the dropouts. Second, some unmeasured variable or latent factor associated with time to events might induce unobserved heterogeneity among subjects. With the help of joint modeling longitudinal measurements, the estimation is expected to be more efficient for parameters in survival endpoint. Third, joint modeling makes it possible to evaluate treatment effects on the two endpoints simultaneously.

Joint modeling has been studied extensively in recent years. Much of the previous work on joint modeling can be classified into three categories: (i) inference for longitudinal measurements with nonignorable missing data due to dropout [4–7]; (ii) inference for time-to-event conditional on the time-dependent covariate [8–12]; and (iii) simultaneous inference for the longitudinal measurements and time-to-event [13, 14]. Most of these joint models assume single failure type with independent censoring. Most recently, Elashoff *et al*. [15, 16] considered a more general joint model, which incorporates competing risks model for survival endpoint, using a likelihood approach. Their model allows dependent censoring by treating it as a competing risk. Their proposed competing risks models could be a cause-specific model [17] or a mixture sub-model [18], in which the probability of failure type and the conditional hazard rate for each failure type were modeled with the logistic model and the proportional hazards model, respectively. They derived EM-based algorithms to obtain the parameter estimates, and a profile likelihood method was used to estimate the standard errors. However, their EM algorithm involves multi-dimensional integration which can be formidable and computationally expensive for high-dimension random effects problems. The rounding error from numerical integration may also result in inaccurate inference for the parameters. Another problem with the likelihood approach is that variance estimation usually requires asymptotic arguments and can be quite complicated to derive [19]. Moreover, there is always the issue of whether the sample size is large enough for the asymptotic approximation to be valid.

The purpose of this paper is to develop a Bayesian framework for joint modeling of longitudinal measurements and competing risks failure time data. Faucett and Thomas [9], Wang and Taylor [10], and Brown and Ibrahim [11] explored a Bayesian method for a single type of failure with independent censoring. To the best of our knowledge, this paper is the first to apply the Bayesian idea to the more general joint model of Elashoff *et al*. [16] with multiple types of failures in the failure time data. Under the Bayesian framework, variance estimates and other posterior summaries can be calculated directly from the posterior samples using standard Markov chain Monte Carlo (MCMC) sampling techniques. Unlike the likelihood approach, the Bayesian method can avoid high-dimensional integration and computational implementation is typically not more difficult as the number of random effects grows. For instance, in the SLS study (Section 4, Table I), the cost of computation for the four models are similar even though the last model involves four random effects. The Bayesian method also enables one to incorporate prior information into the current analysis [20] in a natural way. As illustrated in our simulation (Section 5, Table V), for small samples, incorporation of appropriate informative priors may improve the estimation accuracy of some parameters by decreasing the bias or standard error.

The remainder of the paper is organized as follows. Section 2 gives the formulation of our joint model. Section 3 describes the posterior densities and the sampling procedures. Section 4 presents the analysis of a clinical trial data for Scleroderma lung disease using the

developed method. In Section 5, the performance of our joint model is evaluated with simulation studies. We conclude with a discussion in Section 6.

2. THE JOINT MODEL

Our joint model consists of two submodels: a linear mixed effects model and a competing risk model. Suppose there are m subjects in the study. For subject i, the longitudinal response Y_i follows a linear mixed effects model:

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

where Y_i is $n_i \times 1$, $X_i^{(1)}(t)$ is an $n_i \times p$ design matrix for fixed effects, β is a $p \times 1$ parameter vector of regression coefficients, commonly referred to as fixed effects in the model, Z_i is an $n_i \times Q$ matrix of covariates for the $Q \times 1$ vector of errors U_i, and ε_i is an n_i It is assumed that ε_i is independent of U_i and is normally distributed $N_{ni}(0, \sigma^2 I_{ni})$.

During the followup, each subject may experience one of K distinct types of failure or could be right censored. Occurrence of one type of failure precludes us observing other types of failures. Let $C_i = (T_i, D_i)$ be the competing risks data on subject i, where T_i is the failure time or censoring time, and D_i takes a value from $0, 1, \ldots, K$, with $D_i = 0$ indicating a censored event and $D_i = k$, where $k = 1, \ldots, K$, indicating that subject i fails from the kth type of failure. Dependent censoring can be treated as one of the K types of failure. The censoring mechanism other than the considered dependent censoring is assumed to be independent of the survival time. Let $X_i^{(2)}(t)$ be a $1 \times R$ row vector of covariates potentially related to the competing risks process. A cause-specific hazard model with random effects is assumed for competing risks survival data:

$$\begin{aligned}\lambda^{(k)}\left(t; X_i^{(2)}, v_i, \gamma^{(k)}, \nu^{(k)}\right) &= \lim_{h \to 0} h^{-1} P\left[t \le T_i < t+h, D_i = k | T_i \ge t, X_i^{(2)}(t), v_i\right] \\ &= \lambda_0^{(k)}(t) \exp\left\{X_i^{(2)}(t)\gamma^{(k)} + \nu^{(k)} v_i\right\}\end{aligned} \tag{2}$$

where the function $\lambda^{(k)}\left(t; X_i^{(2)}, v_i, \gamma^{(k)}, \nu^{(k)}\right)$ gives the instantaneous hazard rate from cause k at time t, given the regression vector $X_i^{(2)}(t)$ and the latent unknown factor v_i, in the presence of the other failure types. The regression coefficient $\nu^{(k)}$ represents the effect of the latent variable v_i and $\gamma^{(k)}$ represents the effects of the observed covariates in $X_i^{(2)}(t)$ to the kth type of failure. Suppose that $X_i^{(2)}(t) = \left[X_i^{(21)}, X_i^{(22)}(t)\right]$, $\gamma^{(k)} = \left[\gamma_1^{(k)}, \gamma_2^{(k)}\right]$ where $X_i^{(21)}$ is time-independent and $X_i^{(22)}(t)$ is time-dependent. Our model allows both nonignorable monotone missing data caused by death/dropout and ignorable missing data (missing-at-random) at intermittant visit times. It is possible to have two types of missing data for the longitudinal measurements after the event time.

We assume that U_i and v_i jointly have a multivariate normal distribution:

$$\begin{pmatrix} U_i \\ v_i \end{pmatrix} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_u & \Sigma_{uv} \\ \Sigma_{uv}^T & \sigma_v^2 \end{pmatrix}\right) \tag{3}$$

The parameter $\nu^{(1)}$ is set to 1 to ensure identifiability. We further assume that the kth baseline hazard is a step function, $\lambda_0^{(k)}(t) = \lambda_{0s}^{(k)}$, for $t_{s-1}^{(k)} < t \le t_s^{(k)}$, where $0 < t_1^{(k)} < \cdots < t_{S(k)}^{(k)} < \infty$ is

a partition of $(0, \infty)$. Here $S^{(k)}$ is the total number of pieces of the step function for type k failure and $s=1,\ldots, S^{(k)}$. Note that any hazard function can be approximated by a piecewise step function. More discussion on this issue is given later in the discussion section.

The parameters Σ_{uv} measure the latent association between the longitudinal measure and time-to-event introduced by U and v. If $\Sigma_{uv} = 0$, then the joint model we proposed would reduce to those of separate analysis and the joint model and separate model give identical inference for fixed effects. In addition, recall that dependent censoring can be treated as one of the K competing risks. The magnitude of heterogeneity induced by the unobserved factor v to the first type of event is characterized by σ_v^2. The parameters v measure the latent association between the competing risks. The parameter $v^{(k)}$ measures the latent dependence of the kth type of failure and the first type of failure. For example, in our SLS application, $\sigma_{uv} < 0$ would indicate that there is a latent dependence between the longitudinal response %FVC and the survival process. The significance of the coefficient $\sigma^{(2)}$ would indicate that there is a latent association between the two types of events and that the censoring of the key event treatment failure or death is informative. The negative value of σ_{uv} and the positive value for $v^{(2)}$ would indicate that there is a higher risk of treatment failure and/or disease-related dropout for subjects with more rapidly declining %FVC during the study, which could lead to biased estimation of the time trend and the variance of the random slope.

3. ESTIMATION METHODS

3.1. The likelihood

This section describes our Bayesian estimation procedure. The parameters in the joint model are $(U_i, v_i, i = 1,\ldots,m)$, β, σ^2, Σ_u, Σ_{uv}, σ_v^2, $\gamma=(\gamma^{(1)}, \gamma^{(2)},\ldots, \gamma^{(K)})$, $v=(v^{(2)},\ldots, v^{(K)})$, and $\lambda_0 = \left(\lambda_{01}^{(1)}, \lambda_{02}^{(1)}, \cdots, \lambda_{0S^{(K)}}^{(K)}\right)$.

To ease the computation, we reparameterize (3) as

$$v_i = \theta' U_i + e_i \tag{4}$$

where θ is Q vector, $e_i \sim N(0, \sigma_e^2)$, and $U_i \sim N(0, \Sigma_u)$. It is easy to see that $\sigma_v^2 = \theta' \Sigma_u \theta + \sigma_e^2$ and $\Sigma_{uv} = \Sigma_u \theta$. The reparameterization of (4) provides an attractive alternative for specitfying the noninformative priors for the relationship between two latent factors U_i and v_i. With this reparameterization, one only needs to specify priors for θ and σ_e^2. In contrast, (3) would require specification of noninformative priors for the whole covariance matrix of $\left(U_i', v_i\right)'$ which is notoriously tricky.

Define: $\Omega = ((U_i, e_i, i=1,\ldots,m), \beta, \sigma^{(2)}, \Sigma_u, \gamma, v, \lambda_0, \theta, \sigma_e^2)$. Assume that parameters β, σ^2, Σ_u, γ, v, λ_0, θ, and σ_e^2 have independent priors. Assume further that for each subject, the longitudinal data are independent of the survival data conditional on Ω, X_i, Z_i, and the latent factors U_i and v_i. Then, the contribution of subject i to the conditional likelihood is:

$$
\begin{aligned}
f(Y_i, t_i, d_i | X_i, Z_i, \Omega) &= f\left(Y_i | X_i^{(1)}, Z_i, \Omega\right) f\left(t_i, d_i | X_i^{(2)}, \Omega\right) \\
&= f\left(Y_i | X_i^{(1)}, Z_i, \beta, U_i, \sigma^2\right) f\left(t_i, d_i | X_i^{(2)}, \gamma, v, \lambda_0, \theta, U_i, e_i\right) \\
&= \left(2\pi\sigma^2\right)^{-n_i/2} \exp\left\{-\left(Y_i - X_i^{(1)}\beta - Z_i U_i\right)^T \left(Y_i - X_i^{(1)}\beta - Z_i U_i\right) / 2\sigma^2\right\} \times \prod_{k=1}^{K} \left(\left(\lambda_1^{(k)}(t_i)\right)^{I(d_i=k)} \exp\left\{-H^{(k)}(t_i)\right\}\right)
\end{aligned} \tag{5}
$$

where $\lambda_i^{(k)}(t) = \lambda_0^{(k)}(t) \exp\left\{X_i^{(2)}(t)\gamma^{(k)} + (\theta' U_i + e_i)\nu^{(k)}\right\}$ and $H^{(k)}(t_i) = \int_0^{t_i} \lambda_0^{(k)}(t) \exp\left\{X_i^{(2)}(t)\gamma^{(k)} + (\theta' U_i + e_i)\nu^{(k)}\right\} dt$.

Note that, under the piecewise constant assumption,

$$H^{(k)}(t_i) = \exp\left\{X_i^{(21)}\gamma_1^{(k)} + \left(\theta' U_i + e_i\right)\nu^{(k)}\right\} \sum_{s=1}^{S^{(k)}} H_{is}^{(k)}\left(\lambda^{(k)}, \gamma^{(k)}, \nu^{(k)}\right)$$

(6)

with $H_{is}^{(k)}\left(\lambda^{(k)}, \gamma^{(k)}, \nu^{(k)}\right) = I\left(t_i > t_{s-1}^{(k)}\right) \lambda_{0s}^{(k)} \int_{t_{s-1}^{(k)}}^{\min\left(t_i, t_s^{(k)}\right)} \exp\left\{X_i^{(22)}(t)\gamma_2^{(k)}\right\} dt$, $s = 1, \ldots, S^{(k)}$.

3.2. MCMC sampling procedure

Markov chain Monte Carlo (MCMC) methods are used for posterior sampling [21]. Itinvolves sampling directly from the full distribution, Metropolis–Hastings (MH) sampling [22, 23], and adaptive rejection sampling (ARS) [24], as described below.

The full conditional distribution of the parameters is given in the Appendix. For parameters β, σ^2, Σ_u, σ_e^2, and λ_0 in our joint model, each of the conditional distributions is the product of some standard distribution and the prior. If a conjugate prior is used, drawing random variates from their full conditional distributions is straightforward. For the parameters U_i ($i=1,\ldots,m$) the full conditional density is the product of a normal density from the longitudinal data and a factor from the survival data. We use the one-step MH algorithm to obtain the update in the sampling sequence, and the normal density from the longitudinal data as the proposal density. The parameter U_i is obtained by first sampling a random variate from 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.

For the parameter $\gamma_r^{(k)}$ ($k=1,\ldots,K$, $r=1,\ldots,R$), in the survival model, we use a Metropolis sampler to update the values of these parameters since the direct sampling is not available. For each of these parameters, we propose a normal density as the proposal density, which has the current parameter value as its mean and its standard deviation set to four times the standard error of a maximum partial likelihood estimate from a standard Cox model [10]. The parameter e_i ($i=1,\ldots,m$) is obtained by first sampling a random variate from the normal densities as its assumption and then by using the conditional likelihood contribution from the survival data to determine the acceptance of the new draw. Parameters θ and $\nu^{(k)}$ are obtained using ARS.

The initial values of the parameters for sampling are obtained by modeling the longitudinal data and survival data separately. The initial value for $\lambda_{0s}^{(k)}$ ($s=1,\ldots,S^{(k)}$, $k=1,\ldots,K$) can be obtained by drawing a random variate from the gamma full conditional distribution in the Appendix.

3.3. Inference

We estimate the parameters by the posterior medians. The approximate 95 per cent probability interval is based on the 2.5th percentile and the 97.5th percentile. Standard errors are obtained using the standard deviations of the posterior samples. The procedure has been shown to work well by the application and simulation study in the later sections.

4. APPLICATION

In this section, we apply our method to analyze the SLS data [1] discussed earlier in the Introduction. The SLS is a double-blinded, randomized, placebo-controlled trial on patients with evidence of active alveolitis and scleroderma-related interstitial lung disease. The primary study objective is to determine if oral cyclophosphamide (CYC) compared with placebo can either improve pulmonary function as measured by forced vital capacity (FVC, as % predicted) or reduce the risk of treatment failure or death. Patients received CYC or placebo for 1 year and were followed for an additional year. Initially, patients were given a partial dose, and the dose was gradually increased to the target level within the first 6 months. %FVC was assessed every 3 months. Our analyses use data from 6–21 months from the 140 subjects who completed at least six months of treatment. Baseline covariates such as age and degree of fibrosis in the lung (FIB) were also recorded. It should be noted that the data we use here do not include the measurements of %FVC after the observed treatment failures since the patients might have switched to other treatments after the treatment failures, which differs from the analysis by Elashoff *et al*. [16].

We observed 16 treatment failures or deaths, 37 disease-related dropouts due to worsening disease or AE, and 12 independent dropouts with no evidence that the dropouts were related to disease. The rest were administratively censored. We consider treatment failure or death as risk 1 and disease-related dropout as risk 2. The average number of visits per patient was 6.0. As discussed earlier in Section 1, both treatment failure or death and disease-related dropout could cause nonignorable missing data for the longitudinal measure %FVC.

We applied the joint model described in Section 2 to the SLS data after adjustment of baseline %FVC, age, and degree of fibrosis in the lung to their means. In the sub-model for %FVC, we fitted the following linear mixed effects model with random effects U_i:

$$
\begin{aligned}
FVC_{ij} = {} & \beta_0 + \beta_1 \text{Age}_i \\
& + \beta_2 \text{CYC}_i \\
& + \beta_3 \text{FBI}_i \\
& + \beta_4 \text{FVC}_{0i} \\
& + \beta_5 \text{Age}_i \\
& \times CYC_i \\
& + \beta_6 \text{FBI}_i \\
& \times CYC_i \\
& + \beta_7 \text{FVC}_{0i} \\
& \times CYC_i \\
& + \beta_8 \text{time}_{ij} \\
& + \beta_9 \text{time}_{ij} \\
& \times CYC_i \\
& + Z_{ij} U_i + \epsilon_{ij}
\end{aligned}
\tag{7}
$$

in which, for subject i, Age_i is the age at baseline, FVC_{0i} is the baseline %FVC, FIB_i is the degree of fibrosis in the lung, CYC_i is the CYC treatment group indicator, time_{ij} is the measurement time for visit j in months, U_i represents random effects, and the ε_{ij}'s are mutually independent normal measurement errors with mean 0 and variance σ^2. Here, Z_{ij} includes covariates for random effects. We considered multiple choices for Z_{ij} and selected the model based on the DIC criterion [25]. The results are shown later.

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

$$\lambda^{(1)}(t) = \lambda_0^{(1)}(t) \exp\left\{\gamma_1^{(1)}\text{Age}_i + \gamma_2^{(1)}\text{CYC}_i + \gamma_3^{(1)}\text{FIB}_i + \gamma_4^{(1)}\text{FVC}_{0i} + \gamma_5^{(1)}\text{Age}_i \times \text{CYC}_i + \gamma_6^{(1)}\text{FIB}_i \times \text{CYC}_i + \gamma_7^{(1)}\text{FVC}_{0i} \times \text{CYC}_i + v_i\right\} \quad (8)$$

and

$$\lambda^{(2)}(t) = \lambda_0^{(2)}(t) \exp\left\{\gamma_1^{(2)}\text{Age}_i + \gamma_2^{(2)}\text{CYC}_i + \gamma_3^{(2)}\text{FIB}_{0i} + \gamma_4^{(2)}\text{FVC}_{0i} + \gamma_5^{(2)}\text{Age}_i \times \text{CYC}_i + \gamma_6^{(2)}\text{FIB}_i \times \text{CYC}_i + \gamma_7^{(2)}\text{FVC}_{0i} \times \text{CYC}_i + v^{(2)}v_i\right\} \quad (9)$$

in which the latent variables follow a bivariate normal distribution:

$$\begin{pmatrix} U_i \\ v_i \end{pmatrix} \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_u & \Sigma_{uv} \\ \Sigma_{uv} & \sigma_v^2 \end{pmatrix}\right)$$

in which $\Sigma_{uv} = [\sigma_{u1v}, \sigma_{u1v}, \ldots, \sigma_{uQv}]^{\text{T}}$.

We used a piecewise exponential baseline hazard function with three knots for the event of treatment failure or death and the event disease-related dropout, respectively, and the time points defining the steps are taken to be 3 equally split percentiles of the observed dropout times for the two event types, respectively. We used a combination of weakly informative and standard prior distributions. The corresponding priors for the parameters are given as $\beta_0 \sim \text{N}(70, 400)$ with other β's $\beta_l \sim \text{N}(0, 400)$ such that $l=1,\ldots,9$, $\sigma^2 \sim \text{IG}(0.001, 0.001)$, $\sigma_u^2 \sim \text{IG}(0.001, 0.001)$, $\sigma_e^2 \sim \text{IG}(0.001, 0.001)$, $\theta \sim \text{N}(0, 10^{10})$, $\lambda_{0s}^{(k)} \sim \Gamma(0.1, 0.1)$ such that s $1,\ldots, S^{(k)}$ and $k=1, 2$; $\gamma_r^{(k)} \sim \text{N}(0, 20)$ such that $k=1, 2$ and $r=1,\ldots,7$, and $v^{(k)} \sim \text{N}(0, 20)$ where $k=2$. Recall that θ is defined in (4). In case θ is a vector, we used an independent element of flat normal prior $\text{N}(0, 10^{10})$ for each element of θ.

Table I summarizes the results for some models with different structures of random effects Z_{ij}. The model with a random slope ($[\text{Time}_{ij}]$) has the smallest DIC, indicating that it provides the best fit for the data among the four models. The inference for Σ_{uv} indicates that there exists latent dependence between the longitudinal measurements and the survival data since at least one element in Σ_{uv} is negatively significant in each model. It is worth noting that all the models give a positive significance of $v^{(2)}$, indicating that there is latent association between the two types of failures and censoring for the key event of treatment failure or death by disease-related dropout may be dependent. Subjects with larger latent factor v tend to experience treatment failure and disease-related dropout earlier than those with a lower v.

The results of parameter estimates and 95 per cent confidence intervals based on the selected random slope model with Z_{ij} as ($[\text{Time}_{ij}]$) are summarized in Table II, together with those from a separate analysis. In the separate analysis, the longitudinal and survival data are fitted separately using two marginal sub-models, a linear mixed model (7) and a cause-specific competing risks model with a common frailty (8)–(9), respectively.

It is seen from Table II, based on our joint model, that the degree of fibrosis in the lung is negatively correlated with %FVC ($\widehat{\beta_3} = -1.67$, 95 per cent CI: $(-2.76, -0.58)$). The baseline %FVC is highly correlated to %FVC during the followup ($\widehat{\beta_4} = 0.90$, 95 per cent CI: $(0.81, 1.00)$). The %FVC for the placebo group declines significantly by a factor of 0.22 each month (95 per cent CI for β_8: $(-0.38, -0.05)$). This result differs from the analysis by

Elashoff *et al.* [16], who included the measurements of %FVC after the observed treatment failures and found nonsignificant time trend. After carefully examining the available %FVC values after the observed treatment failures, we find that these measurements tend to get greater than those values prior to the treatment failures, which results in a weakened decline time trend in their analysis. The significance of the interaction term between the treatment group and time indicates that the developing trend of %FVC for the CYC group is different from that of the placebo group. During one month, the %FVC decreases for the CYC group is 0.25 (95 per cent CI for β_9: (0.01,0.48)) less than that of the placebo group. In contrast to a declining %FVC in the placebo group, there is an increasing trend for %FVC in the CYC group ($\widehat{\beta_8}+\widehat{\beta_9}$=0.03, 95 per cent CI: (−0.14,0.19)).

Our joint analysis for the survival endpoint concludes that older subjects have an increased risk in experiencing treatment failure ($\widehat{\gamma}_1^{(1)}$=0.10, 95 per cent CI: (0.02,0.20)) and disease-related dropout ($\widehat{\gamma}_1^{(2)}$=0.05, 95 per cent CI: (0.01,0.09)). Baseline %FVC also affects the risk to experience disease-related dropout ($\widehat{\gamma}_4^{(2)}$= − 0.06, 95 per cent CI: (−0.12,−0.02)), but this effect in the CYC group differs from the placebo group ($\widehat{\gamma}_7^{(2)}$=0.09, 95 per cent CI: (0.02,0.16)).

The joint and separate analyses give different results on the time trend parameters. Based on our joint model, the %FVC for the placebo group declines significantly ($\widehat{\beta_8}$= − 0.22, 95 per cent CI (−0.38,−0.05)). The interaction term between time trend and the treatment group is significant ($\widehat{\beta_9}$=0.25, 95 per cent CI: (0.01,0.48)). However, the separate linear mixed effects model analysis fails to detect a significant decline in %FVC for the placebo group ($\widehat{\beta_8}$= − 0.14, 95 per cent CI (−0.31,0.03)), and does not reach significance for the difference of developing trends between the placebo and CYC groups ($\widehat{\beta_9}$=0.21, 95 per cent CI: (−0.03,0.44)). The posterior distribution of β_8 (time trend for the placebo group) and $\beta_8+\beta_9$ (time trend for the CYC group) are noticeably seen different, as shown in Figures 1 and 2.

Asides from the time trends, the variance of the random slope estimated with the separate analysis ($\widehat{\sigma}_u^2$=0.26, 95 per cent CI (0.20,0.36)) is smaller than that in joint analysis ($\widehat{\sigma}_u^2$=0.30, 95, per cent CI:(0.22,0.41)). Our simulations in the later section show the underestimation of this parameter with the separate analysis in presence of nonignorable missing data due to the response related events.

We also notice that in the survival submodel the estimates and confidence intervals for some parameters derived with the joint model differ greatly from the separate models. With the help of modeling longitudinal outcomes, the joint model is expected to improve the estimation of the parameters in the survival model.

The differences between the separate and joint analysis might be explained by the negative significance for the covariance σ_{uv} (estimate −1.18, 95 per cent CI: (−2.16, −0.62)) between the latent variable of the longitudinal model and that of the survival model. This indicates that there is a dependence between the longitudinal response %FVC and the survival process. The significance of the coefficient $v^{(2)}$ (estimate 0.21, 95 per cent CI: (0.01,0.52)) indicates that there is a latent association between the two types of events. The negative value of σ_{uv} and the positive value for $v^{(2)}$ indicate that there is a higher risk of treatment failure and/or disease-related dropout for subjects with more rapidly declining %FVC during the study, which leads to biased estimation of the time trend and the variance of the random slope.

5. SIMULATION STUDY

First we conducted a simulation to compare the performance of the likelihood and the Bayesian approaches for the joint model we proposed. The longitudinal responses were simulated from a random slope linear mixed model

$$Y\left(t_{ij}\right)=\beta_0+\beta_1X_{1i}+\beta_2t_{ij}+t_{ij}U_i+\varepsilon_{ij} \tag{10}$$

for $j=1,\ldots,n_i$, where $\varepsilon_{ij} \sim N(0,\sigma^2)$. The covariate X_{1i} was simulated from a Bernoulli distribution with success probability of 0.5 acting as a treatment group indicator for a 1:1 allocation in randomization trials. The data for longitudinal measurements were generated for every 0.5 unit from 0 to 5. We simulated two competing risks, risk 1 and 2, from the following cause-specific proportional hazards model with the random effects v_i:

$$\lambda^{(1)}\left(t;X_{1i},X_{2i},U_i,\gamma^{(1)}\right)=\lambda_0^{(1)}\left(t\right)\exp\left\{\gamma_1^{(1)}X_{1i}+\gamma_2^{(1)}X_{2i}+v_i\right\} \tag{11}$$

$$\lambda^{(2)}\left(t;X_{1i},X_{2i},U_i,\gamma^{(2)},\nu^{(2)}\right)=\lambda_0^{(2)}\left(t\right)\exp\left\{\gamma_1^{(2)}X_{1i}+\gamma_2^{(2)}X_{2i}+\nu^{(2)}v_i\right\} \tag{12}$$

where X_{1i} was shared with the longitudinal model and covariate X_{2i} was simulated from a normal distribution with mean 0 and variance 0.1. We used constant baseline hazards of 0.1 and 0.2 for risk 1 and risk 2, respectively, to generate the event time data. The latent variables were generated from the following bivariate normal distribution:

$$\begin{pmatrix} U_i \\ v_i \end{pmatrix} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix},\begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\right)$$

The censoring time was generated from an exponential distribution with a constant hazard rate of 0.1. The parameters are given in Table III. With this setup, the rate of risk 1 is approximately 0.43, the rate of risk 2 is 0.38, and censoring rate is 0.19. Longitudinal responses are missing after the observed or censored event times. The average number of total longitudinal observations is 3.7 per subject.

Vague priors were used in the Bayesian model. We used $N(0,10^{10})$ priors for β_0, β_1, β_2, $\gamma_r^{(k)}$, $\nu^{(2)}$, θ, $IG(10^{-3},10^{-3})$ for σ^2, σ_u^2, σ_e^2, and $\Gamma(10^{-2}, 10^{-2})$ for λ_{0s}. The Jeffreys' [26] priors give similar results and are not shown here. The MCMC sampling in all simulation studies was run using 5 000 iterations, and the estimation results were based on the last 4000 iterations.

For this simulation, sample sizes of 100, 200 and 500 were considered. Table III shows that except for the frailty parameters in the survival submodel, both the Bayesian model with noninformative priors and the likelihood method have small bias and good coverage probabilities in almost all cases. For the frailty parameters, larger samples are needed to get good estimate.

We did another simulation to compare our joint analysis with a separate analysis and a joint analysis that ignores the dependent censoring (i.e. treats the dependent censoring as noninformative censoring). The data were generated same as (10)–(12). We analyzed 200 Monte Carlo data replications with the joint analysis incorporating-dependent censoring (i.e. our joint model treating-dependent censoring as a competing risk), the joint analysis

ignoring-dependent censoring and the separate analysis. We compared bias, coverage rate of the 95 per cent intervals and standard deviations of the posterior medians in Table IV. For the separate analysis, a random slope linear mixed model (10) was fitted to the longitudinal outcome and a cause-specific competing risks model (11)–(12) was fitted to the two types of failures separately. In the joint analysis ignoring-dependent censoring, only (10) and (11) were included, with the second type of risk treated as independent censoring. Sample sizes of 200 and 500 were considered in this simulation study.

We notice from Table IV that the joint analysis incorporating-dependent censoring gives nearly unbiased estimates for all the parameters, and the coverage rates for the 95 per cent confidence intervals are close to the nominal value. In the separate analysis, the time trend β_2 (the time trend for the control group) and the variance of the random slope σ_u^2 are underestimated, even for the larger sample size of 500. This is a consequence of having nonignorable missing data induced by two types of failures. Because the coefficients σ_{uv}, $\nu^{(2)}$ are positive, a higher risk of failures is expected for those subjects with greater ascending rates in the longitudinal outcome. This cannot be handled properly by the separate linear mixed effects model. Moreover, the joint analysis incorporating-dependent censoring as a competing risk provides much more accurate estimates than treating it as noninformative censoring. In addition, the standard deviations of the posterior estimates in the separate analysis for the survival data are larger than those in the joint analysis. The accuracy for estimating the frailty and the efficiency for estimating the other parameters in the survival endpoint are improved by incorporating the longitudinal outcome in our proposed joint model. In the joint analysis ignoring the dependent dropout, β_2 (the time trend for the control group) is also biased (underestimated). Because the nonignorable missing values are due to both failure types, considering one failure type may only partially correct this problem. Moreover, the variance of the latent variable in the survival sub-model is overestimated without considering the second type of failure. This shows that the latent variable might not be correctly estimated if dependent censoring is treated as noninformative censoring.

A third simulation was conducted to demonstrate that appropriate informative priors could improve estimation for the parameters, especially for smaller samples. Sample sizes of 50 and 100 were considered. We considered two methods. The first method used the same noninformative priors as in the previous simulations. The second method used informative priors for some parameters, with the prior means set as the true parameter values. We used $N(1,0.04)$ for β_2, $N(0.8,0.25)$ for $\gamma_2^{(1)}$, $N(1.2,1)$ for $\nu^{(2)}$, $N(0.8,4)$ for θ, $IG(5.56,1)$ for σ_e^2 and the same priors for the other parameters as the previous simulation. We used $N(0,4)$ for β_2, $N(0,25)$ for $\gamma_2^{(1)}$, $N(0,4)$ for $\nu^{(2)}$, $N(0,4)$ for θ, $IG(1,1)$ for σ_e^2 and the same priors for the other parameters as the previous simulation. The results are summarized in Table V.

It is seen that for the relatively small sample sizes of 50 and 100, some parameters such as β_2, σ_v^2, σ_{uv}, $\gamma_r^{(k)}$, and $\nu^{(2)}$ are not well estimated when using noninformative priors. However, with good informative priors, the estimation accuracy is greatly improved. Finally, we note that informative priors should be used with caution since poor prior information may induce additional bias.

Finally, we conducted a simulation using a setup similar to the SLS. The longitudinal response and the competing risks event times were simulated from models (7)–(9) with Z_{ij} as $([Time_{ij}])$, in which the covariates were generated from distributions similar to those in the SLS. All the parameters for generating the data were set close to the estimated values from the joint analysis for the SLS listed in Table II. Weibull baseline hazard functions were used to produce similar risk rates to those in the SLS since we observed increasing estimated

baseline hazard trends for both failure types. The results from the joint analysis and the separate analysis are compared in Table VI using 200 simulated data sets with m equal to 140. We used a piecewise exponential baseline hazard function with three knots for modeling the sample size of 140. The time points defining the intervals are taken to be 33th, 67th, and 100th percentiles of the observed failure times.

Table VI shows that joint analysis incorporating-dependent censoring gives good point estimates and coverage rates for the parameters in the longitudinal submodel. The separate analysis produces comparable estimates for most parameters except a biased time trend β_8 and an underestimated variance σ_u^2, which confirms the analysis in Section 4. These biases do not decrease even for the larger sample size of 500 (simulation results are not reported here). The separate competing risks model estimated the random effect parameters $(v^{(2)}, \sigma_v^2)$ and some others $(\gamma_2^{(1)}, \gamma_3^{(2)}, \gamma_6^{(2)}, \gamma_7^{(2)})$ poorly. With a small sample size of 140, low event rates (12 per cent for risk 1 and 28 per cent for risk 2) and small number of intervals for the piecewise exponential baseline hazard function, even joint analysis incorporating-dependent censoring may not estimate the random effects v_i well, resulting in poor parameter estimates for the survival data $(\sigma_v^2, \gamma_2^{(1)})$. This can be improved when the sample size increases to 500 and the number of knots increases to 8 (simulation results are not reported here). With more data and more number of knots to allow a better approximation to the true baseline hazard function, the frailty at the survival endpoint is better estimated, along with the other parameters modeling the survival data.

6. DISCUSSION

We proposed a Bayesian method for joint modeling of longitudinal and competing risks survival data. The joint model provides a useful method for analysis of the longitudinal data with nonignorable missing data induced by multiple types of events. The competing risks submodel enables one to handle dependent censoring for the key event by treating it as a competing risk. It also allows simultaneous inference on both longitudinal and survival endpoints. Moreover, by combining the information from the longitudinal data, the joint model improves the accuracy and efficiency of the parameter estimates for the survival endpoint compared with a separate survival analysis. The developed Bayesian method has some appealing features. Computationally, it can easily handle high-dimensional random effects, which is difficult under the frequentist framework (e.g. [15, 27]). In addition, standard errors and confidence intervals can be conveniently obtained from the posterior samples. It also allows incorporation of prior information.

Our method requires a specification of the piecewise baseline hazard in the competing risks submodel (2). In practice, the intervals for the baseline hazard can be defined based on the quantiles of the observed event times. Our simulations with the data generated with the Weibull baseline hazard demonstrated this works well. We also studied the appropriate number of intervals for the function with simulation studies. We observed that an increase in number might decrease the estimation bias for the parameters in the survival submodel with the data generated with a Weibull (nonconstant) baseline hazard rate. We also observed inflated standard errors for these parameters with the data generated with an exponential (constant) baseline hazard rate. The number of intervals has a small effect on the inference for the parameters modeling the longitudinal measurements. In practice, we suggest to first fit the data with a large number of pieces, and then combine the adjacent intervals with similar baseline estimates.

Our model can be extended to handle clustered data. Clustered data arise frequently from multi-site clinical trials in which each site can be viewed as a cluster, from studies with

families data, or from studies with recurrent events for each subject. A simple method is to add an additional random effect in our joint model to adjust for heterogeneity across the clusters. Our computational MCMC algorithm can be adopted without much modification.

We assume multivariate normal distribution of random effects in the joint model. Robustness against the departure from the normality assumption has been studied by several authors [19, 28, 29]. These authors have reported similar findings that inference seems generally robust against the normality assumption for the random effects. We also have done some simulation, which is not reported here, and confirm their findings with symmetric distributions. On the other hand, if the underlying distribution is asymmetric, such as exponential, we also observe biased estimation for those parameters directly related to the random effects. To allow more flexibility and robustness for random effects, Brown and Ibrahim [11] proposed a semiparametric Bayesian hierarchical joint model with a single failure type and it can be extended to our situation.

Our limited experience suggests that our method is pretty robust with respect to the normality assumption of measurement errors, even when underlying error distribution is asymmetric, such as exponential or unbalanced normal mixtures. However, our method is observed to be sensitive to outliers. In a sequel, we will develop a robust joint model for longitudinal and survival data.

Acknowledgments

APPENDIX: FULL CONDITIONAL DENSITIES

We used Gibbs sampling to sample from the joint posterior distribution of the parameters: $(U_i, \beta, \sigma^2, \Sigma_u, \gamma_r^{(k)}, v^{(k)}, \lambda_{0s}^{(k)}, e_i, \sigma_e^2)$. We use the notation [.] and [.|.] to denote marginal and conditional densities, respectively.

- Sample $[\sigma^2|.] \propto \text{InvGamma}\,(\alpha_{\sigma^2}, \beta_{\sigma^2})[\sigma^2]$, where $\alpha_{\sigma^2} = \sum_{i=1}^m n_i/2 - 1$, and $\beta_{\sigma^2} = \sum_{i=1}^m \left(Y_i - X_i^{(1)}\beta - Z_i U_i\right)^{\mathrm{T}} \left(Y_i - X_i^{(1)}\beta - Z_i U_i\right)/2$.

- Sample $[\beta|.] \propto \text{N}((X^{\mathrm{T}}X)^{-1}X(Y - ZU), (X^{\mathrm{T}}X)^{-1}\sigma^2)[\beta]$.

- Sample $[\Sigma_u|.] \propto \text{InvWish}(m - Q - 1, S^{-1})[\Sigma_u]$ where $S = \sum_{i=1}^m U_i U_i^{\mathrm{T}}$.

- Sample $[U_i|.] \propto \text{N}$
  $\left(\mu_{U_i|Y_i}, \Sigma_{U_i|Y_i}\right) * \prod_{k=1}^K \exp\left\{\left(\theta' U_i + e_i\right) v^{(k)} I\,(d_i = k) - H^{(k)}\,(t_i)\right\}, i = 1, \ldots, m, i = 1, \ldots, m$
  where $\mu_{U_i|Y_i} = \left(Z_i^{\mathrm{T}} Z_i + \Sigma_u^{-1}\sigma^2\right)^{-1} Z_i^{\mathrm{T}}\,(Y_i - X_i\beta)$ and $\Sigma_{U_i|Y_i} = \left(Z_i^{\mathrm{T}} Z_i/\sigma^2 + \Sigma_u^{-1}\right)^{-1}$ and $H^{(k)}\,(t_i)$ is defined as (6), MH.

- Sample $\left[\lambda_{01}^{(k)}, \cdots, \lambda_{0S^{(k)}}^{(k)}|.\right] \propto \prod_{s=1}^{S^{(k)}} \Gamma\left(\delta_s^k, \beta_{\lambda_{0s}^{(k)}}\right)\left[\lambda_{01}^{(k)}, \cdots, \lambda_{0S^{(k)}}^{(k)}\right], k = 1, \ldots, K$ where $\delta_s^k = \sum_{i=1}^m I\left(d_i = k, t_{s-1}^{(k)} < t_i \le t_s^{(k)}\right) + 1$ and

$$\beta_{\lambda_{0s}^{(k)}} = \sum_{i=1}^{m} \exp\left\{X_i^{(21)}\gamma_1^{(k)} + \left(\theta' U_i + e_i\right)\nu^{(k)}\right\} I\left(t_i > t_{s-1}^{(k)}\right) \int_{t_{s-1}^{(k)}}^{\min\left(t_i, t_s^{(k)}\right)} \exp\left\{X_i^{(22)}(t)\gamma_2^{(k)}\right\} \mathrm{d}t.$$

Here δ_s^k is the number of events occurring in the time interval $\left(t_{s-1}^{(k)}, t_s^{(k)}\right]\left(s=1,\ldots,S^{(k)}\right)$ for failure type k.

- Sample $\left[\gamma_r^{(k)}|.\right] \propto \exp\left\{\gamma_r^{(k)} \sum_{i=1}^{m} I(d_i=k) X_{ir}^{(2)}(t_i)\right\}\left(\prod_{i=1}^{m} S^{(k)}(t_i)\right)\left[\gamma_r^{(k)}\right]$, $r=1,\ldots,R$

  $k=1,\ldots K$ where $S^{(k)}(t_i) = \exp\left\{-\int_0^{t_i}\lambda_i^{(k)}(t)\,\mathrm{d}t\right\} = \exp\left\{-H^{(k)}(t_i)\right\}$, MH.

- Sample $\left[\nu^{(k)}|.\right] \propto \exp\left\{\nu^{(k)} \sum_{i=1}^{m} I(d_i=k)\left(\theta' U_i + e_i\right)\right\}\left(\prod_{i=1}^{m} S^{(k)}(t_i)\right)\left[\nu^{(k)}\right]$, $k=2,\ldots,K$

  where $S^{(k)}(t_i) = \exp\left\{-\int_0^{t_i}\lambda_i^{(k)}(t)\,\mathrm{d}t\right\} = \exp\left\{-H^{(k)}(t_i)\right\}$, ARS.

- Sample $\left[\theta_q|.\right] \propto \prod_{k=1}^{K} \exp\left\{\theta_q \nu^{(k)} \sum_{i=1}^{m} I(d_i=k) U_{iq}\right\}\left(\prod_{i=1}^{m} S^{(k)}(t_i)\right)\left[\theta_q\right]$, $q=1,\ldots,Q$

  where $S^{(k)}(t_i) = \exp\left\{-\int_0^{t_i}\lambda_i^{(k)}(t)\,\mathrm{d}t\right\} = \exp\left\{-H^{(k)}(t_i)\right\}$, ARS.

- Sample $[e_i|.] \propto N\left(0, \sigma_e^2\right) * \prod_{k=1}^{K} \exp\left\{e_i \nu^{(k)} I(d_i=k) - H^{(k)}(t_i)\right\}$, $i=1,\ldots,m$ where $H^{(k)}(t_i)$ is defined as (6), MH.

- Sample $\left[\sigma_e^2|.\right] \propto \mathrm{InvGamma}(\alpha_{\sigma_e^2}, \beta_{\sigma_e^2})\left[\sigma_e^2\right]$, where $\alpha_{\sigma_e^2} = m/2 - 1$, and $\beta_{\sigma_e^2} = \sum_{i=1}^{m} e_i^2/2$.

REFERENCES

1. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, Arriola E, Silver R, Strange C, Bolster M, Seibold JR, Riley DJ, Hsu VM, Varga J, Schraufnagel DE, Theodore A, Simms R, Wise R, Wigley F, White B, Steen V, Read C, Mayes M, Parsley Ed, Mubarak K, Connolly MK, Golden J, Olman M, Fessler B, Rothfield N, Metersky M, The Scleroderma Lung Study Research Group. Cyclophosphamide versus placebo in scleroderma lung disease. The New England Journal of Medicine. 2006; 354:2655–2666. [PubMed: 16790698]

2. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38:963–974. [PubMed: 7168798]

3. Cox DR. Regression models and life tables. Journal of the Royal Statistical Society, Series B. 1972; 34:187–200.

4. Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. Biometrics. 1988; 44:175–188.

5. Schluchter MD. Methods for the analysis of informatively censored longitudinal data. Statistics in Medicine. 1992; 11:1861–1870. [PubMed: 1480878]

6. Little RJA. Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association. 1993; 88:125–134.

7. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. Journal of the American Statistical Association. 1995; 90:1112–1121.

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

9. Faucett CJ, Thomas DC. Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Statistics in Medicine. 1996; 15:1663–1685. [PubMed: 8858789]

10. Wang Y, Taylor JMG. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. Journal of the American Statistical Association. 2001; 96:895–905.

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

12. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica. 2004; 14:809–834.

13. Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. Biostatistics. 2000; 4:456–480.

14. Zeng D, Cai J. Simultaneous modelling of survival and longitudinal data with an application to repeated quality of life measures. Lifetime Data Analysis. 2005; 11:151–174. [PubMed: 15940822]

15. Elashoff RM, Li G, Li N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. Statistics in Medicine. 2007; 26:2813–2835. [PubMed: 17124698]

16. Elashoff RM, Li G, Li N. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. Biometrics. 2008; 64:762–771. [PubMed: 18162112]

17. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. Biometrics. 1978; 34:541–554. [PubMed: 373811]

18. Ng SK, McLachlan GJ. An EM-based semi-parametric mixture model approach to the regression analysis of competing-riss data. Statistics in Medicine. 2003; 22:1011–1097.

19. Hsieh F, Tseng YK, Wang JL. Joint modeling of survival and longitudinal data: likelihood approach revisited. Biometrics. 2006; 62:1037–1043. [PubMed: 17156277]

20. Ibrahim, JG.; Chen, M.; Sinha, D. Bayesian Survival Analysis. Springer; New York: 2001.

21. Geman S, Geman D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984; 6:721–741.

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

23. Chib S, Greenberg E. Understanding the Metropolis–Hastings algorithm. Biometrika. 1995; 49:327–335.

24. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. Applied Statistics. 1992; 41:337–348.

25. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit (with Discussion). Journal of the Royal Statistical Society, Series B. 2002; 64(4):583–640.

26. Jeffreys, H. Theory of Probability. Oxford University Press; Clarendon Press; Oxford: 1961.

27. Ratcliffe S, Guo W, Have TRT. Joint modeling of longitudinal and survival data via a common frailty. Biometrika. 2004; 60:892–899.

28. 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]

29. Zeng D, Cai J. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. The Annals of Statistics. 2005; 33:2132–2163.
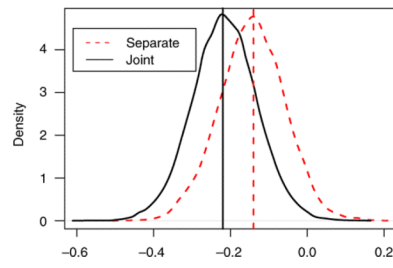
Figure 1.
Posterior densities of time trend for the placebo group in the SLS study from the separate models and the proposed joint model. The vertical lines represent the posterior medians. The estimates and CIs for the two methods are: −0.14(−0.31 0.03), −0.22(−0.38−0.05), respectively.
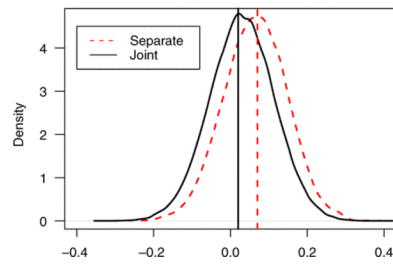
Figure 2.
Posterior densities of time trend for the CYC group in the SLS study from the separate models and the proposed joint model. The vertical lines represent the posterior medians. The estimates and CIs for the two methods are: 0.07(−0.10, 0.22), 0.03(−0.14, 0.19), respectively.

Table I

Results for the different models examined with specification as (7)–(9).

Z_{ij}	# of RE	Parameter	Latent dependence		DIC
			Inference		
[1]	2	$\sigma_{u_1 u}$	$-12.40\ (-25.38, -5.41)$		5777.1
		$\nu^{(2)}$	$0.24\ (0.01, 0.73)$		
[time$_{ij}$]	2	$\sigma_{u_1 u}$	$-1.18\ (-2.16, -0.62)$		4921.1
		$\nu^{(2)}$	$0.21\ (0.01, 0.52)$		
[1,time$_{ij}$]	3	$\sigma_{u_1 u}$	$-0.84\ (-7.20, 5.34)$		5529.8
		$\sigma_{u_2 u}$	$-1.36\ (-3.03, -0.58)$		
		$\nu^{(2)}$	$0.27\ (0.04, 0.64)$		
[1,time$_{ij}$,CYC$_i$]	4	$\sigma_{u_1 u}$	$-0.65\ (-8.21, 4.43)$		5821.6
		$\sigma_{u_2 u}$	$-1.49\ (-3.09, -0.64)$		
		$\sigma_{u_1 u}$	$0.67\ (-7.46, 10.46)$		
		$\nu^{(2)}$	$0.25\ (0.04, 0.69)$		

Table II

Results from analyzing the SLS data using a random slope model. Entries in boldface indicate different results from the joint and separate analysis.

Factor	Separate analysis Estimate (95 per cent CI)	Joint analysis Estimate (95 per cent CI)
Longitudinal outcome %FVC		
Intercept (β_0)	66.83 (65.40 68.21)	67.18 (65.87 68.52)
Age (β_1)	−0.03 (−0.12 0.06)	−0.03 (−0.11 0.06)
CYC group (β_2)	−0.82 (−2.75 1.10)	−0.99 (−2.88 0.90)
Fibrosis (β_3)	−1.67 (−2.74 −0.62)	−1.67 (−2.76 −0.58)
Baseline %FVC (β_4)	0.91 (0.82 1.00)	0.90 (0.81 1.00)
Age × CYC group (β_5)	0.07 (−0.07 0.21)	0.07 (−0.07 0.22)
Fibrosis × CYC group (β_6)	1.36 (−0.19 2.94)	1.30 (−0.25 2.89)
Baseline %FVC × CYC group (β_7)	0.11 (−0.03 0.25)	0.11 (−0.03 0.25)
Time(month) (β_8)	**−0.14 (−0.31 0.03)**	**−0.22 (−0.38 −0.05)**
Time (month) × CYC group (β_9)	**0.21 (−0.03 0.44)**	**0.25 (0.01 0.48)**
σ^2	20.22 (17.91 22.86)	19.79 (17.56 22.38)
σ_u^2	**0.26 (0.20 0.36)**	**0.30 (0.22 0.41)**
Time to treatment failure or death		
Age ($\gamma_1^{(1)}$)	0.09 (−0.01 0.45)	0.10 (0.02 0.20)
CYC group ($\gamma_2^{(1)}$)	−1.30 (−6.13 1.63)	−1.33 (−3.61 0.61)
Fibrosis ($\gamma_3^{(1)}$)	−0.14 (−2.49 1.34)	−0.45 (−1.77 0.68)
Baseline %FVC ($\gamma_4^{(1)}$)	0.10 (−0.01 0.55)	0.03 (−0.06 0.13)
Age × CYC group ($\gamma_5^{(1)}$)	−0.10 (−0.60 0.10)	−0.13 (−0.30 0.01)
Fibrosis × CYC group ($\gamma_6^{(1)}$)	−0.70 (−4.38 2.42)	0.12 (−1.89 2.25)
Baseline %FVC × CYC group ($\gamma_7^{(1)}$)	−0.17 (−0.84 0.02)	−0.05 (−0.21 0.12)
Time to disease-related dropout		
Age ($\gamma_1^{(2)}$)	0.04 (0.00 0.10)	0.05 (0.01 0.09)
CYC group ($\gamma_2^{(2)}$)	0.44 (−0.43 1.41)	0.40 (−0.41 1.28)
Fibrosis ($\gamma_3^{(2)}$)	0.09 (−0.49 0.68)	0.07 (−0.44 0.58)
Baseline %FVC ($\gamma_4^{(2)}$)	−0.07 (−0.14 −0.02)	−0.06 (−0.12 −0.02)
Age × CYC group ($\gamma_5^{(2)}$)	−0.01 (−0.09 0.06)	−0.03 (−0.09 0.03)
Fibrosis × CYC group ($\gamma_6^{(2)}$)	0.11 (−0.67 0.91)	0.18 (−0.55 0.91)
Baseline %FVC × CYC group ($\gamma_7^{(2)}$)	0.10 (0.03 0.19)	0.09 (0.02 0.16)

Factor	Separate analysis Estimate (95 per cent CI)	Joint analysis Estimate (95 per cent CI)
Random effects for survival endpoint		
$v^{(2)}$	−0.14 (−5.18 4.52)	0.21 (0.01, 0.52)
σ_v^2	13.24 (0.00 142.91)	5.07 (1.52, 18.91)
Covariance of u_i and v_i		
σ_{uv}	—	−1.18 (−2.16, −0.62)

Table III

Bias, coverage rate of the 95 per cent interval, standard deviation (SD), based on the likelihood method and the Bayesian model with noninformative priors.

n	Parameter	Likelihood method			Bayesian method		
		Bias	Coverage	SD	Bias	Coverage	SD
100	$\beta_0 = 10$	0.001	0.965	0.055	−0.003	0.960	0.058
	$\beta_1 = -1.5$	0.001	0.955	0.072	0.008	0.930	0.078
	$\beta_2 = 1$	−0.027	0.960	0.105	−0.039	0.895	0.137
	$\sigma^2 = 0.25$	0.001	0.965	0.020	0.001	0.965	0.020
	$\sigma_u^2 = 0.5$	−0.020	0.910	0.110	0.006	0.975	0.117
	$\sigma_{uv} = 0.4$	−0.049	0.915	0.286	−0.048	0.870	0.266
	$\sigma_v^2 = 0.5$	0.219	0.990	0.822	0.110	0.920	0.619
	$\gamma_1^{(1)} = -1$	−0.059	0.960	0.406	0.033	0.945	0.374
	$\gamma_2^{(1)} = 0.8$	0.066	0.970	0.604	0.078	0.945	0.626
	$\gamma^{(2)} = -1.5$	−0.051	0.965	0.504	−0.088	0.960	0.434
	$\gamma_2^{(2)} = 0.5$	−0.016	0.960	0.680	0.064	0.945	0.740
	$v^{(2)} = 1.2$	0.162	0.930	1.113	0.056	0.890	1.570
200	$\beta_0 = 10$	0.002	0.970	0.036	0.002	0.945	0.044
	$\beta_1 = -1.5$	−0.002	0.955	0.052	0.002	0.980	0.051
	$\beta_2 = 1$	−0.019	0.955	0.095	−0.009	0.955	0.087
	$\sigma^2 = 0.25$	>0.001	0.965	0.014	0.001	0.945	0.014
	$\sigma_u^2 = 0.5$	−0.007	0.940	0.078	0.014	0.960	0.084
	$\sigma_{uv} = 0.4$	−0.013	0.980	0.181	0.002	0.950	0.142
	$\sigma_v^2 = 0.5$	0.080	0.980	0.352	0.034	0.975	0.314
	$\gamma_1^{(1)} = -1$	−0.029	0.975	0.266	0.007	0.960	0.245

n	Parameter	Likelihood method			Bayesian method		
		Bias	Coverage	SD	Bias	Coverage	SD
	$\gamma_2^{(1)} = 0.8$	0.020	0.965	0.384	0.070	0.925	0.422
	$\gamma_1^{(2)} = -1.5$	−0.075	0.955	0.395	−0.018	0.950	0.278
	$\gamma_2^{(2)} = 0.5$	0.013	0.940	0.514	0.022	0.935	0.425
	$\nu^{(2)} = 1.2$	0.118	0.930	0.779	0.051	0.980	0.435
500	$\beta_0 = 10$	−0.002	0.965	0.027	0.003	0.960	0.026
	$\beta_1 = -1.5$	0.002	0.965	0.038	−0.003	0.970	0.033
	$\beta_2 = 1$	−0.014	0.965	0.077	0.008	0.950	0.057
	$\sigma^2 = 0.25$	>0.001	0.965	0.010	>0.001	0.960	0.009
	$\sigma_u^2 = 0.5$	−0.007	0.910	0.057	0.013	0.935	0.057
	$\sigma_{uv} = 0.4$	0.007	0.965	0.118	0.019	0.965	0.090
	$\sigma_v^2 = 0.5$	0.067	0.930	0.228	0.024	0.945	0.164
	$\gamma_1^{(1)} = -1$	−0.021	0.950	0.162	0.011	0.940	0.156
	$\gamma_2^{(1)} = 0.8$	0.023	0.980	0.229	0.014	0.965	0.233
	$\gamma_1^{(2)} = -1.5$	−0.030	0.950	0.231	<0.001	0.930	0.162
	$\gamma_2^{(2)} = 0.5$	0.011	0.955	0.289	0.037	0.945	0.259
	$\nu^{(2)} = 1.2$	−0.003	0.925	0.388	−0.008	0.955	0.227

Table IV

Bias, coverage rate of the 95 per cent interval, standard deviation (SD), for the separate analysis, the joint analysis ignoring-dependent censoring and the joint analysis incorporating-dependent censoring.

n	Parameter	Separate analysis			Joint analysis ignoring dependent censoring			Joint analysis incorporating dependent censoring		
		Bias	Coverage	SD	Bias	Coverage	SD	Bias	Coverage	SD
200	β_0	0.003	0.935	0.043	<0.001	0.9450	0.043	0.002	0.945	0.044
	$\beta_1 = -1.5$	0.025	0.945	0.052	0.017	0.965	0.052	0.002	0.980	0.051
	$\beta_2 = 1$	**−0.203**	**0.210**	**0.070**	**−0.100**	**0.780**	0.077	−0.009	0.955	0.087
	$\sigma^2 = 0.25$	0.001	0.960	0.013	0.001	0.960	0.013	0.001	0.945	0.014
	$\sigma_u^2 = 0.5$	**−0.049**	0.935	0.071	−0.011	0.955	0.0140	0.014	0.960	0.084
	$\sigma_{uv} = 0.4$				0.023	0.925	0.169	0.002	0.950	0.142
	$\sigma_v^2 = 0.5$	−0.034	0.845	**0.690**	**0.236**	**0.875**	0.400	0.034	0.975	0.314
	$\gamma_1^{(1)} = -1$	0.073	0.950	**0.256**	0.034	0.955	**0.257**	0.007	0.960	0.245
	$\gamma_2^{(1)} = 0.8$	0.027	0.950	**0.442**	0.061	0.955	0.413	0.070	0.925	0.422
	$\gamma_1^{(2)} = -1.5$	−0.092	0.930	**0.338**	—	—	—	−0.018	0.950	0.278
	$\gamma_2^{(2)} = 0.5$	−0.015	0.935	**0.507**	—	—	—	0.022	0.935	0.425
	$\nu^{(2)} = 1.2$	**0.145**	0.935	**4.918**	—	—	—	0.051	0.980	0.435
500	$\beta_0 = 10$	0.005	0.950	0.025	0.002	0.975	0.025	0.003	0.960	0.026
	$\beta_1 = -1.5$	0.019	0.9500	0.033	0.011	0.950	0.033	−0.003	0.970	0.033
	$\beta_2 = 1$	**−0.203**	**0.005**	**0.044**	**−0.093**	**0.560**	0.053	0.008	0.950	0.057
	$\sigma^2 = 0.25$	<0.001	0.970	0.008	<0.001	0.970	0.008	<0.001	0.960	0.009
	$\sigma_u^2 = 0.5$	**−0.050**	**0.810**	0.048	−0.013	0.920	0.055	0.013	0.935	0.057
	$\sigma_{uv} = 0.4$				0.022	0.915	0.107	0.019	0.965	0.090
	$\sigma_v^2 = 0.5$	−0.029	0.935	**0.220**	**0.200**	**0.810**	0.275	0.024	0.945	0.164
	$\gamma_1^{(1)} = -1$	0.012	0.920	**0.163**	0.033	0.925	**0.164**	0.011	0.940	0.156

n	Parameter	Separate analysis			Joint analysis ignoring dependent censoring			Joint analysis incorporating dependent censoring		
		Bias	Coverage	SD	Bias	Coverage	SD	Bias	Coverage	SD
	$\gamma_2^{(1)} = 0.8$	0.001	0.950	**0.250**	0.043	0.950	**0.250**	0.014	0.965	0.233
	$\gamma_1^{(2)} = -1.5$	−0.011	0.955	**0.173**	—	—	—	<0.001	0.930	0.162
	$\gamma_2^{(2)} = 0.5$	0.034	0.935	**0.284**	—	—	—	0.037	0.945	0.259
	$v^{(2)} = 1.2$	**0.108**	0.940	**1.106**	—	—	—	−0.008	0.955	0.227

Entries in boldface indicate different results from the three analysis methods.

Table V

Bias, coverage rate of the 95 per cent interval, standard deviation (SD), for the noninformative and informative priors.

n	Parameter	Noninformative priors			Informative prior		
		Bias	Coverage	SD	Bias	Coverage	SD
50	$\beta_0 = 10$	−0.002	0.945	0.083	−0.001	0.940	0.084
	$\beta_1 = -1.5$	0.024	0.960	0.112	0.007	0.960	0.112
	$\beta_2 = 1$	**− 0.114**	**0.890**	**0.169**	**− 0.036**	**0.980**	**0.094**
	$\sigma^2 = 0.25$	0.006	0.955	0.028	0.004	0.955	0.027
	$\sigma_u^2 = 0.5$	0.005	0.945	0.176	−0.003	0.970	0.148
	$\sigma_{uv} = 0.4$	**− 0.155**	**0.900**	**0.355**	**− 0.001**	**0.975**	**0.230**
	$\sigma_v^2 = 0.5$	**0.435**	**0.970**	**1.231**	**0.117**	**0.985**	**0.365**
	$\gamma_1^{(1)} = -1$	0.090	0.910	0.635	0.039	0.950	0.517
	$\gamma_2^{(1)} = 0.8$	**0.244**	**0.935**	**1.005**	**0.050**	**1.000**	**0.226**
	$\gamma_1^{(2)} = -1.5$	−0.379	0.905	2.328	−0.340	0.925	2.680
	$\gamma_2^{(2)} = 0.5$	0.062	0.955	1.202	0.092	0.930	1.096
	$v^{(2)} = 1.2$	**− 1.071**	**0.925**	**1.753**	**− 0.078**	**0.980**	**0.493**
100	$\beta_0 = 10$	−0.003	0.960	0.058	−0.005	0.950	0.057
	$\beta_1 = -1.5$	0.008	0.930	0.078	0.012	0.940	0.076
	$\beta_2 = 1$	**− 0.039**	**0.895**	**0.137**	**− 0.027**	**0.965**	**0.084**
	$\sigma^2 = 0.25$	0.001	0.965	0.020	0.001	0.945	0.020
	$\sigma_u^2 = 0.5$	0.006	0.975	0.117	−0.006	0.980	0.103
	$\sigma_w = 0.4$	**− 0.048**	**0.870**	**0.266**	**− 0.001**	**0.975**	**0.165**
	$\sigma_v^2 = 0.5$	**0.110**	**0.920**	**0.619**	**0.070**	**0.970**	**0.277**
	$\gamma_1^{(1)} = -1$	0.033	0.945	0.374	−0.006	0.955	0.342

n	Parameter	Noninformative priors			Informative prior		
		Bias	Coverage	SD	Bias	Coverage	SD
	$\gamma_2^{(1)} = 0.8$	**0.078**	**0.945**	**0.626**	**0.019**	**0.990**	**0.252**
	$\gamma_1^{(2)} = -1.5$	−0.088	0.960	0.434	−0.052	0.960	0.428
	$\gamma_2^{(2)} = 0.5$	0.064	0.945	0.740	0.046	0.955	0.668
	$v^{(2)}=1.2$	**0.056**	**0.890**	**1.570**	**0.034**	**0.985**	**0.394**

Entries in boldface indicate different results from the two analysis methods.

Table VI

Simulation results for SLS data (n=140).

Factor	Separate analysis			Joint analysis incorporating dependent censoring		
	Estimate	Coverage	SD	Estimate	Coverage	SD
β_0=67	66.911	0.930	0.450	67.022	0.970	0.414
β_1=0	0.008	0.965	0.029	0.001	0.930	0.032
β_2=−1	−1.016	0.920	0.642	−1.074	0.940	0.599
β_3=−1.7	−1.754	0.940	0.362	−1.715	0.940	0.360
β_4=1	0.997	0.955	0.032	1.000	0.960	0.033
β_5=0	−0.009	0.920	0.047	−0.003	0.925	0.048
β_6=1.4	1.469	0.960	0.499	1.407	0.935	0.512
β_7=0.1	0.098	0.925	0.048	0.100	0.935	0.048
β_8=−0.25	**− 0.126**	**0.710**	0.077	−0.249	0.960	0.096
β_9=0.25	0.230	0.950	0.118	0.266	0.955	0.123
σ^2=20	20.145	0.975	1.141	19.946	0.985	1.146
$\sigma_u^2 = 0.3$	**0.254**	**0.860**	0.042	0.299	0.945	0.053
$\gamma_1^{(1)} = 0.1$	0.098	0.875	0.064	0.118	0.960	0.052
$\gamma_2^{(1)} = -1.2$	−2.901	0.885	12.066	−1.932	0.875	1.447
$\gamma_3^{(1)} = -0.4$	−0.538	0.900	1.505	−0.513	0.910	0.603
$\gamma_4^{(1)} = 0$	0.001	0.935	0.115	0.005	0.935	0.051
$\gamma_5^{(1)} = -0.1$	−0.114	0.880	0.289	−0.114	0.925	0.092
$\gamma_6^{(1)} = 0.1$	0.107	0.915	4.169	0.048	0.925	1.079
$\gamma_7^{(1)} = -0.05$	−0.056	0.915	0.334	−0.070	0.945	0.095
$\gamma_1^{(2)} = 0.05$	0.071	0.940	0.131	0.051	0.950	0.026
$\gamma_2^{(2)} = 0.4$	0.415	0.920	1.690	0.446	0.930	0.462

Factor	Separate analysis			Joint analysis incorporating dependent censoring		
	Estimate	Coverage	SD	Estimate	Coverage	SD
$\gamma_3^{(2)} = 0.06$	0.305	0.925	1.405	0.107	0.950	0.291
$\gamma_4^{(2)} = -0.06$	−0.116	0.875	0.692	−0.046	0.870	0.028
$\gamma_5^{(2)} = -0.03$	−0.029	0.945	0.083	−0.028	0.930	0.033
$\gamma_6^{(2)} = 0.2$	0.392	0.905	1.487	0.196	0.945	0.394
$\gamma_7^{(2)} = 0.1$	0.380	0.900	2.568	0.080	0.900	0.032
$\nu^{(2)}=0.2$	1.227	0.815	28.930	0.179	0.930	0.156
$\sigma_v^2 = 5.2$	16.473	0.475	140.058	9.012	0.935	5.630
$\sigma_{uv}=-1.2$	—	—	—	−1.399	0.965	0.579

Entries in boldface indicate different results from the two analysis methods.