Published in final edited form as:

Biom J. 2009 February ; 51(1): 19–30. doi:10.1002/bimj.200810491.

Robust Joint Modeling of Longitudinal Measurements and Competing Risks Failure Time Data

Ning Li^{1,*}, Robert M. Elashoff^{1,2}, and Gang Li²

¹ Department of Biomathematics, University of California at Los Angeles, 10833 Leconte Ave, Box 951766, Los Angeles, CA 90095-1766, USA

² Department of Biostatistics, School of Public Health, University of California at Los Angeles, Los Angeles, CA 90095, USA

Summary

Existing methods for joint modeling of longitudinal measurements and survival data can be highly influenced by outliers in the longitudinal outcome. We propose a joint model for analysis of longitudinal measurements and competing risks failure time data which is robust in the presence of outlying longitudinal observations during follow-up. Our model consists of a linear mixed effects sub-model for the longitudinal outcome and a proportional cause-specific hazards frailty sub-model for the competing risks data, linked together by latent random effects. Instead of the usual normality assumption for measurement errors in the linear mixed effects sub-model, we adopt a *t*-distribution which has a longer tail and thus is more robust to outliers. We derive an EM algorithm for the maximum likelihood estimates of the parameters and estimate their standard errors using a profile likelihood method. The proposed method is evaluated by simulation studies and is applied to a scleroderma lung study.

Keywords

Cause-specific hazard; EM algorithm; Joint modeling; Longitudinal data; Non-ignorable missing data; Robust inference

1 Introduction

Joint modeling of longitudinal measurements and survival data has broad applications in biomedical studies in which we observe both a longitudinal outcome during follow-up and the occurrence of certain events, such as onset of a disease, death, discontinuation of treatment, dropout, etc. In many cases the longitudinal measurements and the event times are known to be correlated, or they share some important covariates. For example, lower longitudinal measurements on CD4 count are correlated with a higher risk for AIDS and death. In schizophrenia studies, longitudinal measurements are collected on the positive and negative syndrome scale (PANSS). Patients with higher PANSS scores indicate more severe psychopathology and therefore have a higher probability to discontinue the treatment because of low efficacy. Another example is the scleroderma lung study (Tashkin et al., 2006) in which the longitudinal outcome Forced Vital Capacity (FVC, % predicted) is correlated with treatment failure, deaths or dropouts.

*Corresponding author: e-mail: ningli@ucla.edu, Phone: 1 310 206 8928, Fax: 1 310 267 2611.

Much work has been done for joint modeling of longitudinal measurements and survival data in the last decade. Joint models have been used as a means to handle non-ignorable missing data due to informative dropout for longitudinal analysis (Schluchter, 1992; DeGruttola and Tu, 1994; Little, 1995; Hogan and Laird, 1997; Henderson, Diggle and Dobson, 2000; Elashoff, Li and Li, 2007a, b). In addition, it has been used in survival regression analysis with time-dependent covariates which are measured only intermittently (Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Wang and Taylor, 2001; Xu and Zeger, 2001; Song, Davidian and Tsiatis, 2002; Brown and Ibrahim, 2003; Tseng, Hsieh and Wang, 2005). Finally, it also enables statistical inference on both endpoints simultaneously (Henderson et al., 2000; Zeng and Cai, 2005; Elashoff et al., 2007a, b).

All the aforementioned work assumes a normal random error for the longitudinal sub-model, which is not robust against outlying observations. However, outliers have been often encountered in practice as shown by Figure 1 for the scleroderma lung study. In Sections 3 and 4 we use data from the scleroderma lung study and simulations to illustrate that the joint model with normal random errors can be easily influenced by a small portion of outliers. In this paper, we propose a joint model that is robust against outlying longitudinal measurements by assuming a t -distributed measurement error in the linear mixed effects sub-model. As shown in Section 2, with a t -distributed random error, our model implicitly reweights the observations according to their residuals when estimating the fixed effects at the longitudinal endpoint, which effectively downweights outlying observations. Compared to other robust methods such as the Huber's robust regression (Huber, 1973), the t -distribution based method is mathematically more tractable and computationally simple. We note that the t -distribution has first been used by several authors in linear regression models to guard against outliers (Sutradhar and Ali, 1986; Lange, Little and Taylor, 1989). However, to the best of our knowledge, it has not been studied for complex models such as joint models for longitudinal and survival data. Our robust joint model will have a broad application since it provides a useful tool to deal with both non-ignorable missing data and outliers, which may be present in longitudinal measurements of most biomedical studies.

The rest of the paper is organized as follows. The model and the inference procedure are described in Section 2. Section 3 contains application of the robust joint model to the scleroderma lung study. Section 4 includes simulation studies in which the robust joint model is compared to the joint model with a normal distribution for the measurement error in the presence of outlying data points in the longitudinal measurements. Section 5 gives some concluding remarks and discussions.

2 The Robust Joint Model Formulation and Estimation

2.1 The robust joint model

Our robust joint model is comprised of two linked sub-models: a linear mixed effects sub-model for the longitudinal measurements and a cause-specific hazards sub-model for the competing risks failure time data. The linkage between the two aspects is modeled through the association between random effects. Let $Y_i(t)$ be the longitudinal outcome measured at time t for subject $i = 1, 2, \dots, n$, where n is the total number of subjects in study, and $C_i = (T_i, D_i)$ denote the competing risks data on subject i , with T_i being the failure/censoring time and D_i taking value in $\{0, 1, \dots, g\}$, where $D_i = 0$ indicates a censored event and $D_i = k$ shows that subject i fails from the k th type of failure, $k = 1, \dots, g$. Throughout, the censoring mechanism is assumed to be independent of the survival time. Dependent (or informative) censoring can be treated as one of the g types of failures. To be more specific, the model is characterized in terms of the following two components:

$$\begin{cases} Y_i(t) = X_i^{(1)}(t)^T \beta + \tilde{X}_i^{(1)}(t)^T b_i + \varepsilon_i(t), \\ \lambda_k(t; X_i^{(2)}(t), u_i, \gamma_k, \nu_k) \\ = \lim_{h \rightarrow 0} \frac{P[t \leq T_i < t+h, D_i = k | T_i \geq t, X_i^{(2)}(t), u_i]}{h} \\ = \lambda_{0k}(t) \exp\{X_i^{(2)}(t)^T \gamma_k + \nu_k u_i\}. \end{cases} \quad (1) \quad (2)$$

In the linear mixed effects sub-model (1), the vectors $X_i^{(1)}(t)$ and $\tilde{X}_i^{(1)}(t)$ are associated covariates and are allowed to change over time, β represents the fixed effects of $X_i^{(1)}(t)$ and b_i is subject-specific effects of $\tilde{X}_i^{(1)}(t)$. Instead of assuming $\varepsilon_i(t) \sim N(0, \sigma^2)$, we propose $\varepsilon_i(t) \sim t(0, \sigma^2, \kappa)$, where κ represents the degrees of freedom. The normal distribution is included as a special case when $\kappa \rightarrow \infty$. For simplicity, here we assume κ is prespecified. Our limited experience suggests that $\kappa = 3$ or 4 works well in practice. In addition, we assume that $b_i \perp \varepsilon_i(t)$ and that $\varepsilon_i(t_1) \perp \varepsilon_i(t_2)$ for any $t_1 \neq t_2$. Sub-model (2) specifies the distribution of the competing risks survival data, with $\lambda_k(t; X_i^{(2)}(t), u_i, \gamma_k, \nu_k)$ being the instantaneous rate for failures of type κ at time t in the presence of all other failure types, given the vector of covariates $X_i^{(2)}(t)$ and the frailty u_i . Here $\lambda_{0k}(t)$ is a completely unspecified baseline hazard function for risk κ , where $\kappa = 1, \dots, g$. We model the cause-specific hazards of the competing risk failure times in which the correlation between the failure times are accounted for by the random effect u_i and the parameter ν . This is an extension of the cause-specific hazards model proposed by Prentice et al. (1978). We assume that b_i and u_i jointly have a multivariate normal distribution:

$$\theta_i = \begin{pmatrix} b_i \\ u_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{bb} & \Sigma_{bu}^T \\ \Sigma_{bu} & \sigma_u^2 \end{pmatrix} \right).$$

Therefore, the correlation between b_i and u_i models the *latent* association between the longitudinal measurements and competing risks data that cannot be accounted for by the

observed covariates if $X_i^{(1)}(t)$ (or $\tilde{X}_i^{(1)}(t)$) and $X_i^{(2)}(t)$ have elements in common. We can test the hypothesis $\Sigma_{bu} = 0$ for the *latent* association between the two endpoints. At last, we assume that, conditional on all the covariates and random effects, the longitudinal measurements are independent of the competing risks survival data. We denote $\gamma = (\gamma_1^T, \dots, \gamma_g^T)^T$ and $\nu = \nu_1, \dots, \nu_g$. The parameter ν_1 is set to 1 to ensure identifiability.

2.2 The likelihood and the EM algorithm

Suppose the longitudinal outcome $Y_i(t)$ is observed at time point t_{ij} for $j = 1, \dots, n_i$, and denote $Y_i = (Y_{i1}, \dots, Y_{in_i})$. Note that the set $(t_{i1}, \dots, t_{in_i})$ can be different among subjects, due to different event times and the fact that some patients may miss one or more visits. We assume that the missing values in the longitudinal measurements caused by reasons other than the g types of events are missing at random. Let $C_i = (T_i, D_i)$ denote the competing risks data on subject i . It is important to note that the joint distribution of (Y, C) is completely determined by $f(Y|\theta, \Psi)$, $f(C|\theta, \Psi)$, and $f(\theta|\Psi)$ as specified before, where $\Psi = (\beta, \sigma^2, \gamma, \nu, \Sigma, \lambda_{01}(t), \dots, \lambda_{0g}(t))$ is the vector containing all the unknown parameters from (1)–(2). The full likelihood function for Ψ , conditional on the observed data (Y_i, C_i) for $i = 1, \dots, n$ and the covariates, is thus

$$\begin{aligned}
L(\Psi; Y, C) &\propto \prod_{i=1}^n f(Y_i, C_i | \Psi) \\
&= \prod_{i=1}^n \int_{\theta} f(Y_i | C_i, \theta, \Psi) f(C_i | \theta, \Psi) f(\theta | \Psi) d\theta \\
&= \prod_{i=1}^n \int_{\theta} f(Y_i | \theta, \Psi) f(C_i | \theta, \Psi) f(\theta | \Psi) d\theta \\
&= \prod_{i=1}^n \int_{\theta} \left\{ \prod_{j=1}^{n_i} \frac{\Gamma(\frac{\kappa+1}{2})}{\sqrt{\pi \kappa \sigma^2} \Gamma(\frac{\kappa}{2})} \left[1 + \frac{1}{\sigma^2 \kappa} (Y_{ij} - \right. \right. \\
&\quad \left. \left. X_i^{(1)}(t_{ij})^T \beta - \tilde{X}_i^{(1)}(t_{ij})^T b) \right]^2 \right\}^{-\frac{\kappa+1}{2}} \\
&\quad \times \left[\prod_{k=1}^g \lambda_k(T_i; X_i^{(2)}(T_i), u, \gamma_k, \nu_k)^{I(D_i=k)} \right] \\
&\quad \times \exp \left\{ - \int_0^{T_i} \left[\sum_{k=1}^g \lambda_k(t; X_i^{(2)}(t), u, \gamma_k, \nu_k) \right] dt \right\} \\
&\quad \times \frac{1}{\sqrt{2\pi} |\Sigma|} \exp \left(- \frac{1}{2} \theta^T \Sigma^{-1} \theta \right) d\theta,
\end{aligned} \tag{3}$$

where Γ is the *Gamma* function. The second equality follows from the assumption that Y and C are independent conditional on all the covariates and the random effects. Write the baseline cumulative hazard function as $\Lambda_{0\kappa}(t) = \int \lambda_{0\kappa}(s) ds$, which belongs to a space consisting of all the increasing functions with $\lambda_{0\kappa}(0) = 0$. The maximum likelihood estimate for $\Lambda_{0\kappa}$ can be chosen as a step function with jumps at observed event times due to risk κ , where $\kappa = 1, \dots, g$.

It is obvious that the estimation of β and σ^2 is intractable in this likelihood function. It can be simplified using the fact that the distribution $t(0, \sigma^2, \kappa)$ can be derived by mixing the sample ε_{ij} with a scaling variate w_{ij} , such that

$$\varepsilon_{ij} | w_{ij} \sim N(0, \sigma^2 / w_{ij}), w_{ij} \sim \chi_{\kappa}^2 / \kappa, \tag{4}$$

where w_{ij} are i.i.d. *Gamma* random variables (Lange et al., 1989). Given w_{ij} and θ_i , the complete-data likelihood function can be written as follows:

$$\begin{aligned}
&L(\Psi; Y, C, \theta, w) \\
&\propto \prod_{i=1}^n \left\{ \prod_{j=1}^{n_i} \frac{\sqrt{w_{ij}}}{\sqrt{2\pi\sigma^2}} \exp \left[- \frac{w_{ij}}{2\sigma^2} (Y_{ij} - X_i^{(1)}(t_{ij})^T \beta - \tilde{X}_i^{(1)}(t_{ij})^T b)^2 \right] \right. \\
&\quad \left. \left[\prod_{k=1}^g \lambda_k(T_i; X_i^{(2)}(T_i), u_i, \gamma_k, \nu_k)^{I(D_i=k)} \right] \right. \\
&\quad \left. \times \exp \left\{ - \int_0^{T_i} \left[\sum_{k=1}^g \lambda_k(t; X_i^{(2)}(t), u_i, \gamma_k, \nu_k) \right] dt \right\} \right. \\
&\quad \left. \times \prod_{j=1}^{n_i} \frac{(\frac{\kappa}{2})^{\frac{\kappa}{2}} w_{ij}^{\frac{\kappa}{2}-1}}{\Gamma(\frac{\kappa}{2})} \exp \left(- \frac{\kappa}{2} w_{ij} \right) \times \frac{1}{\sqrt{2\pi} |\Sigma|} \exp \left(- \frac{1}{2} \theta_i^T \Sigma^{-1} \theta_i \right) \right\}.
\end{aligned} \tag{5}$$

The details of the EM algorithm is deferred to the Appendix. We point out that with t -distributed random errors, w_{ij} reweights the observations according to their residuals (see equation (13)), and thus we are able to obtain robust estimation by downweighting the outliers.

The parameter vector Ψ can be split into two components, the parametric component $\Omega = (\beta, \sigma^2, \gamma, \nu, \Sigma)$ and the collection of nonparametric baseline hazard functions $\Lambda = (\lambda_{01}(t), \dots, \lambda_{0g}(t))$. We propose to approximate the variance-covariance matrix of Ω by inverting the empirical Fisher information obtained from the profile likelihood where the baseline hazards functions have been profiled out (Lin, McCulloch and Rosenheck, 2004). Let $l^{(i)}(\Omega, Y, C)$ denote the observed score vector from the profile likelihood on the i th subject evaluated at Ω . The observed information matrix of Ω can be approximated by

$$\sum_{i=1}^n l^{(i)}(\hat{\Omega}; Y, C) l^{(i)}(\hat{\Omega}; Y, C)^T. \quad (6)$$

3 Example

The scleroderma lung study enrolled 158 patients with scleroderma-related interstitial lung disease, which were randomized to receive either CYC (2 mg/kg) (79 patients) or identical-appearing placebo (79 patients) for 12 months (Tashkin et al., 2006). A second year of follow-up was performed to determine if CYC effects persisted after stopping treatment. The primary outcome is forced vital capacity (FVC, % predicted) which was measured every 3 months from the baseline. A treatment failure occurred if a patient's %FVC fell by 15% or greater compared to the baseline after at least 3 months into treatment. The patients may dropout or die before the completion of the study and the average number of visits (including baseline) per patient is 7.3.

Comparison of baseline- and fibrosis-adjusted %FVC scores between the two groups revealed overlapping values prior to 6 months due to the dose-escalation protocol for CYC and the anticipated delay in treatment effect. In addition, the beneficial effects of CYC on pulmonary function continued to increase after stopping treatment, maximized at 18 months, but eventually dissipated at 24 months. For illustration purposes, we analyze 6 – 18 months %FVC scores at the longitudinal endpoint and the sample size is 141. There are 14 treatment failures or deaths (5 in CYC group and 9 in the placebo group), 32 informatively censored events (18 CYC and 14 placebo), and 5 noninformatively censored events (1 CYC and 4 placebo), in which the censored events were due to withdrawal in these patients. We pool together deaths and treatment failures into a single risk because the deaths were primarily due to extreme treatment failures in the forced vital capacity endpoint. We define "informatively censored events" if the patient withdrew due to disease or treatment related reasons, such as worsening disease, adverse event (AE) and serious adverse event (SAE). There could be %FVC measurements available after informatively censored events because these patients were encouraged to continue with scheduled visits and procedures, but the measurements may not necessarily be taken at all the subsequent visits. We define "noninformatively censored events" if there was no evidence showing the withdrawal was due to disease or treatment related reasons.

We propose a joint analysis of the longitudinal measurements %FVC and the time to treatment failure or death and the time to informative dropout, primarily based on the following three reasons: (1) we are interested in evaluating the effects of CYC on both longitudinal measurements of %FVC and the time to treatment failure or death simultaneously; (2) analysis of the longitudinal outcome %FVC alone can be complicated by treatment withdrawal or death, since patients with worse scleroderma lung disease condition tend to withdraw from the study early or die, hence are lost to follow up. The resultant non-ignorable missing data may lead to biased inference if we use the standard techniques such as mixed effects models (Henderson et al., 2000; Elashoff et al., 2007a, b). In the joint model we model the correlation between the

longitudinal outcome and the event times, so that we are able to adjust our inference for possible non-ignorable missing data; (3) the informatively censored events could be related to the risk of treatment failure or death, a competing risks framework should be adopted at the survival endpoint to handle their correlations.

We considered two factors in our joint model when assessing the CYC treatment effects: baseline %FVC (FVC_0) and baseline lung fibrosis (FIB_0). The latter was an important risk factor and had an interaction effect with the treatment as suggested by our preliminary analyses in which Huber's robust regression was carried out for %FVC scores at month 12. Figure 1(a) and (b) display the longitudinal profile of %FVC over time for the two groups. It suggests potential outliers as indicated by the big jumps in the plots, and we identified 6 possible outlying datapoints by examining the residuals. We are particularly concerned how these datapoints would affect the inference from the joint model. In table 1 the first two columns, we compared the results from our joint model with and without the outliers, in which we assumed a normal distribution for the measurement error. We considered a random slope model, because inclusion of both a random intercept and a random slope provided a marginal improvement of the overall fit. Without the outliers, the time trend in the placebo group is -0.16 , a more steep decrease in %FVC scores than that estimated from the joint model using all the datapoints. In addition, a significant interaction between the baseline %FVC and the treatment is identified by excluding the outliers. We also notice a fairly amount of changes in the estimated effect size for other fixed effects at the longitudinal endpoint. Since the outliers may introduce excessive noise in the data, the estimated standard errors for all the β 's become smaller after we excluded these datapoints. As a result, the p-value for the overall effects of CYC on %FVC scores, tested by the null hypothesis $\beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$, decreases from 0.1019 to 0.0265. The covariance σ_{bu} between the random effects b_i and u_i is significant in both columns. Because b_i and u_i are negatively correlated ($\sigma_{bu} < 0$) and v_2 is positive, it indicates that subjects with higher increasing rates in %FVC tend to have lower risks for both treatment failure or death and dropout. In addition, positive v_2 suggests that for subjects with a higher risk to dropout, there would also be a greater risk to treatment failure or death. There are comparable estimates at the survival endpoint, and the treatment effect is not significant for the risk of treatment failure or death by testing $\gamma_{23} = \gamma_{24} = \gamma_{25} = 0$. All the above comparisons suggest that inference from the joint model with a normal distribution for the measurement error is not robust against potential outliers in the longitudinal data, although the outliers may not have much impact on the estimation of the survival data.

The third column of Table 1 shows the estimates from the robust joint model with a t -distribution ($\kappa = 3$) for the measurement error using all the datapoints. There is a significant decreasing rate of the %FVC scores in the placebo group ($\beta_1 = -0.15$), and its interaction with CYC ($\beta_7 = 0.18$) suggests that on average %FVC increases about 0.03 every month in the CYC group. Similar to the first column, the method is able to detect a significant interaction effect between the baseline %FVC and the treatment. Testing the hypothesis $\beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$ for the overall CYC effects gives p-value < 0.0001 . There is also a negative covariance σ_{bu} between the random effects b_i and u_i and it is significant at 0.05 level. Even in the presence of the noise introduced by the potential outliers, the joint model using the t -distribution gets smaller standard errors for the fixed effects at the longitudinal endpoint, which is consistent with the simulations studied in the next section. All the three models generate very similar results for the fixed effects at the competing risks endpoint. The overall fit of the model is assessed using the likelihood. With the outliers, it is shown that the model assuming normal measurement errors has log of likelihood -2122 , lower than -2080 , the log of likelihood from the model using the t -distribution. For the robust joint model, we also assessed the model fit by varying the degree of freedom of the t -distribution, which shows that the likelihood reaches its maximum when $\kappa = 3$.

4 Simulation Studies

We conducted simulations to compare the robust joint model and the joint model assuming a normal distribution for the measurement error in terms of bias and standard errors. The data were simulated under a structure very similar to the scleroderma lung study. We seek to identify similar patterns of the estimates as shown in Table 1.

To be more specific, the longitudinal measurements were simulated from the following random slope model

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 FVC_{0i} + \beta_3 FIB_{0i} + \beta_4 CYC_i + \beta_5 FVC_{0i} \times CYC_i + \beta_6 FIB_{0i} \times CYC_i + \beta_7 t_{ij} \times CYC_i + b_i t_{ij} + \varepsilon_{ij}, \quad (7)$$

where $t_{ij} = 0, 3, \dots, 18$ was the scheduled visiting time (the datapoints eventually included in the joint model analysis were 6 – 18 months as in Table 1) and the baseline covariates FVC_{0i} , FIB_{0i} and CYC_i were sampled from distributions close to what we observed in the real data. The i.i.d. measurement errors ε_{ij} were simulated from a zero-mean normal distribution. To mimic the 6 outlying datapoints in the scleroderma lung study, the measurement errors of the outliers was generated from uniform distributions $U(-100, -50)$ and $U(50, 100)$ with rates 0.1% and 0.9%, respectively, and they had a higher probability to occur in later follow-up times. We simulated two competing risks for event times, say risk 1 and risk 2, with the following cause-specific hazards:

$$\lambda_1(t) = \lambda_{01}(t) \exp(\gamma_{11} FVC_{0i} + \gamma_{12} FIB_{0i} + \gamma_{13} CYC_i + \gamma_{14} FVC_{0i} \times CYC_i + \gamma_{15} FIB_{0i} \times CYC_i + u_i), \quad (8)$$

$$\lambda_2(t) = \lambda_{02}(t) \exp(\gamma_{21} FVC_{0i} + \gamma_{22} FIB_{0i} + \gamma_{23} CYC_i + \gamma_{24} FVC_{0i} \times CYC_i + \gamma_{25} FIB_{0i} \times CYC_i + v_i), \quad (9)$$

where the covariates were the same as in (7) and the baseline hazard functions were simulated from Weibull distributions as an approximation to the scleroderma lung study. The random effects b_i in (7) and u_i in (8–9) had a zero-mean bivariate normal distribution. All the parameters in model (7–9) were set to their estimated values from the robust joint model in Table 1.

Table 2 compares the results between the robust joint model and the joint model with normal-distributed random error using 100 Monte Carlo samples with $n = 141$. The degree of freedom for the t -distribution in the robust joint model was 3. We label the empirical standard error as SE, the median of the estimated standard error as Est.SE and the 95% confidence interval coverage probability as CP. The time trend β_1 is overestimated in the normal-error joint model, with an average at about -0.048 (and thus the bias is 0.102). This quantity is nearly unbiased when the t -distribution ($\kappa = 3$) for the measurement error was employed. The estimated standard errors from the latter decrease to around half in magnitude for the β 's, but are about the same for the γ 's, which are quite consistent with the estimates in Table 1. This is expected in the presence of outliers which produce extensive noise in the longitudinal data, but the normal distribution is not capable of removing that noise. Beyond these observations, we also notice that under the sample size $n = 141$, the standard error of σ_b^2 tends to be underestimated using the joint model with normal-distributed random error. The parameter v_2 is estimated poorly in both models, which suggests we should not overinterpret this quantity in the

application in Section 3. However, the bias of v_2 does not seem to affect the estimation of other parameters in the joint model.

In our other simulations with the joint model assuming normal-distributed random error (results not shown), the outliers in the longitudinal data can cause large bias not only in β and v_2 , but in σ_b^2 , and at the same time the standard errors of these and other parameters can be underestimated. In contrast, the robust joint model produces much lower bias in these parameters and the standard errors are more reasonably estimated. However, we do observe that the robust joint model may lose a little efficiency in estimating β when the outliers are absent.

5 Discussion

In this paper we propose a robust joint model for longitudinal measurements and competing risks failure times, in which the robustness against potential outliers in the longitudinal measurements is achieved by downweighting the influential datapoints using a t -distribution for the measurement error in the linear mixed effects sub-model. As in our previous work, the association between the two endpoints is modeled by the joint distribution between hidden random effects, which motivates an EM algorithm for parameter estimation. Our simulation studies and the real data application have shown that the joint model equipped with the t -distribution has a superb robust feature in the presence of outliers, although there is a small loss of efficiency when the outliers are absent. The $\kappa = 3$ or 4 works well in our limited empirical studies. In practice, one may also estimate κ from the data using the maximum likelihood method.

In the scleroderma lung study, one of the risks in the survival endpoint, treatment failure or death, is partly determined by the longitudinal outcome %FVC, which may violate the assumption that the longitudinal outcome and the event times are independent given the random effects. However, we observe in some sensitivity analysis the estimates may not be affected when the proportion of treatment failures is around 10%, which is similar to what we have in the scleroderma lung study. There could be biased estimates when the proportion is higher.

Song et al. (2002) and Hsieh, Tseng and Wang (2006) have reported robust features of the maximum likelihood estimates against departure from the normality assumption of the random effects under the joint modeling setting. We observe from our own simulations that the MLE's are robust when the random effects are from symmetric distributions, e.g., symmetric bimodal or trimodal normal distributions, or t -distributions. However, large bias is observed if the random effects follow some asymmetric distribution such as a *Gamma* or unbalanced mixture of normal distribution.

In general, inference for longitudinal measurements in joint analysis depends on the posited missing data mechanism, but in the presence of non-ignorable missing data, the assumption for the latent association is untestable. Some authors have conducted sensitivity analysis to examine the property of the proposed method when the modeling assumption is misspecified (Kenward, 1998; Verbeke, Molenberghs and Thijs, 2001; Minini and Chavance, 2004; Gad and Ahmed, 2007). This issue can be further explored in the setting of our joint model.

References

1. Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* 2003;59:221–228. [PubMed: 12926706]
2. DeGruttola V, Tu XM. Modeling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics* 1994;50:1003–1014. [PubMed: 7786983]

3. Elashoff RM, Li G, Li N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine* 2007a;26:2813–2835. [PubMed: 17124698]
4. Elashoff RM, Li G, Li N. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*. 2007b10.1111/j.1541-0420.2007.00952.x
5. Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* 1996;15:1663–1685. [PubMed: 8858789]
6. Gad AM, Ahmed AS. Sensitivity analysis of longitudinal data with intermittent missing values. *Statistical Methodology* 2007;4:217–226.
7. Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. *Biostatistics* 2000;4:465–480. [PubMed: 12933568]
8. Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* 1997;16:259–272. [PubMed: 9004396]
9. Hsieh F, Tseng YK, Wang JL. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* 2006;62:1037–1043. [PubMed: 17156277]
10. Huber PJ. Robust regression: asymptotics, conjectures and Monte Carlo. *Annals of Statistics* 1973;1:799–821.
11. Kenward MG. Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in Medicine* 1998;17:2723–2732. [PubMed: 9881418]
12. Lange KL, Little RJA, Taylor JMG. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* 1989;84:881–896.
13. Lin H, McCulloch CE, Rosenheck RA. Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* 2004;60:295–305. [PubMed: 15180654]
14. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* 1995;90:1112–1121.
15. Minini P, Chavance M. Sensitivity analysis of longitudinal normal data with drop-outs. *Statistics in Medicine* 2004;23:1039–1054. [PubMed: 15057877]
16. 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]
17. Press, WH.; Teutolsky, SA.; Vetterling, WT.; Flannery, BP. *Numerical Recipes in FORTRAN. The Art of Scientific Computing*; 2; Cambridge University Press; 1992.
18. Schluchter MD. Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* 1992;11:1861–1870. [PubMed: 1480878]
19. 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]
20. Sutradhar BC, Ali MM. Estimation of parameters of a regression model with a multivariate t error variable. *Communications in Statistics-Theory and Methods* 1986;15:429–450.
21. 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 E, Mubarak K, Connolly MK, Golden J, Olman M, Fessler B, Rothfield N, Metersky M. for 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]
22. Tseng YK, Hsieh F, Wang JL. Joint modelling of accelerated failure time and longitudinal data. *Biometrika* 2005;92:587–603.
23. Verbeke G, Molenberghs G, Thijs H. Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics* 2001;57:7–14. [PubMed: 11252620]
24. Wang Y, Taylor JMG. Joint modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* 2001;96:895–905.
25. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997;53:330–339. [PubMed: 9147598]
26. Xu J, Zeger SL. Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* 2001;50:375–387.

27. 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]

Appendix

Appendix: the EM Algorithm

E-step

Because the expectation of complete-data log-likelihood $l_i(\Psi, w_{ij}, \theta_i)$ for subject i is linear (up to a certain constant) in w_{ij} , we can write $E_{w_{ij}, \theta_i | Y_i, C_i, \Psi^{(m)}} [l_i(\Psi, w_{ij}, \theta_i)] = E_{\theta_i | Y_i, C_i, \Psi^{(m)}} [l_i(\Psi, \theta_i, E_{w_{ij} | \theta_i, Y_i, C_i, \Psi^{(m)}}(w_{ij}))]$, where indicates that the parameter estimates are from the m th iteration of the EM algorithm. In the E-step of the $(m + 1)$ th iteration, for subject i , we have

$$E_{w_{ij} | \theta_i, Y_i, C_i, \Psi^{(m)}}(w_{ij}) = \frac{\kappa + 1}{\kappa + \delta_{ij}^2(\Psi^{(m)})}, \quad (10)$$

where

$$\delta_{ij}^2(\Psi^{(m)}) = (Y_{ij} - X_i^{(1)}(t_{ij})^T \beta - \tilde{X}_i^{(1)}(t_{ij})^T b_i)^2 / \sigma^2, \quad (11)$$

which is a function of the hidden variable b_i . Therefore, it eventually boils down to evaluate the expected values of all functions of θ_i , say $h(\theta_i)$. In the E-step of the $(m + 1)$ th iteration,

$$\begin{aligned} E_{\theta_i | Y_i, C_i, \Psi^{(m)}} [h(\theta_i)] &= \int h(\theta_i) f(\theta_i | Y_i, C_i, \Psi^{(m)}) d\theta_i \\ &= \frac{\int h(\theta_i) f(Y_i, C_i, \theta_i | \Psi^{(m)}) d\theta_i}{f(Y_i, C_i | \Psi^{(m)})} \\ &= \frac{\int h(\theta_i) f(Y_i | \theta_i, \Psi^{(m)}) f(C_i | \theta_i, \Psi^{(m)}) f(\theta_i | \Psi^{(m)}) d\theta_i}{\int f(Y_i | \theta_i, \Psi^{(m)}) f(C_i | \theta_i, \Psi^{(m)}) f(\theta_i | \Psi^{(m)}) d\theta_i}. \end{aligned} \quad (12)$$

The integrals can be evaluated using Gauss-Hermite quadrature (Press et al., 1992).

M-step

Use E to stand for $E_{\theta_i | Y_i, C_i, \Psi^{(m)}}$. We have

$$\begin{aligned} \beta^{(m+1)} &= \left[\sum_{i=1}^n \sum_{j=1}^{n_i} E(w_{ij}) X_i^{(1)}(t_{ij}) X_i^{(1)}(t_{ij})^T \right]^{-1} \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} [E(w_{ij}) Y_{ij} - \right. \\ &\quad \left. E(w_{ij}) \tilde{X}_i^{(1)}(t_{ij})^T b_i] X_i^{(1)}(t_{ij}) \right\}, \end{aligned} \quad (13)$$

$$\sigma^{2(m+1)} = \frac{1}{\sum_{i=1}^n n_i} \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} E(w_{ij}) (Y_{ij} - X_i^{(1)}(t_{ij})^T \beta^{(m+1)} - \tilde{X}_i^{(1)}(t_{ij})^T b_i)^2 \right\}, \quad (14)$$

$$\sum_{bb}^{(m+1)} = \frac{1}{n} \sum_{i=1}^n E(b_i b_i^T), \quad (15)$$

$$\sigma_u^{2(m+1)} = \frac{1}{n} \sum_{i=1}^n E(u_i^2), \quad (16)$$

and

$$\sum_{bu}^{(m+1)} = \frac{1}{n} \sum_{i=1}^n E(b_i u_i). \quad (17)$$

Suppose there are q_κ distinct failure times due to the κ th cause and write $t_{\kappa 1} \leq \dots \leq t_{\kappa q_\kappa}$ for $\kappa = 1, \dots, g$. Let $R(t_{\kappa j})$ be the risk set at time $t_{\kappa j}$, and $d_{\kappa j}$ be the number of failures due to cause κ at time $t_{\kappa j}$. The cumulative baseline hazard function for cause κ is

$$H_{0\kappa}^{(m+1)}(t_{\kappa q}) = \sum_{j=1}^q \frac{d_{\kappa j}}{\sum_{r \in R(t_{\kappa j})} \exp(X_r^{(2)}(t_{\kappa j})^T \gamma_k^{(m)}) E(\exp(v_k^{(m)} u_r))}. \quad (18)$$

No closed-form solutions exist for γ and v . They can be updated using the following one-step Newton-Raphson algorithm in each iteration:

$$\gamma_k^{(m+1)} = \gamma_k^{(m)} + I_{\gamma_k}^{(m)-1} S_{\gamma_k}^{(m)} \quad (19)$$

where $\kappa = 1, \dots, g$, with $I_{\gamma_k}^{(m)}$ and $S_{\gamma_k}^{(m)}$ being

$$\begin{aligned} I_{\gamma_k}^{(m)} &= \sum_{i=1}^n \sum_{t_{\kappa j} \leq T_i} \lambda_{0\kappa}(t_{\kappa j})^{(m+1)} \exp(X_i^{(2)}(t_{\kappa j})^T \gamma_k^{(m)}) \times \\ &E(\exp(v_k^{(m)} u_i)) X_i^{(2)}(t_{\kappa j}) X_i^{(2)}(t_{\kappa j})^T, \end{aligned} \quad (20)$$

$$\begin{aligned} S_{\gamma_k}^{(m)} &= \sum_{i=1}^n [I(D_i = \kappa) X_i^{(2)}(T_i) - \sum_{t_{\kappa j} \leq T_i} \lambda_{0\kappa}(t_{\kappa j})^{(m+1)} \exp(X_i^{(2)}(t_{\kappa j})^T \gamma_k^{(m)}) \times \\ &E(\exp(v_k^{(m)} u_i)) X_i^{(2)}(t_{\kappa j})], \end{aligned} \quad (21)$$

$$v_k^{(m+1)} = v_k^{(m)} + S_{v_k}^{(m)} / I_{v_k}^{(m)} \quad (2)$$

where $\kappa = 2, \dots, g$, with $I_{v_k}^{(m)}$ and $S_{v_k}^{(m)}$ being

$$I_{v_k}^{(m)} = \sum_{i=1}^n \sum_{t_{kj} \leq T_i} \lambda_{0k}(t_{kj})^{(m+1)} \exp(X_i^{(2)}(t_{kj})^T \gamma_k^{(m+1)}) \times E(u_i^2 \exp(v_k^{(m)} u_i)), \quad (23)$$

$$S_{v_k}^{(m)} = \sum_{i=1}^n [I(D_i=k)E(u_i) - \sum_{t_{kj} \leq T_i} \lambda_{0k}(t_{kj})^{(m+1)} \exp(X_i^{(2)}(t_{kj})^T \gamma_k^{(m+1)}) \times E(u_i \exp(v_k^{(m)} u_i))]. \quad (24)$$

Fig. 1. Figure 1. (a)–(b) Profile plots of %FVC for CYC group vs. the placebo group: ○ for treatment failure or death; △ or informatively censored events; □ for noninformatively censored events.

Table 1
Analysis of 6 – 18 months scleroderma lung study data

	Normal Distribution		<i>t</i> -distribution ($\kappa = 3$)
	<u>Exclude the outliers</u>	<u>Include the outliers</u>	<u>Include the outliers</u>
	Estimate (SE)	Estimate (SE)	Estimate (SE)
<u>Longitudinal outcome %FVC</u>			
Time (β_1)	-0.16 (0.09)	-0.12 (0.12)	-0.15 (0.06)[†]
FV C_0 (β_2)	0.88 (0.04) [†]	0.92 (0.05) [†]	0.88 (0.03) [†]
FIB ₀ (β_3)	-1.86 (0.60) [†]	-2.03 (0.84) [†]	-2.42 (0.38) [†]
CYC (β_4)	-1.07 (0.94)	-1.25 (1.29)	-0.79 (0.77)
FV $C_0 \times CY C$ (β_5)	0.14 (0.07)[†]	0.11 (0.08)	0.14 (0.04)[†]
FIB ₀ $\times CY C$ (β_6)	2.20 (0.86) [†]	2.07 (1.04) [†]	2.68 (0.52) [†]
Time $\times CY C$ (β_7)	0.21 (0.13)	0.24 (0.16)	0.18 (0.09)[†]
p-value for $H_0: \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$	0.0265	0.1019	<0.0001
<u>Cause-specific hazards (informatively censored events)</u>			
FV C_0 (γ_{11})	-0.07 (0.03) [†]	-0.07 (0.03) [†]	-0.07 (0.03) [†]
FIB ₀ (γ_{12})	0.21 (0.34)	0.22 (0.33)	0.21 (0.34)
CYC (γ_{13})	0.31 (0.45)	0.31 (0.46)	0.33 (0.44)
FV $C_0 \times CY C$ (γ_{14})	0.10 (0.04) [†]	0.11 (0.04) [†]	0.10 (0.04) [†]
FIB ₀ $\times CY C$ (γ_{15})	0.08 (0.42)	0.09 (0.42)	0.09 (0.42)
<u>Cause-specific hazards (treatment failure or death)</u>			
FV C_0 (γ_{21})	0.02 (0.04)	0.01 (0.05)	0.02 (0.03)
FIB ₀ (γ_{22})	0.20 (0.42)	0.24 (0.43)	0.32 (0.37)
CYC (γ_{23})	-0.77 (0.96)	-0.77 (1.22)	-0.58 (0.83)
FV $C_0 \times CY C$ (γ_{24})	-0.06 (0.09)	-0.05 (0.09)	-0.05 (0.08)
FIB ₀ $\times CY C$ (γ_{25})	-0.52 (1.47)	-0.48 (1.51)	-0.50 (1.58)
p-value for $H_0: \gamma_{23} = \gamma_{24} = \gamma_{25} = 0$	0.67	0.82	0.85
<u>Random effects</u>			
v_2	3.19 (1.94)	2.80 (2.60)	7.28 (5.05)
σ_b^2	0.27 (0.04)	0.26 (0.04)	0.25 (0.03)
σ_u^2	0.20 (0.11)	0.24 (0.21)	0.03 (0.02)
<u>Covariance of b_i and u_i</u>			
σ_{bu}	-0.23 (0.06) [†]	-0.24 (0.08) [†]	-0.09 (0.02) [†]
[†] p-value < 0.05			

Table 2
Comparison of the joint model with a normal distribution versus that with a t -distribution for ε_{ij} where the simulated data have a similar structure to the scleroderma lung study.

Parameter	True	Normal Distribution				<i>t</i> -distribution ($\kappa = 3$)			
		Bias	SE	Est. SE	CP	Bias	SE	Est. SE	CP
<i>Longitudinal</i>									
Fixed effects									
β_0	67.20	-0.219	1.983	2.361	0.95	-0.090	1.066	0.993	0.93
β_1	-0.15	0.102	0.211	0.247	0.93	0.005	0.145	0.131	0.95
β_2	0.88	-0.001	0.058	0.106	0.98	-0.004	0.051	0.055	0.96
β_3	-2.42	0.102	0.769	1.178	1.00	0.072	0.575	0.585	0.94
β_4	-0.79	-0.094	2.301	3.380	1.00	-0.069	1.458	1.420	0.92
β_5	0.14	-0.008	0.082	0.155	1.00	-0.006	0.067	0.077	0.95
β_6	2.68	-0.117	1.093	1.637	0.98	-0.095	0.831	0.833	0.93
β_7	0.18	-0.001	0.264	0.333	0.99	0.020	0.189	0.182	0.95
Random effects									
σ_b^2	0.25	0.031	0.158	0.083	0.89	-0.035	0.046	0.048	0.96
<i>Survival</i>									
Fixed effects									
γ_{11}	-0.07	-0.011	0.030	0.033	0.97	-0.009	0.024	0.032	0.97
γ_{12}	0.21	-0.004	0.272	0.325	0.97	0.024	0.242	0.316	0.99
γ_{13}	0.33	0.049	0.448	0.426	0.97	-0.005	0.397	0.427	0.98
γ_{14}	0.10	0.013	0.038	0.044	0.99	0.011	0.034	0.043	0.98
γ_{15}	0.09	0.056	0.330	0.434	0.98	0.003	0.320	0.425	1.00
γ_{21}	0.02	-0.002	0.047	0.060	0.99	-0.004	0.041	0.054	0.93
γ_{22}	0.32	0.027	0.530	0.729	1.00	-0.007	0.505	0.607	0.97
γ_{23}	-0.58	0.007	1.173	1.183	0.98	-0.071	1.375	0.998	0.95
γ_{24}	-0.05	-0.001	0.093	0.107	0.99	0.010	0.087	0.094	0.96
γ_{25}	-0.50	-0.045	0.803	1.152	0.99	0.073	1.093	1.042	0.98

Parameter	True	Normal Distribution				<i>t</i> -distribution (<i>k</i> = 3)			
		Bias	SE	Est. SE	CP	Bias	SE	Est. SE	CP
Random effects									
ν_2	7.28	-1.495	6.949	8.944	0.86	-3.867	8.244	10.454	0.79
σ_u^2	0.03	0.004	0.018	0.049	1.00	-0.004	0.005	0.018	1.00
Covariance									
σ_{bu}	-0.086	-0.011	0.054	0.082	0.98	0.012	0.016	0.032	0.96

Note: nominal coverage probability (CP) is 0.95.