Journal of Statistical Research *https://doi.org/10.47302/jsr.2021550113*
2021, Vol. 55, No. 1, pp. 187-205 *ISSN 0256 - 422 X*

INFERENCE WITH JOINT MODELS UNDER MISSPECIFIED RANDOM EFFECTS DISTRIBUTIONS

ABDUS SATTAR*

Department of Population and Quantitative Health Sciences
Case Western Reserve University, 10900 Euclid Ave, Cleveland, OH, USA 44106
Email: sattar@case.edu

SANJOY K. SINHA

School of Mathematics and Statistics, Carleton University, Ottawa, ON K1S 5B6 Canada
Email: sinha@math.carleton.ca

SUMMARY

Joint models are often used to analyze survival data with longitudinal covariates or biomarkers. Latent random effects that are used to describe the relationship between longitudinal and survival outcomes are typically assumed to follow a multivariate Gaussian distribution. A joint likelihood analysis of the data provides valid inferences under a correctly specified random effects distribution. However, the maximum likelihood method may produce biased estimators under a misspecified random effects distribution, and hence may provide invalid inferences. In this paper, we explore the empirical properties of the maximum likelihood estimators under various types of random effects, and propose a skew-normal distribution to address uncertainties in random effects. An extensive Monte Carlo study shows that our proposed method provides robust and efficient estimators under various types of model misspecifications. We also present an application of the proposed method using a large clinical dataset obtained from the genetic and inflammatory markers of sepsis (GenIMS) study.

Keywords and phrases: Frailty; Joint model; Longitudinal data; Misspecified random effect; Mixed model; Skew-normal distribution.

AMS Classification: MSC 2000: Primary 62F10; secondary 62F35

1 Introduction

One of the fundamental aspects of joint modeling is the parameterization of the dependence structure among multivariate outcomes. Commonly used parameterization approaches include (a) interaction and lagged effects, (b) time-dependent slopes, (c) cumulative effects, and (d) random effects (Rizopoulos, 2015). A comprehensive list of association structures used in the joint modeling of longitudinal and survival data may be found in Hickey et al. (2016). In this paper, we focus on studying

* Corresponding author

joint models based on random effects. Joint models with random effects provide a general framework for describing the dependence structure (Verbeke and Davidian, 2009). However, the standard likelihood analysis may be sensitive to deviations from underlying distributional assumptions and hence suggests potential lack of robustness (Diggle et al., 2009). The impact of misspecified random effects distributions on joint modeling is unclear, and has not received much attention from the scientific community.

Misspecification of random effects distribution has been studied in the context of linear and generalized linear mixed models (Neuhaus et al., 2013). These models generally require a parametric specification of the random effects distribution, e.g., the normal distribution. Random effects may be incorporated into a mixed model as a subject specific random intercept only or subject specific random intercept and slope. Misspecification of one or more of the random effects terms has been studied in the literature (Neuhaus et al., 1992; Verbeke and Lesaffre, 1997; Heagerty and Kurland, 2001; McCulloch and Neuhaus, 2011). Depending on the random effect and its misspecification, there could be a certain degree of bias in the estimation of some parameters (Litire et al., 2008; Neuhaus et al., 2013). Pantazis and Touloumi (2007) study the impact of misspecification of the random effects distribution on modeling informatively censored bivariate longitudinal data using simulations, where the study shows that estimators of fixed effects regression parameters are consistent, but standard errors of the estimates can be underestimated. Li et al. (2012) consider joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times using nonparametric random effects distributions, where the authors consider bimodal, gamma, and uniform random effects distributions to demonstrate the robustness of the proposed nonparametric method. The skew-normal random effects distribution has been used in the context of longitudinal models (Arellano-Valle et al., 2005; Rastegaran and Zadkarami, 2015). Arellano-Valle et al. (2005) relaxed the normality assumption of random effects and error distribution for a linear mixed effects model by using the skew-normal distribution. Rastegaran and Zadkarami (2015) apply the skew-normal random effects distribution for modeling a longitudinal ordinal categorical response variable in the presence of non-ignorable missing data. Kim and Albert (2016) introduce a skewed multivariate random effects distribution in joint modeling of longitudinal continuous and discrete outcomes.

Random effects in joint models play an important role in studying the association between longitudinal and survival data, where "shared random effects" are often used to establish the relationship between the two outcome processes. Also, "association parameters" are used to determine the strength of relationship between the outcomes (Chen et al., 2011; Sattar et al., 2015; Sattar and Sinha, 2017; Alam et al., 2021). A comprehensive review of the joint modeling of longitudinal and survival data can be found in Tsiatis and Davidian (2004), Gould et al. (2015), Hickey et al. (2016), and Furgal et al. (2019). Typically, random effects linking various outcome processes are assumed to follow a normal distribution. Under a correctly specified normal distribution for the random effects, the joint modeling approach provides consistent and efficient estimators of the model parameters (Hogan and Laird, 1998). Effects of violations of the normality assumption on the random effects have not been studied much in the literature in the context of joint models. In this paper, we study the impact of misspecified random effects distributions on maximum likelihood estimators in joint models. We aim to provide a general framework for analyzing joint models assuming a flexible

skew-normal distribution for random effects. As a flexible approach, the proposed skew-normal distribution is able to capture potential asymmetry in the random effects, and hence can provide robust and efficient estimators.

Our motivation for studying the impact of a misspecified random effects distribution on joint models arises from multiple endpoints of a clinical study, named genetic and inflammatory marker of sepsis (GenIMS) study (Kellum et al., 2007). The GenIMS study, a large cohort study of patients with community acquired pneumonia (CAP), investigates the association among a group of inflammatory and coagulation biomarkers, and their effects on a clinical endpoint of "pneumonia-to-death" of sepsis patients. The patients were enrolled in the study through the emergency departments of 28 hospitals during 2001–2003. A goal of the study was to understand the dynamic behavior of the biomarkers in relation to severe sepsis, a primary cause of pneumonia-to-death. Blood samples were drawn from the hospitalized patients daily for the first seven days of hospitalization and weekly thereafter. The primary outcome variables in the GenIMS study were the incidence of severe sepsis and the 90-day mortality of patients. In this paper, we consider jointly analyzing a coagulation biomarker, named D-dimer, and the 90-day mortality endpoint. We also examine the relationship between a set of clinically relevant predictors of longitudinally measured D-dimer biomarker and the survival outcome. We use a linear mixed model for the longitudinal D-dimer data, where the clinical covariates of interests include "inpatient coagulation medication (Anticoa)" and "prior antibiotics use (Ant7pres)". In modeling the survival outcome, pneumonia-to-death, we use a similar set of covariates with a focus on the Charlson comorbidity score. The Charlson score includes diabetic information which is an important risk factor for many diseases including sepsis and death. Further details about the biomarkers and the GenIMS data analysis are provided in Section 4.

In summary, our study contributes to the joint modeling literature in several ways: it (1) investigates impacts of the misspecification of random effects, (2) extends the normal random effect approach to a flexible skew-normal approach, and (3) demonstrates the usefulness of the proposed method using both Monte Carlo simulations and applications. The proposed method is also applicable to a general joint modeling framework for multiple outcomes. The paper is organized as follows. In Section 2, we present the model and notation for joint models, and also investigate asymptotic properties of the estimators in joint models under misspecified random effects distributions. Section 3 investigates the finite-sample properties of the estimators based on a simulation study. Section 4 presents the findings of the real data analysis from the GenIMS study. We conclude the paper with some discussion in Section 5.

2 Model and Notations

In the following, we introduce submodels for the longitudinal and survival outcomes, where a linear mixed model is used for longitudinal data and a frailty model is used for survival data. The two outcome processes are linked through a set of shared random effects. We introduce a flexible skew-normal distribution as a robust alternative to the commonly used normal distribution to accommodate potential asymmetry in the random effects.

2.1 Longitudinal models

Suppose y_{ij} represents the longitudinal outcome from subject i measured at time t_{ij} ($i = 1, \ldots, N$; $j = 1, \ldots, n_i$). Assume that the outcome variable may be described as a function of available covariates by the linear mixed model

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{v}_i + \epsilon_{ij}, \tag{2.1}$$

where $\mathbf{x}_{ij} = (x_{1ij}, \ldots, x_{pij})'$ is a p-dimensional vector of covariates associated with the fixed effects and $\mathbf{z}_{ij} = (z_{1ij}, \ldots, z_{qij})'$ is a q-dimensional vector of covariates associated with the random effects. The random effects $\mathbf{v}_i$ are assumed to be independent and follow a density function $f_{v_i}(\mathbf{v}_i|\boldsymbol{\theta})$ with the mean vector $\mathbf{0}$ and covariance matrix $\mathbf{G}(\boldsymbol{\theta})$, depending upon a vector of variance components $\boldsymbol{\theta}$. The random errors ϵ_{ij} are assumed to be independent and follow a density function $f_{\epsilon_{ij}}(\epsilon_{ij}|\sigma^2_\epsilon)$ with the mean 0 and variance component σ^2_ϵ. Also, ϵ_{ij} are assumed to be independent of the random effects $\mathbf{v}_i$.

Model (2.1) may be written in the matrix form as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{v}_i + \boldsymbol{\epsilon}_i, \tag{2.2}$$

for $i = 1, \ldots, N$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})'$, $\mathbf{X}_i$ is an $n_i \times p$ design matrix for the fixed effects with its jth row being equal to $\mathbf{x}_{ij}$, $\mathbf{Z}_i$ is an $n_i \times q$ design matrix for the random effects with its jth row being equal to $\mathbf{z}_{ij}$, and the vector of random errors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})'$.

From (2.2), the conditional density of the ith response vector $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})'$ given the random effects $\mathbf{v}_i$ may be written as

$$f_{y_i|v_i}(\mathbf{y}_i|\mathbf{v}_i, \boldsymbol{\beta}, \sigma^2_\epsilon) = (2\pi\sigma^2_\epsilon)^{-n_i/2}\exp\left\{-\frac{1}{2\sigma^2_\epsilon}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{v}_i)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{v}_i)\right\}. \tag{2.3}$$

The conditional mean of the response vector $\mathbf{y}_i$ given the random effects $\mathbf{v}_i$ is obtained as $E(\mathbf{y}_i|\mathbf{v}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{v}_i$, and its conditional variance is obtained as $\text{Var}(\mathbf{y}_i|\mathbf{v}_i) = \sigma^2_\epsilon\mathbf{I}_{n_i}$, where $\mathbf{I}_{n_i}$ is an $n_i \times n_i$ identity matrix. The marginal mean and variance of the response vector $\mathbf{y}_i$ are obtained as $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$ and $\text{Var}(\mathbf{y}_i) = \mathbf{V}_i = \mathbf{Z}_i\mathbf{G}(\boldsymbol{\theta})\mathbf{Z}'_i + \sigma^2_\epsilon\mathbf{I}_{n_i}$.

Note that when making an inference based on only the longitudinal outcomes, one can fit the linear mixed model (2.2) using the standard weighted least squares method. For given random effects variance components $\boldsymbol{\alpha} = (\boldsymbol{\theta}', \sigma^2_\epsilon)'$, the best linear unbiased estimator (BLUE) of the fixed effects regression coefficients $\boldsymbol{\beta}$ may be obtained as

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N}\mathbf{X}'_i\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}'_i\mathbf{V}_i^{-1}\mathbf{y}_i\right), \tag{2.4}$$

whereas the best linear unbiased predictor (BLUP) of the random effects $\mathbf{v}_i$ may be obtained as

$$\tilde{\mathbf{v}}_i(\boldsymbol{\alpha}) = \mathbf{G}\mathbf{Z}'_i\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\tilde{\boldsymbol{\beta}}). \tag{2.5}$$

The variance components $\boldsymbol{\alpha}$ are unknown in practice. Several methods are available for estimating variance components, which include the widely used maximum likelihood (ML) and restricted maximum likelihood (REML) methods. For example, under the assumption that the random effects $\mathbf{v}_i$

and random errors ϵ_{ij} are normally distributed, the ML estimators $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{q+1})'$ may be obtained by solving the likelihood score equations

$$\sum_{i=1}^{N} \left\{ (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})' \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\alpha}_l} \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) - \text{tr}\left(\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\alpha}_l} \right) \right\} = \mathbf{0}, \tag{2.6}$$

using a numerical method, for $l = 1, \ldots, q+1$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)' \equiv (\alpha_1, \ldots, \alpha_q)'$ and $\sigma_\epsilon^2 = \alpha_{q+1}$.

The empirical best linear unbiased estimators of $\boldsymbol{\beta}$ and empirical best linear unbiased predictors (EBLUPs) of $\mathbf{v}_i$ may be obtained by replacing $\boldsymbol{\alpha}$ with the estimator $\hat{\boldsymbol{\alpha}}$, so that $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}})$ and $\hat{\mathbf{v}}_i = \tilde{\mathbf{v}}_i(\hat{\boldsymbol{\alpha}})$.

2.2 Frailty models

A frailty model is commonly used to study the heterogeneity among individuals when fitting a survivor function or a hazard function. Let s_i denote the true survival time (event time) and c_i the censored survival time for the ith individual ($i = 1, \ldots, N$). Let $T_i = \min(s_i, c_i)$ denote the observed survival time and δ_i the censoring status of the survival time for the ith individual, with $\delta_i = 0$ if T_i is censored, and $\delta_i = 1$, otherwise. Consider a set of m baseline covariates $\mathbf{w}_i = (w_{i1}, \ldots, w_{im})'$ observed from the ith individual. To study the effects of covariates, we consider a frailty model in the form

$$h_i(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi}) = h_0(T_i|\boldsymbol{\tau}) \exp(\mathbf{w}_i'\boldsymbol{\gamma} + \boldsymbol{\varphi}'\mathbf{v}_i), \tag{2.7}$$

for $i = 1, \ldots, N$, where $h_i(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi})$ denotes the conditional hazard function of the ith individual given the frailty $\mathbf{v}_i$ and $h_0(T_i|\boldsymbol{\tau})$ denotes the baseline hazard function depending on unknown parameters $\boldsymbol{\tau}$. The frailties $\mathbf{v}_i$ ($i = 1, \ldots, N$) are assumed to be independent and follow a density $f_{v_i}(\mathbf{v}_i|\boldsymbol{\theta})$ with the mean vector $\mathbf{0}$ and covariance matrix $\mathbf{G}(\boldsymbol{\theta})$. Given frailty random effects $\mathbf{v}_i$, from (2.7) the conditional density of the observed survival data (T_i, δ_i) for the ith individual may be obtained as

$$\begin{aligned} f_{T_i,\delta_i|v_i}(T_i, \delta_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi}) &= \left[f_{T_i|v_i}(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi}) \right]^{\delta_i} \left[S_i(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi}) \right]^{1-\delta_i} \\ &= \left[h_i(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi}) \right]^{\delta_i} S_i(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi}), \end{aligned}$$

where $f_{T_i|v_i}(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi})$ denotes the conditional density of the survival time T_i given the frailty $\mathbf{v}_i$ and $S_i(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi})$ denotes the conditional survivor function given the frailty $\mathbf{v}_i$.

Note that the aforementioned longitudinal and survival outcome processes are linked through the shared random effects $\mathbf{v}_i$, where the association parameters $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_q)'$ determines the degree of association between the two outcome processes.

2.3 Joint models

Recall the longitudinal and survival submodels discussed earlier. Given the observed data (T_i, δ_i, y_i), the contribution of the ith subject to the log-likelihood of $\boldsymbol{\xi} = (\boldsymbol{\gamma}', \boldsymbol{\tau}', \boldsymbol{\varphi}', \boldsymbol{\beta}', \boldsymbol{\theta}', \sigma_\epsilon^2)'$ is given by

$$l_i(\boldsymbol{\xi}|T_i, \delta_i, \mathbf{y}_i) = \log \int \Big[f_{T_i,\delta_i|v_i}(T_i, \delta_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi})\ f_{y_i|v_i}(\mathbf{y}_i|\mathbf{v}_i, \boldsymbol{\beta}, \sigma_\epsilon^2)\ f_{v_i}(\mathbf{v}_i|\boldsymbol{\theta}) \Big] d\mathbf{v}_i, \quad (2.8)$$

for $i = 1, \ldots, N$, where the conditional density of the survival time T_i is given by

$$\begin{aligned} & f_{T_i,\delta_i|v_i}(T_i, \delta_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi}) \\ &= \big[h_i(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau})\big]^{\delta_i} S_i(T_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}) \\ &= \big[h_0(T_i|\boldsymbol{\tau})\ \exp(\mathbf{w}_i'\boldsymbol{\gamma} + \boldsymbol{\varphi}'\mathbf{v}_i)\big]^{\delta_i} \times \exp\left\{ - \int_0^{T_i} h_0(s;\tau)\ \exp(\mathbf{w}_i'\boldsymbol{\gamma} + \boldsymbol{\varphi}'\mathbf{v}_i)\, ds \right\}, \end{aligned}$$

and the conditional density $f_{y_i|v_i}(\mathbf{y}_i|\mathbf{v}_i, \boldsymbol{\beta}, \sigma_\epsilon^2)$ of the longitudinal responses $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})'$ is given by

$$f_{y_i|v_i}(\mathbf{y}_i|\mathbf{v}_i, \boldsymbol{\beta}, \sigma_\epsilon^2) = \prod_{j=1}^{n_i} f_{y_{ij}|v_i}(y_{ij}|\mathbf{v}_i, \boldsymbol{\beta}, \sigma_\epsilon^2).$$

The density $f_{y_{ij}|v_i}(y_{ij}|\mathbf{v}_i, \boldsymbol{\beta}, \sigma_\epsilon^2)$ is assumed to be normal with mean $\mu_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{v}_i$ and variance σ_ϵ^2. Here the two outcome processes are linked through the random effects $\mathbf{v}_i$ and association parameters $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_q)'$. When $\boldsymbol{\varphi} = \mathbf{0}$, a joint analysis of the two outcome processes would be equivalent to their separate analyses.

The maximum likelihood estimator of $\boldsymbol{\xi}$ may be obtained by maximizing the observed data log-likelihood

$$l(\boldsymbol{\xi}) = \sum_{i=1}^{N} l_i(\boldsymbol{\xi}|T_i, \delta_i, \mathbf{y}_i) \quad (2.9)$$

with respect to $\boldsymbol{\xi}$ using a numerical method. Equivalently, we can solve the likelihood score equation $\Phi(\boldsymbol{\xi}) = \mathbf{0}$, where the score function $\Phi(\boldsymbol{\xi})$ may be obtained as

$$\Phi(\boldsymbol{\xi}) \quad = \quad \sum_{i=1}^{N} \int A(\boldsymbol{\xi}, \mathbf{v}_i)\ f_{v_i}(\mathbf{v}_i|T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\xi})\ d\mathbf{v}_i, \quad (2.10)$$

with $A(\boldsymbol{\xi}, \mathbf{v}_i)$ being the "complete data" score function, given by

$$A(\boldsymbol{\xi}, \mathbf{v}_i) = \frac{\partial}{\partial \boldsymbol{\xi}} \Big[\log f_{T_i,\delta_i|v_i}(T_i, \delta_i|\mathbf{v}_i, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\varphi}) + \log f_{y_i|v_i}(\mathbf{y}_i|\mathbf{v}_i, \boldsymbol{\beta}, \sigma_\epsilon^2) + \log f_{v_i}(\mathbf{v}_i|\boldsymbol{\theta}) \Big]. \quad (2.11)$$

The function $f(\mathbf{v}_i|T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\xi})$ represents the conditional density of the random effects $\mathbf{v}_i$, given the observed data $(T_i, \delta_i, \mathbf{y}_i)$. This density function does not have a closed form, in general, and numerical methods are needed to compute the log-likelihood, score function and Fisher information.

The score equation $\Phi(\boldsymbol{\xi}) = \mathbf{0}$ may be solved numerically using an iterative method, such as the Newton-Raphson method. Given some initial estimate $\boldsymbol{\xi}^{(0)}$, the Newton-Raphson method obtains the estimates using the iterative equations

$$\boldsymbol{\xi}^{(r+1)} = \boldsymbol{\xi}^{(r)} - \left\{\Phi^{(1)}\left(\boldsymbol{\xi}^{(r)}\right)\right\}^{-1}\left\{\Phi\left(\boldsymbol{\xi}^{(r)}\right)\right\},$$

for $r = 0, 1, 2, \ldots$, where $\Phi(\boldsymbol{\xi}^{(r)})$ is the likelihood score function $\Phi(\boldsymbol{\xi})$, evaluated at $\boldsymbol{\xi}^{(r)}$, and $\Phi^{(1)}(\boldsymbol{\xi}^{(r)})$ is the first derivative of the score function $\Phi(\boldsymbol{\xi})$ with respect to $\boldsymbol{\xi}$, evaluated at $\boldsymbol{\xi}^{(r)}$. We can show that

$$\begin{aligned}
\Phi^{(1)}(\boldsymbol{\xi}) &= \frac{\partial}{\partial \boldsymbol{\xi}}\Phi(\boldsymbol{\xi}) \\
&= \sum_{i=1}^{N}\int \frac{\partial}{\partial \boldsymbol{\xi}} A(\boldsymbol{\xi}, \mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\xi})\, d\mathbf{v}_i \\
&\quad + \sum_{i=1}^{N}\int A(\boldsymbol{\xi}, \mathbf{v}_i)\, A'(\boldsymbol{\xi}, \mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\xi})\, d\mathbf{v}_i \\
&\quad - \sum_{i=1}^{N}\int A(\boldsymbol{\xi}, \mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\xi})\, d\mathbf{v}_i \int A'(\boldsymbol{\xi}.\mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\xi})\, d\mathbf{v}_i.
\end{aligned}$$

At convergence, we obtain the maximum likelihood (ML) estimates $\hat{\boldsymbol{\xi}}$ of the model parameters $\boldsymbol{\xi}$. Under the assumption that the density $f_{v_i}(\mathbf{v}_i|\boldsymbol{\theta})$ of the random effects $\mathbf{v}_i$ is correctly specified, the ML estimator $\hat{\boldsymbol{\xi}}$ follows an asymptotic normal distribution with the mean $\boldsymbol{\xi}$ and an approximate variance $V(\hat{\boldsymbol{\xi}}) = I^{-1}(\boldsymbol{\xi})$, where $I(\boldsymbol{\xi})$ is the observed Fisher information, given by $I(\boldsymbol{\xi}) = -\Phi^{(1)}(\boldsymbol{\xi})$.

2.4 Joint analysis under misspecified random effects

It is well-known that the maximum likelihood estimators are the most efficient under correctly specified distributions. However, under misspecified distributions, the ML method may provide systematic bias in the estimation. In this section, we investigate the properties of the maximum likelihood estimator $\hat{\boldsymbol{\xi}}$ for the case when the density $f_{v_i}(\mathbf{v}_i|\boldsymbol{\theta})$ of the random effects $\mathbf{v}_i$ is incorrectly specified.

Recall that the maximum likelihood estimator $\hat{\boldsymbol{\xi}}_N \equiv \hat{\boldsymbol{\xi}}$ are obtained by solving

$$N^{-1}\sum_{i=1}^{N}\int A(\boldsymbol{\xi}, \mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\xi})\, d\mathbf{v}_i\Big|_{\hat{\boldsymbol{\xi}}_N} = \mathbf{0}, \tag{2.12}$$

where $f_{v_i}(\mathbf{v}_i|T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\xi})$ denotes the conditional distribution of $\mathbf{v}_i$, given the observed data $(T_i, \delta_i, \mathbf{y}_i)$ for the ith individual. If the marginal density $f_{v_i}(\mathbf{v}_i|\boldsymbol{\theta})$ of $\mathbf{v}_i$ is misspecified, then the ML estimator $\hat{\boldsymbol{\xi}}_N$ converges to $\boldsymbol{\xi}^*$ that minimizes the Kullback-Leibler information criterion (White, 1982), or equivalently, satisfies the estimating equation

$$\lim_{N\to\infty} N^{-1}E\left[\sum_{i=1}^{N}\int A(\boldsymbol{\xi}, \mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\xi})\, d\mathbf{v}_i\right]_{\boldsymbol{\xi}^*} = \mathbf{0}, \tag{2.13}$$

where the expectation E is taken with respect to the underlying "true distribution" with parameters $\boldsymbol{\xi}_0$.

Following White (1982), we can show that $\hat{\boldsymbol{\xi}}_N$ follows an asymptotic normal distribution, i.e.,

$$\sqrt{N}(\hat{\boldsymbol{\xi}}_N - \boldsymbol{\xi}^*) \sim N(\mathbf{0}, C(\boldsymbol{\xi}^*)), \tag{2.14}$$

where $C(\boldsymbol{\xi}^*)$ is obtained from a sandwich-type variance-covariance matrix $C(\boldsymbol{\xi})$, given by $C(\boldsymbol{\xi}) = M(\boldsymbol{\xi})^{-1}Q(\boldsymbol{\xi})M(\boldsymbol{\xi})^{-1}$, evaluated at $\boldsymbol{\xi} = \boldsymbol{\xi}^*$, with

$$M(\boldsymbol{\xi}) = \lim_{N\to\infty} N^{-1}E\left[\sum_{i=1}^{N}\frac{\partial}{\partial\boldsymbol{\xi}}\int A(\boldsymbol{\xi},\mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i,\delta_i,\mathbf{y}_i;\boldsymbol{\xi})\, d\mathbf{v}_i\right],$$

and

$$Q(\boldsymbol{\xi}) = \lim_{N\to\infty} N^{-1}E\left[\sum_{i=1}^{N}\int A(\boldsymbol{\xi},\mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i,\delta_i,\mathbf{y}_i;\boldsymbol{\xi})\, d\mathbf{v}_i \times \int A'(\boldsymbol{\xi},\mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i,\delta_i,\mathbf{y}_i;\boldsymbol{\xi})\, d\mathbf{v}_i\right].$$

The above variance-covariance matrix $C(\boldsymbol{\xi})$ may be approximated by

$$C_N(\boldsymbol{\xi}) = M_N(\boldsymbol{\xi})^{-1}\, Q_N(\boldsymbol{\xi})\, M_N(\boldsymbol{\xi})^{-1}, \tag{2.15}$$

where

$$M_N(\boldsymbol{\xi}) = N^{-1}\sum_{i=1}^{N}\frac{\partial}{\partial\boldsymbol{\xi}}\int A(\boldsymbol{\xi},\mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i,\delta_i,\mathbf{y}_i;\boldsymbol{\xi})\, d\mathbf{v}_i,$$

and

$$Q_N(\boldsymbol{\xi}) = N^{-1}\sum_{i=1}^{N}\int A(\boldsymbol{\xi},\mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i,\delta_i,\mathbf{y}_i;\boldsymbol{\xi})\, d\mathbf{v}_i \int A'(\boldsymbol{\xi},\mathbf{v}_i)\, f_{v_i}(\mathbf{v}_i|T_i,\delta_i,\mathbf{y}_i;\boldsymbol{\xi})\, d\mathbf{v}_i.$$

Under appropriate regularity conditions as given in White (1982), we can show that as $N \to \infty$, $C_N(\hat{\boldsymbol{\xi}}_N) \to C(\boldsymbol{\xi}^*)$ a.s. In the next section, we study the asymptotic bias $(\boldsymbol{\xi}^* - \boldsymbol{\xi}_0)$ of the estimator $\hat{\boldsymbol{\xi}}_N$ based on a Monte Carlo study. Specifically, we calculate the empirical bias $[(1/S)\sum_{s=1}^{S}\hat{\boldsymbol{\xi}}_N^{(s)} - \boldsymbol{\xi}_0]$ based on S replicates of data sets, where $\hat{\boldsymbol{\xi}}_N^{(s)}$ is the estimator $\hat{\boldsymbol{\xi}}_N$ obtained from the sth replicate $(s = 1, \ldots, S)$. The empirical results are obtained under various misspecified random effects distributions, where we study the robustness properties of the proposed estimators under skew-normal random effects.

3 Simulation Study

We ran a series of simulations using a linear mixed model for longitudinal outcomes and a Weibull frailty model for survival outcomes. Specifically, the longitudinal outcomes were obtained from the linear mixed model

$$y_{ij} = \beta_0 + \beta_1 \text{Visit}_{ij} + \beta_2 \text{Treat}_i \times \text{Visit}_{ij} + \beta_3 x_i + v_i + \epsilon_{ij}, \tag{3.1}$$

for $i = 1, \ldots, N,\ j = 1, \ldots, n_i$, where the random errors ϵ_{ij} are assumed to be independent $N(0, \sigma_\epsilon^2)$. The random effects v_i are assumed to follow a distribution with mean 0 and variance σ_v^2, independent of ϵ_{ij}. The group indicator Treat_i is 0 if the ith subject is in the control group and 1, if in the treatment group, where we consider equal number of subjects in each group. The values of the baseline covariate x_i, representing measurements on a biomarker, were generated from the normal distribution $N(\mu_x, \sigma_x^2)$ with $\mu_x = 1.8$ and $\sigma_x^2 = 2.9$. Also, Visit_{ij} represents the follow-up times at which the longitudinal measurements are obtained.

The survival outcomes were obtained from the Weibull frailty model

$$h_i(T_i|v_i, \boldsymbol{\gamma}, \lambda, \tau, \varphi) = \lambda\tau T_i^{\tau-1} \exp\left(\gamma_1 \text{Treat}_i + \gamma_2 x_i + \varphi v_i\right), \tag{3.2}$$

for $i = 1, \ldots, N$. The above two outcome processes are linked through the shared random effects v_i, where the "association parameter" φ measures the strength of association between the two processes.

The data were generated using two combinations of sample sizes, $N = 250$ and 500. Each simulation run was based on $S = 1000$ replicates of data sets. The model parameters were fixed at $(\beta_0, \beta_1, \beta_2, \beta_3) = (10.0, -0.25, 1.5, 0.5)$, $(\sigma_v^2, \sigma_\epsilon^2) = (1.0, 0.5)$, $(\log\lambda, \tau) = (-4.6, 0.8)$, $(\gamma_1, \gamma_2) = (1.5, 0.5)$, and $\varphi = 0.5$. The censored survival times were obtained by setting $\log\lambda = -4.7$, which resulted in roughly 48% right-censored values.

To study the effects of misspecified random effects, v_i were generated from the following three combinations of distributions:

$$\begin{aligned}
\text{Normal:} \qquad & v_i \sim N(0, \sigma_v^2), \\
\text{Chi-square:} \qquad & v_i = \sigma_v \left(\frac{u_i - E(u_i)}{\sqrt{\text{Var}(u_i)}} \right), \text{ with } u_i \sim \text{Chi-square}(7\,\text{df}), \\
\text{Gamma:} \qquad & v_i = \sigma_v \left(\frac{u_i - E(u_i)}{\sqrt{\text{Var}(u_i)}} \right), \text{ with } u_i \sim \text{Gamma}(2, 0.8).
\end{aligned}$$

Models (3.1) and (3.2) were fitted under the following two combinations of random effects distributions:

$$\begin{aligned}
\text{Normal:} \qquad & v_i \sim N(0, \sigma_v^2) \quad \text{(naive method)}, \\
\text{Skew-normal:} \qquad & v_i = \sigma_v (u_i - E(u_i))/\sqrt{\text{Var}(u_i)} \quad \text{(robust method)},
\end{aligned}$$

where the skew-normal density of u_i is given by

$$f_{u_i}(u_i|\sigma_v^2, \delta) = \frac{2}{\sigma_v} \phi\left(\frac{u_i}{\sigma_v}\right) \Phi\left(\delta\left(\frac{u_i}{\sigma_v}\right)\right),$$

with $\phi(z)$ and $\Phi(z)$ being the standard normal and cumulative standard normal distributions, respectively. Note that the choice of the shape (skewness) parameter $\delta = 0$ leads to the normal distribution $N(0, \sigma_v^2)$. We estimate δ and other model parameters simultaneously by the joint maximum likelihood approach discussed earlier. We then study the empirical properties of the estimators under both correctly specified and misspecified random effects distributions.

Table 1 presents empirical biases, mean squared errors (MSEs) and coverage probabilities of the maximum likelihood estimators for $N = 250$. Table 2 repeats the results for $N = 500$. The estimates were obtained under the assumption that the random effects followed normal or skew-normal, when the true distribution is, in fact, either symmetric (normal) or skew-symmetric (chi-square or gamma). It is clear from the tables that the estimates of the regression coefficients $(\beta_0, \beta_1, \beta_2, \beta_3)$ and (γ_1, γ_2) in the longitudinal and survival submodels are roughly unbiased under all scenarios considered. The coverage probabilities of the estimates are also close to the nominal 95% confidence level.

When the true distribution of v_i is normal, the ML method based on the normality assumption appears to provide better estimates, as compared to the skew-normal distribution. This is expected since the ML estimates are generally the most efficient under the correctly specified model. However, it is interesting to note that the proposed skew-normal distribution also provides estimates that are almost as efficient as those obtained from the correctly specified normality assumption. For example, when estimating (β_3, γ_2), effects of x_i on longitudinal and survival outcomes, the efficiencies of the estimates from the skew-normal approach are obtained as $(0.1849/0.1856, 0.5012/0.5000) \times 100 = (99.6\%, 100.2\%)$, as shown in Table 1 for $N = 250$. We can expect to lose such a small efficiency from the skew-normal approach when the true distribution is, in fact, normal. However, our main interest is in the non-normal distribution of v_i, where the skew-normal approach is expected to perform better than the normal approach.

When the true distribution of v_i is chi-square, the skew-normal approach appears to provide generally better estimates as compared to the normal approach. For example, when estimating (β_3, γ_2), the skew-normal approach provides corresponding efficiencies of $(0.1718/0.1354, 0.5271/0.5077) \times 100 = (126.9\%, 103.8\%)$, as shown in Table 1 for $N = 250$. Also, for $N = 500$, Table 2 shows corresponding efficiencies of the estimates as $(0.0867/0.0677, 0.2415/0.2341) \times 100 = (128.1\%, 103.2\%)$.

We observe a similar pattern when the true distribution of v_i is gamma. For example, when estimating (β_3, γ_2), the skew-normal approach provides corresponding efficiencies of $(0.1689/0.1218, 0.5212/0.4907) \times 100 = (138.7\%, 106.2\%)$, as shown in Table 1 for $N = 250$. For $N = 500$, Table 2 shows the corresponding efficiencies as $(0.0842/0.0594, 0.2633/0.2466) \times 100 = (141.8\%, 106.8\%)$.

The gain in efficiency by the skew-normal approach is even more dramatic when estimating the variance component and association parameter (σ_v^2, φ) under the misspecified model. For example, when the true distribution of v_i is chi-square, it is clear from Table 1 that the skew-normal approach provides corresponding efficiencies of the estimates of (σ_v^2, φ) as $(2.2395/1.7361, 2.3081/1.6342) \times 100 = (129.0\%, 141.2\%)$.

In summary, the proposed skew-normal approach appears to be robust against misspecified random effects distributions. One can encounter considerable loss of efficiency from the ordinary normal approach if the random effects distribution is, in fact, non-normal. In such cases, the skew-normal approach should be considered as a robust and efficient approach to analyzing joint models.

Table 1: Empirical percentage relative biases, mean squared errors (MSEs) and coverage probabilities (CPs) of the maximum likelihood estimators obtained under misspecified random effects distribution. Parameters are fixed at $(\beta_0, \beta_1, \beta_2, \beta_3) = (10.0, -0.25, 1.5, 0.5)$, $(\sigma_v^2, \sigma_\epsilon^2) = (1.0, 0.5)$, $(\log\lambda, \tau) = (-4.6, 0.8)$, $(\gamma_1, \gamma_2) = (1.5, 0.5)$, and $\varphi = 0.5$. Number of subjects $N =$ **250**.

			Fitted model: Normal			Fitted model: Skew-normal		
Effect (v_i)	Model	Coef	%Bias	100×MSE	CP	%Bias	100×MSE	CP
Normal	Long	β_0	0.0396	1.0848	0.944	0.0354	1.0918	0.946
		β_1	1.3340	0.0120	0.946	1.3431	0.0119	0.952
		β_2	0.6977	0.1202	0.940	0.7194	0.1203	0.938
		β_3	0.7267	0.1849	0.944	0.7114	0.1856	0.942
		σ_v^2	1.5090	1.3819	0.954	0.6332	1.4259	0.962
		σ_ϵ^2	0.0021	0.0957	0.956	0.2705	0.0952	0.962
	Surv	$\log\lambda$	3.1361	15.6034	0.962	3.0395	15.4625	0.958
		τ	3.4927	0.4638	0.950	3.4158	0.4600	0.954
		γ_1	1.7661	4.8398	0.954	1.8178	4.8421	0.954
		γ_2	1.3998	0.5012	0.940	1.1126	0.5000	0.944
		φ	8.2848	1.6533	0.942	8.0126	1.6215	0.946
Chi-sq	Long	β_0	0.1702	1.0408	0.938	0.0408	0.9130	0.944
		β_1	1.1291	0.0116	0.956	1.0828	0.0109	0.950
		β_2	0.2265	0.1156	0.960	0.3843	0.1028	0.952
		β_3	0.2477	0.1718	0.946	0.5145	0.1354	0.936
		σ_v^2	9.7347	2.2395	0.762	4.0706	1.7361	0.896
		σ_ϵ^2	1.5906	0.1073	0.934	0.1571	0.0941	0.952
	Surv	$\log\lambda$	4.1703	17.5516	0.956	3.9072	16.8945	0.964
		τ	4.2017	0.5108	0.950	4.0630	0.4990	0.950
		γ_1	2.5822	4.9831	0.958	2.4463	4.9218	0.962
		γ_2	3.1719	0.5271	0.956	2.8380	0.5077	0.956
		φ	14.6827	2.3081	0.924	8.1480	1.6342	0.948
Gamma	Long	β_0	0.2336	1.0139	0.944	0.0547	0.8861	0.938
		β_1	0.8918	0.0119	0.942	0.5455	0.0105	0.942
		β_2	0.1315	0.1153	0.946	0.2785	0.0962	0.958
		β_3	0.6613	0.1689	0.948	0.4543	0.1218	0.950
		σ_v^2	11.0333	2.4494	0.706	7.4246	1.9467	0.826

Continued on next page

Table 1 – *Continued from previous page*

			Fitted model: Normal			Fitted model: Skew-normal		
Effect (v_i)	Model	Coef	%Bias	100×MSE	CP	%Bias	100×MSE	CP
		σ^2_ϵ	2.1181	0.1103	0.940	0.1271	0.0934	0.948
	Surv	$\log\lambda$	3.6902	16.8862	0.948	3.3809	15.9310	0.950
		τ	3.6948	0.5059	0.950	3.6869	0.5011	0.950
		γ_1	1.8704	4.8890	0.950	1.7088	4.8325	0.950
		γ_2	4.1824	0.5212	0.944	3.8619	0.4907	0.954
		φ	14.2490	2.3849	0.914	7.3589	1.5997	0.924

Table 2: Empirical percentage relative biases, mean squared errors (MSEs) and coverage probabilities (CPs) of the maximum likelihood estimators obtained under misspecified random effects distribution. Parameters are fixed at $(\beta_0, \beta_1, \beta_2, \beta_3) = (10.0, -0.25, 1.5, 0.5)$, $(\sigma^2_v, \sigma^2_\epsilon) = (1.0, 0.5)$, $(\log\lambda, \tau) = (-4.6, 0.8)$, $(\gamma_1, \gamma_2) = (1.5, 0.5)$, and $\varphi = 0.5$. Number of subjects $N =$ **500**.

			Fitted model: Normal			Fitted model: Skew-normal		
Effect (v_i)	Model	Coef	%Bias	100×MSE	CP	%Bias	100×MSE	CP
Normal	Long	β_0	0.0915	0.5515	0.954	0.0838	0.5509	0.958
		β_1	1.5145	0.0068	0.920	1.4812	0.0068	0.920
		β_2	0.4520	0.0530	0.938	0.5127	0.0538	0.936
		β_3	0.2382	0.0914	0.968	0.3263	0.0916	0.966
		σ^2_v	1.1513	0.6972	0.938	0.7181	0.7100	0.940
		σ^2_ϵ	0.0511	0.0478	0.962	0.0323	0.0477	0.960
	Surv	$\log\lambda$	2.9720	8.5403	0.944	2.9315	8.4901	0.942
		τ	3.1240	0.2521	0.938	3.0985	0.2513	0.938
		γ_1	1.7599	2.4166	0.930	1.5324	2.4010	0.932
		γ_2	1.7542	0.2501	0.954	1.6169	0.2490	0.956
		φ	7.1983	0.8488	0.940	6.8569	0.8320	0.948
Chi-sq	Long	β_0	0.0698	0.5211	0.964	0.0268	0.4607	0.964
		β_1	1.1031	0.0062	0.956	1.1245	0.0059	0.944
		β_2	0.1826	0.0528	0.954	0.3311	0.0473	0.948
		β_3	0.0925	0.0867	0.952	0.5127	0.0677	0.964
		σ^2_v	7.5146	1.2343	0.794	1.6451	0.8533	0.924

Continued on next page

Table 2 – *Continued from previous page*

			Fitted model: Normal			Fitted model: Skew-normal		
Effect (v_i)	Model	Coef	%Bias	100×MSE	CP	%Bias	100×MSE	CP
		σ_ϵ^2	1.3141	0.0551	0.950	0.0941	0.0475	0.956
	Surv	$\log\lambda$	2.5892	8.0840	0.956	2.3658	7.7542	0.958
		τ	2.9776	0.2473	0.950	2.8237	0.2394	0.956
		γ_1	2.2858	2.4745	0.944	2.1673	2.4414	0.942
		γ_2	0.4531	0.2415	0.962	0.2851	0.2341	0.962
		φ	11.5990	1.1565	0.876	5.2110	0.7503	0.914
Gamma	Long	β_0	0.1770	0.5285	0.920	0.1258	0.4311	0.934
		β_1	1.1329	0.0062	0.920	0.9569	0.0055	0.922
		β_2	0.1617	0.0526	0.972	0.2429	0.0433	0.958
		β_3	0.3894	0.0842	0.954	0.8621	0.0594	0.944
		σ_v^2	11.6540	1.9978	0.614	7.5705	1.2889	0.794
		σ_ϵ^2	1.9810	0.0614	0.952	0.1123	0.0465	0.962
	Surv	$\log\lambda$	2.9267	8.5055	0.946	2.5556	7.9383	0.952
		τ	2.9849	0.2482	0.940	2.9634	0.2451	0.934
		γ_1	0.7488	2.3576	0.956	0.6135	2.3270	0.956
		γ_2	2.8847	0.2633	0.946	2.3389	0.2466	0.948
		φ	12.9543	1.2985	0.890	5.9240	0.7856	0.930

4 Application: GenIMS Data Analysis

Here we present an application of the proposed method using the GenIMS data introduced in Section 1. The study enrolled a total of 2320 patients aged 18 years or older who were diagnosed with the community-acquired pneumonia (CAP). This large observational cohort study was conducted in western Pennsylvania, Connecticut, southern Michigan, and western Tennessee in the USA through the emergency departments of 28 academic and community hospitals during December 2001 – November 2003. Although 2320 individuals were enrolled initially, after discharge from emergency departments and changes in primary diagnosis, the inpatient CAP cohort consists of a group of 1886 patients. During the study, the cohort recorded a total of 212 deaths. Details about the study can be found in Kellum et al. (2007).

A primary objective of the GenIMS study was to investigate the association between the risk factors (covariates) and 90-day mortality of the patients. Blood samples were taken from the patients immediately after their enrollment in the study, daily for the first 7 days and weekly thereafter while the patient remained in the hospital. In this study, survival outcomes T_i were obtained as

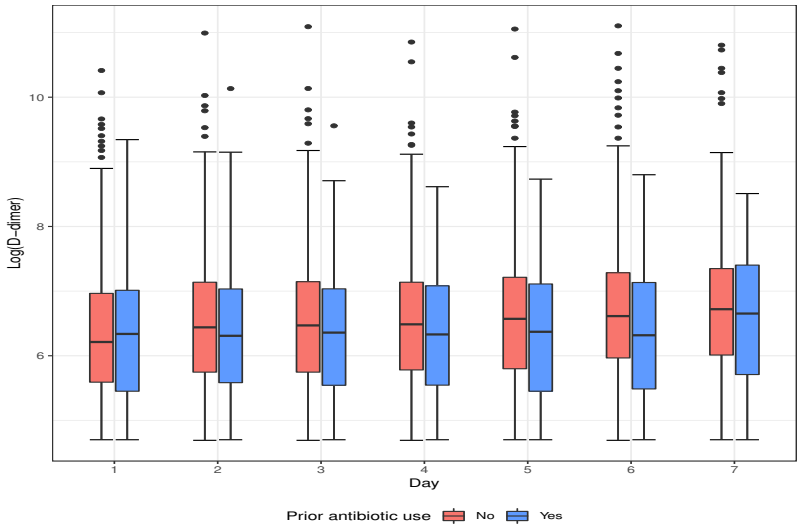

Figure 1: Logarithm of D-dimer levels of sepsis patients from the GenIMS study. Boxplots of overall D-dimer levels are shown for seven hospital days (Day 1–7)

the 90-day mortality of the pneumonia patients, whereas longitudinal measurements y_{ij} on a D-dimer biomarker were taken from each patient daily for a period of up to seven days. The baseline covariates considered for the analysis include "Age10" (age in years/10), "Sex" (0=male, 1=female), and binary indicators of "Charlson", "Ant7pres", "Statin", and "Anticoa".

Figure 1 displays boxplots of the logarithm of the longitudinal D-dimer levels of the sepsis patients for seven hospital days (Day 1–7). The D-dimer levels are shown for two groups of patients divided according to their prior antibiotics use (Ant7pres: yes or no). The boxplots indicate that the D-dimer levels are generally low for patients taking the antibiotics.

For the longitudinal measurements y_{ij}, we use a linear mixed model with a random intercept term, given by

$$y_{ij} = \beta_0 + \beta_1 \text{Day}_{ij} + \beta_2 \text{Age10}_i + \beta_3 \text{Sex}_i + \beta_4 \text{Ant7pres}_i + \beta_5 \text{Anticoa}_i + \beta_6 \text{Statin}_i + v_i + \epsilon_{ij}, \quad (4.1)$$

for $i = 1, \ldots, N$, where the random errors ϵ_{ij} are assumed to be independently distributed as $N(0, \sigma_\epsilon^2)$. The intercept random effects v_i are assumed to be independent and follow a distribution with mean 0 and variance σ_v^2, independent of ϵ_{ij}.

For the survival times T_i, we use a frailty model in the framework of the Weibull proportional hazards model with the scale parameter λ and shape parameter τ, given by

$$h_i(T_i|v_i, \gamma, \lambda, \tau, \varphi) = \lambda\tau T_i^{\tau-1} \exp(\gamma_1 \text{Age10}_i + \gamma_2 \text{Sex}_i + \gamma_3 \text{Charlson}_i + \gamma_4 \text{Ant7pres}_i + \gamma_5 \text{Statin}_i + \varphi v_i), \quad (4.2)$$

where h_i is the hazard function for the ith patient.

Table 3 presents the maximum likelihood estimates, their standard errors, z-values, and 95% lower and upper confidence limits (LL and UL) of the model parameters obtained by the naive method assuming normal random effects and also by the proposed robust method assuming skew-normal random effects. The estimates of the covariate effects and variance components obtained by the two methods are somewhat similar. However, the standard errors of the estimates in the longitudinal model as obtained by the skew-normal approach are generally smaller than those obtained by the normal approach. For example, when estimating $(\beta_3, \beta_4, \beta_5, \beta_6)$, the skew-normal approach provides corresponding standard errors of the estimates as $(0.0630, 0.0790, 0.0651, 0.0775)$, whereas the normal approach provides corresponding standard errors of $(0.0641, 0.0810, 0.0669, 0.0803)$. It is also worth noting that the estimate of the skewness parameter δ appears to be positive, $\hat{\delta} = 2.051$, with a standard error of 0.4047, which justifies the use of the skew-normal distribution for the random effects.

From the joint analysis, it appears that the D-dimer levels increase over time, are higher for older patients, and decrease with the use of antibiotics. From the survival model, it appears that older patients, males, and patients with larger Charlson score are at higher risk of death from pneumonia. The treatments Ant7pres and Statin appear to reduce the risk of death. Also, the frailty variance component σ_v^2 appears to be significant with an estimate of 0.8535 and a corresponding standard error of 0.0467. The association parameter φ also appears to be highly significant, which justifies the use of the joint model for analyzing the two outcome processes.

5 Discussion

The purpose of this paper was to provide a suitable robust alternative to the naive normal random effects approach to estimating parameters in joint models. We have studied both asymptotic and finite-sample properties of the estimates under misspecified random effects. Our Monte Carlo study shows that the proposed skew-normal random effects approach provides estimates that are generally more efficient than those obtained under the naive normal random effects approach. The GenIMS data analysis presented earlier shows that the estimates of the model parameters are similar by the two methods, but the skew-normal approach provides smaller standard errors as compared to the normal approach. As an exploratory data analysis, one can fit the given data using both methods. If the results are similar by the two methods, then normal random effects may be used for simplicity.

In our numerical analysis, we have considered a univariate shared random effect term for which it is not so difficult to calculate the joint likelihood, score function and Fisher information using a numerical integration method. For multidimensional random effects, however, an exact likelihood analysis requires intensive computation involving irreducibly high-dimensional integrals. We intend to develop an approximate likelihood method to reduce the computational burden in a future study.

Table 3: Joint analysis of pneumonia-to-death and longitudinal D-dimer outcomes from GenIMS study under normal and skew-normal random effects distributions.

	Fitted model: Normal					Fitted model: Skew-normal				
Variable	Est.	SE	z-value	LL	UL	Est.	SE	z-value	LL	UL
Longitudinal model										
Intercept (β_0)	6.0570	0.1517	39.927	5.7597	6.3543	6.0210	0.1509	39.900	5.7252	6.3168
Day (β_1)	0.0220	0.0040	5.467	0.0142	0.0298	0.0219	0.0040	5.429	0.0141	0.0297
Age10 (β_2)	0.0579	0.0201	2.874	0.0185	0.0973	0.0638	0.0201	3.175	0.0244	0.1032
Sex (β_3)	–0.0474	0.0641	–0.740	–0.1730	0.0782	–0.0463	0.0630	–0.735	–0.1698	0.0772
Ant7pres (β_4)	–0.1641	0.0810	–2.027	–0.3229	–0.0053	–0.1784	0.0790	–2.259	–0.3332	–0.0236
Anticoa (β_5)	0.0214	0.0669	0.320	–0.1097	0.1525	0.0102	0.0651	0.157	–0.1174	0.1378
Statin (β_6)	–0.0496	0.0803	–0.617	–0.2070	0.1078	–0.0602	0.0775	–0.777	–0.2121	0.0917
σ_v^2	0.8467	0.0428	19.785	0.7628	0.9306	0.8535	0.0467	18.258	0.7620	0.9450
σ_ϵ^2	0.1836	0.0046	40.278	0.1746	0.1926	0.1835	0.0046	40.290	0.1745	0.1925
Survival model										
Scale (log λ)	–8.7627	0.7750	–11.306	–10.2817	–7.2437	–8.7841	0.7760	–11.320	–10.3051	–7.2631
Shape (τ)	0.6767	0.0658	10.279	0.5477	0.8057	0.6764	0.0658	10.278	0.5474	0.8054
Age10 (γ_1)	0.4424	0.0874	5.062	0.2711	0.6137	0.4464	0.0875	5.101	0.2749	0.6179
Sex (γ_2)	–0.3759	0.2160	–1.741	–0.7993	0.0475	–0.3849	0.2160	–1.782	–0.8083	0.0385
Charlson (γ_3)	0.6973	0.2752	2.534	0.1579	1.2367	0.6961	0.2754	2.527	0.1563	1.2359
Ant7pres (γ_4)	–0.4368	0.3170	–1.378	–1.0581	0.1845	–0.4402	0.3165	–1.391	–1.0605	0.1801
Statin (γ_5)	–0.5520	0.2874	–1.921	–1.1153	0.0113	–0.5515	0.2863	–1.926	–1.1126	0.0096
Association (φ)	0.6895	0.0983	7.017	0.4968	0.8822	0.6759	0.0959	7.046	0.4879	0.8639
Skewness (δ)						2.0541	0.4047	5.076	1.2609	2.8473

Acknowledgments

The research of Sanjoy K. Sinha is partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors thank Dr. Derek Angus and the Clinical Research, Investigation, and Systems Modeling of Acute Illness Center laboratory for access to the GenIMS data. The authors are indebted to the nurses, respiratory therapists, phlebotomists, physicians, and other health care professionals, as well as the patients and their families who supported this trial. A complete list of GenIMS investigators is available at www.ccm.upmc.edu/genims investigators. The GenIMS study was funded via grant R01 GM61992 by the National Institute of General Medical Sciences. No conflict of interest is declared.

References

Alam, K., Maity, A., Sinha, S. K., Rizopoulos, D., and Sattar, A. (2021), "Joint modeling of mixed longitudinal and time-to-event outcomes," *Lifetime Data Analysis*, 27, 64–90.

Arellano-Valle, R. B., Ozan, S., Bolfarine, H., and Lachos, V. H. (2005), "Skew normal measurement error models," *Journal of Multivariate Analysis*, 96, 265–281.

Chen, L. M., Ibrahim, J. G., and H, C. (2011), "Sample size and power determination in joint modeling of longitudinal and survival data," *Statistics in Medicine*, 30, 2295–2309.

Diggle, P. J., Henderson, R., and Philipson, P. (2009), "Random effects models for joint analysis of repeatedmeasurement and time-to-event outcomes," *In Longitudinal Data Analysis (eds G. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs), pp. 349–366. Boca Raton: Chapman and Hall–CRC*.

Furgal, A. K. C., Sen, A., and Taylor, J. M. G. (2019), "Review and comparison of computational approaches for joint longitudinal and time-to-event models," *International Statistical Review*, 87, 393–418.

Gould, A. L., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., and Bois, F. Y. (2015), "Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group," *Statistics in Medicine*, 34, 2181–2195.

Heagerty, P. J. and Kurland, B. F. (2001), "Misspecified maximum likelihood estimates and generalised linear mixed models," *Biometrika*, 88, 973–985.

Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2016), "Joint modelling of time-to-event and multivariate longitudinal outcomes: Recent developments and issues," *BMC Medical Research Methodology*, 16.

Hogan, J. W. and Laird, N. M. (1998), "Increasing efficiency from censored survival data by using random effects to model longitudinal covariates," *Statistical Methods in Medical Research*, 7, 28–48.

Kellum, J. A., Kong, L., Fink, M. P., Weissfeld, L. A., Yealy, D. M., Pinsky, M. R., Fine, J., Krichevsky, A., Delude, R. L., and Angus, D. C. (2007), "Understanding the inflammatory cytokine response in pneumonia and sepsis: results of the genetic and inflammatory markers of sepsis (GenIMS) study," *Archives of Internal Medicine*, 167, 1655–1663.

Kim, S. and Albert, P. S. (2016), "A class of joint models for multivariate longitudinal measurements and a binary event," *Biometrics*, 72, 917–925.

Li, N., Elashoff, R. M., Li, G., and Tseng, C. H. (2012), "Joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times with nonparametric distributions for random effects," *Statistics in Medicine*, 31, 1707–1721.

Litire, S., Alonso, A., and Molenberghs, G. (2008), "The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models," *Statistics in Medicine*, 27, 3125–3144.

McCulloch, C. E. and Neuhaus, J. M. (2011), "Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter," *Statistical Science*, 26, 388–402.

Neuhaus, J. M., Hauck, W. W., and Kalbfleisch, J. D. (1992), "The effects of mixture distribution misspecification when fitting mixed-effects logistic models," *Biometrika*, 79, 755–762.

Neuhaus, J. M., McCulloch, C. E., and Boylan, R. (2013), "Estimation of covariate effects in generalized linear mixed models with a misspecified distribution of random intercepts and slopes," *Statistics in Medicine*, 32, 2419–2429.

Pantazis, N. and Touloumi, G. (2007), "Robustness of a parametric model for informatively censored bivariate longitudinal data under misspecification of its distributional assumptions: A simulation study," *Statistics in Medicine*, 26, 5473–5485.

Rastegaran, A. and Zadkarami, M. R. (2015), "A skew-normal random effects model for longitudinal ordinal categorical responses with missing data," *Journal of Applied Statistics*, 42, 114–126.

Rizopoulos, D. (2015), "Joint modeling of longitudinal and time-to-event data," *In Handbook of Missing Data Methodology (eds G. Molenberghs, G. Fitzmaurice, M. Kenward, A. Tsiatis, and G. Verbeke), pp.117-136. Boca Raton: Chapman and Hall–CRC*.

Sattar, A. and Sinha, S. K. (2017), "Joint modeling of longitudinal and survival data with a covariate subject to a limit of detection," *Statistical Methods in Medical Research*, 28, 486–502.

Sattar, A., Sinha, S. K., Wang, X., and Li, Y. (2015), "Frailty models for pneumonia to death with a left-censored covariate," *Statistics in Medicine*, 34, 2266–2280.

Tsiatis, A. and Davidian, M. (2004), "Joint modeling of longitudinal and time-to-event data: an overview," *Statistica Sinica*, 14, 809–834.

Verbeke, G. and Davidian, M. (2009), "Joint models for longitudinal data: Introduciton and overview," *In Longitudinal Data Analysis, Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds). CRC Press: New York*.

Verbeke, G. and Lesaffre, E. (1997), "The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data," *Computational Statistics and Data Analysis*, 23, 541–556.

White, H. (1982), "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–25.

Received: March 2, 2021

Accepted: April 24, 2021