Noname manuscript No.
(will be inserted by the editor)

Simultaneous Bayesian modelling of skew-normal longitudinal measurements with non-ignorable dropout

Received: date / Accepted: date

Abstract Most often in genetic improvement studies, repeated measurements are observed on an individual animal, and these repeated measurements are often skewed. From the practical viewpoint, logarithm transformations of variables are usually adopted to reduce skewness, and this works satisfactorily in many cases. In most longitudinal datasets, however, because of the high rate of missingness, skewness often remains after transformation, the achievement of joint normality for each component of separately transformed variables, which are often difficult to interpret, is unrealistic. For this purpose, a more general form of distributions for considering skewness in the model should be used. In this paper, we used Bayesian joint modelling of longitudinal and survival data when data set presents skewness. A skew-normal mixed-effects model for longitudinal measurements and a Cox proportional hazard model for time to event variable were considered. We performed some simulation studies to investigate the performance of the proposed method to skewness in random effects, different dropout rates and sample sizes. Furthermore, we illustrated the proposed method using Nigerian indigenous chickens (NIC) dataset. The longitudinal outcomes of NIC data set were skewed, and presented left censored dropout. We assumed different model structures for the analysis of this data set and considered two versions of the deviance information criteria (DIC): namely, the conditional criteria (given the random effects) and marginal criteria (averaging over the random effects) in selecting the "true" model. These criteria were computed using the importance sampling method.

Keywords Joint mixed linear model · importance sampling · marginal deviance · repeated measurements · time-to-event

Address(es) of author(s) should be given

1 Introduction

In many longitudinal studies, such as genetic improvement programme (GIP), subjects (e.g., chicks, piglets etc.) are followed-up repeatedly over a period of time and response data collected, for example, an animal geneticist might be interested in how different breeds of animal gain weight over a period of time. These repeated measurements require follow-up which may be stopped by a dependent terminal event (such as death and or loss to follow-up) whose probability of occurrence is non-ignorable. Modelling the repeated measures and event-time outcomes separately as previously done in Laird and Ware (1982); Cox (1972) has been questioned in the literature (Ibrahim et al.; 2010) as this method is inefficient and can introduce bias to effect size estimates for two correlated outcomes (Hickey et al.; 2016).

Joint modelling of longitudinal and survival data have received more attention during the past two decades (Ibrahim et al.; 2010; Rizopoulos; 2012; Wulfsohn and Tsiatis; 1997; Henderson et al.; 2000; Mchunu et al.; 2020; Li and Su; 2017; Chan and Grant; 2016; Zhang and Yuan; 2012) and recently Alsefri et al. (2020) gave methodological reviews of the joint modelling of longitudinal and time-to-event data. The joint modelling approach has been suggested as it constructs two sub-models for the longitudinal and the event time data, linked by a set of subject-specific random effects (Wang et al.; 2002). Although the validity of normality assumption has been questioned in literature (Verbeke and Lesaffre; 1997), and its violation could lead to misleading inferences. However, several previous studies assumed either /or both the error terms in the models for the longitudinal response and the measurement errors in covariates follow normal distributions due to mathematical tractability and computational convenience.

Longitudinal outcomes such as animal body weight in GIP, CD4 counts in the HIV research are often skewed in practice. As such, in the joint modelling of longitudinal and time to event data, the skewness in the data should not be ignored. Several studies had looked at these issues, for example, Li et al. (2009) and Huang et al. (2010) proposed robust joint modelling where a student's t distribution in different structures of joint modelling of longitudinal and survival data was applied. Thus alternative flexible distributions, such as the multivariate skew-elliptical (SE) distribution, were proposed in Azzalini and Capitanio (1999); Ma et al. (2004). Huang and Dagne (2011) addressed the issues by jointly model the response and covariate process using a Bayesian approach to non-linear mixed effect models with covariate measurement errors and a skew-normal distribution.

Similarly, Baghfalaki and Ganjali (2015) discussed joint modelling of longitudinal and survival data when skewness exits in the data sets and used the multivariate skew-normal distribution approach of Sahu et al. (2003). The authors considered this method to be more flexible when using Bayesian approach. However, simultaneous modelling of longitudinal and survival data with skewness when the cause of death is experimentally related has not been given much attention, especially in the animal genetic improvement programme.

In this paper, we discussed the robust inference of Bayesian joint modelling of longitudinal and survival data with skewness. We focused on the scenario where the cause of death is experimentally related as we obtained in our motivation example. Also, a skew-normal mixed effect model and a Cox proportional hazard model (as a semiparametric model) with step baseline hazard in a frailty model structure are considered for the joint modelling. We used Bayesian approach and JAGS (Plummer et al.; 2003) for implementation of the models, where two versions of DIC (i.e. the marginal and conditional deviance information criteria) was used for model selection.
The choice of the two versions of DIC is motivated by the fact that with models allowing non-ignorable dropout, when the presence of dropout influences inference, we must take account of the missing data mechanism. Celeux et al. (2006) discussed different DIC definitions for missing data models, in the context of mixtures of distributions and random-effects models. The authors considered six versions of DICs, among which the conditional DIC (cDIC) and marginal DIC (mDIC) were studied. The cDIC (given the random effects) is popular among researchers given its easy computation. At the same time, the mDIC (integrating out) requires more computational efforts as it has not been implemented in any statistical software. Recently, efforts have been made (Ariyo, Quintero, Muñoz, Verbeke and Lesaffre; 2019; Ariyo, Lesaffre, Verbeke and Quintero; 2019) to compute the mDIC using dedicated R functions. However, this is not yet popular among the applied Statisticians. We employ a general Bayesian framework for estimating parameters in an asymmetric joint linear mixed model and survival time and compare its performance with the corresponding symmetric model using mDIC. The proposed approach will be investigated using simulation studies and Nigerian indigenous chickens (NIC) data set.
The plan of the paper is as follows. In Section 2, we introduced the multivariate skew-normal distribution used in this paper, described some models and notations and joint models for skew-normal for longitudinal data with dropout profile. Section 3 includes the Bayesian model selection; here, conditional and marginal DIC were discussed. In Section 4, simulation studies were conducted to assess the proposed model. At the same time, we applied the method to NIC data set and the results presented in Section 5. The concluding remarks are given in Section 6.

2 Methodology

2.1 Multivariate Skew-Normal Distribution

One of the commonly used multivariate skew-normal distributions, in the Bayesian context, was introduced by Sahu et al. (2003). An $n-$dimensional random vector $\mathbf{Y}$ follows an n-variate skew-normal (SN) distribution ($\mathbf{Y} \sim SN_{n,k}(\mu_{\mathbf{0}}, \mathbf{H}, \mathbf{\Delta})$) with location vector $\mu_{\mathbf{0}} \in \mathbb{R}^n$, scale matrix $\mathbf{H}$ (an $n \times n$ positive definite matrix) and $n \times k$ skewness matrix $\mathbf{\Delta}$, if density function is

given by

$$f\left(\mathbf{y}|\mu_{\mathbf{0}},\mathbf{H},\mathbf{\Delta}\right)=2^{k}\phi_{n}\left(\mathbf{y}|\mu_{\mathbf{0}},\mathbf{H}+\mathbf{\Delta}\mathbf{\Delta}^{T}\right)\times$$
$$\Phi_{k}\left(\mathbf{\Delta}^{T}\left(\mathbf{H}+\mathbf{\Delta}\mathbf{\Delta}^{T}\right)^{-1}\left(\mathbf{y}-\mu_{\mathbf{0}}\right)|0,\left(\mathbf{I}_{k}+\mathbf{\Delta}^{T}\mathbf{H}^{-1}\mathbf{\Delta}\right)^{-1}\right), \tag{1}$$

where $\phi_n(\mathbf{y}|\mu,\mathbf{H})$ and $\Phi_n(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{H})$ are respectively the density function and the cumulative distribution functions of the multivariate normal distribution $N_n(\mu,\mathbf{H})$ obtained from $\boldsymbol{y}$. When $\boldsymbol{\mu}=\mathbf{0}$ and $\mathbf{H}=\mathbf{I_n}$ (the $n\times n$ identity matrix), we denote these functions as ϕ_n and Φ_n. If we substitute $\mathbf{\Delta}=\mathbf{0}$, equation (1) reduces to the usual symmetric multivariate normal distribution $N_n(\mu_{\mathbf{0}},\mathbf{H})$. The mean and covariance matrix are given by $E(\boldsymbol{Y})=\boldsymbol{\mu_0}+\sqrt{\frac{2}{\pi}}\boldsymbol{\delta}$, $\mathrm{Var}(Y)=\mathbf{H}+\left(1-\frac{2}{\pi}\right)\Delta^2$, where $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_n)^T$ is a skewness parameter vector (Huang and Dagne; 2011). If $\boldsymbol{\delta}=0$, the density (1) reduces to the usual multivariate normal distribution, whereas for positive values of $\boldsymbol{\delta}$ we obtain a positively (right) skewed distribution and for negative values we obtain a negatively (left) skewed distribution.

2.2 Models and Notation

The linear mixed model has gained tremendous attention for modelling longitudinal data. For a longitudinal model, we denote $\mathbf{Y_i}$ to be $(n_i\times 1)$ vector of responses for the i^{th} subject at times $t_{i1},\ldots,t_{in_i}$, where $i=1,2,\ldots,n$. We consider the following linear mixed model

$$y_{ij}=\boldsymbol{x}_{1i}^{\boldsymbol{T}}\boldsymbol{\beta}_1+\boldsymbol{z}_{\mathbf{1}\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{b}_{1i}+\epsilon_{ij},\quad i=1,2,\ldots n,\quad j=1,2,\ldots,n_i, \tag{2}$$

where components of $\boldsymbol{\epsilon}_i=(\epsilon_{i1},\ldots,\epsilon_{in_i})^T$ are measurements errors, $\boldsymbol{\beta}_1$ is a $p-$dimensional vector of fixed effects parameters. $\boldsymbol{b}_{1i}$ is a $q-$ dimensional vector of random effects independent of $\boldsymbol{\epsilon_i}$. A standard assumption is that the random effects $\mathbf{b}_{1i}$ and the residual component $\boldsymbol{\epsilon_i}$ have a normal distribution, i.e

$$\boldsymbol{b}_{1i}\sim N_q(\mathbf{0},\mathbf{G}),\epsilon_i\sim N_{n_i}(0,\Psi)$$

where $\mathbf{G}=\mathbf{G}(\alpha)$ and $\boldsymbol{\Psi}_i=\boldsymbol{\Psi}_i(\gamma)$, $(i=1,\ldots,n)$, are dispersion matrices, depending on parameters α and γ respectively. Although model (2) offers great flexibility for modelling the within-subjects correlation in longitudinal data, the model is robust for fixed effects as pointed out in Komarek and Lesaffre (2008) but suffers from lack of robustness against departures from distributional assumptions for random effects (Arellano-Valle et al.; 2007). We further assume that $\boldsymbol{\epsilon_i}\sim SN_{n_i}\left(\mathbf{0},\boldsymbol{\Psi},\boldsymbol{\Delta}_\epsilon\right)$ and $\boldsymbol{b_{1i}}\sim SN_q\left(\mathbf{0},\mathbf{G},\boldsymbol{\Delta_b}\right)$. To seek for differentiability (Arellano-Valle et al.; 2007), we assume $\Psi_i=\sigma^2_{\epsilon_i}I_{n_i}$, $\Delta_{\epsilon_i}=\delta_{\epsilon_i}I_{n_i}$.

For survival model, let T_i^* be the true event and C_i be the censoring time which may be informative or non-informative (see Rizopoulos; 2012, for details). As was the case in our motivating example, survival of chicken at the

end of twenty weeks was considered an "event". However, some chickens died before the "event". This is often refer to the left censoring as some chicken do not survival to the ends of the study. The true event $T_i = min(T_i^*, C_i)$ represent the estimated survival time for the i^{th} individual.

Let δ_i^* denote a censoring indicator such that;

$$\delta_i^* = \begin{cases} 0 & \text{if} \quad T_i^* = C_i \quad \text{Right censoring} \\ 1 & \text{if} \quad T_i^* > C_i \quad \text{Left censoring} \ . \end{cases}$$

Therefore, the observed data for the outcome consist of the pairs

$$\left\{ \left(T_i, \delta_i^* \right), i = 1, 2, \ldots, n \right\}.$$

We considered a frailty model which is linked to the longitudinal model through some shared random effects in survival modelling. Let $\mathbf{x}_i$ and $\mathbf{z}_i$ be p_2 and q_2 - dimensional vectors of covariates respectively. As such, the hazard function in proposed model is given as:

$$h\left(t_i | \mathbf{x_{2i}}, \mathbf{z}_{2i}, \mathbf{b}_{2i}\right) = h_0(t_i) \exp\left\{ \mathbf{x}_{2i}^T \boldsymbol{\beta_2} + \mathbf{z}_{2i}^T \mathbf{b}_{2i} \right\},$$

where $h_0(t_i)$ is the baseline hazard function for the reference subject with all covariates equal to 0. $\boldsymbol{\beta}_2 = (\beta_{2i}, \ldots, \beta_{2p_2})^T$ is a p_2-dimensional vector of time to event fixed effect parameters and we assume $\mathbf{b}_{2i} \sim N_{q_2}(\mathbf{0}, \mathbf{D_2})$. Thus, the density function of survival time for the i^{th} individual is

$$h^{\delta_i^*}\left(\mathbf{t_i} | \mathbf{x}_{2i}, \mathbf{z}_{2i}, \mathbf{b}_{2i}\right) \times \exp\left\{ -H_0(t_i) \exp\left\{ \mathbf{x}_{2i}^T \boldsymbol{\beta}_2 + \mathbf{z}_{2i}^T \mathbf{b}_{2i} \right\} \right\}, \tag{3}$$

where $H_0(\boldsymbol{t_i}) = \int_0^t h_0(u) du$ and the covariates are assumed to be time independent. The baseline hazard appears in the likelihood (3) and so must be estimated.

The repeated outcome $\boldsymbol{y_i}$ can be partitioned into

$$\boldsymbol{y_{i,obs}} = \left\{ y_i(t_{ij}) : t_{ij} < T_i, j = 1, 2, \ldots, n_i \right\},$$

which contains all observed measurements for the i^{th} individual before dropout occurs at T_i and $\boldsymbol{y_{i,mis}} = \left\{ y_i(t_{ij}) : t_{ij} \geq T_i, j = 1, 2, \ldots, n_i \right\}$ which contains the measurement of individuals that should have been taken until the end of the study.

2.3 Joint models for skew-normal longitudinal data with dropout

The skew-normal joint modelling of longitudinal and dropout process, as an extension of the usual normal joint modelling, leads to the following hierarchical model (see also Sahu et al.; 2003; Arellano-Valle et al.; 2007; Huang and Dagne; 2011; Rizopoulos; 2012; Baghfalaki and Ganjali; 2015):

$$\begin{cases} \boldsymbol{Y_i} | \boldsymbol{b}_{1i}, \boldsymbol{\beta_1}, \sigma_\epsilon^2, \delta_\epsilon \sim SN_{n_i}\left(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \boldsymbol{Z}_{1i}\boldsymbol{b}_{1i} - \sqrt{\frac{2}{\pi}}\delta_\epsilon 1_{n_i}, \sigma_\epsilon^2 \boldsymbol{I}_{n_i}, \delta_\epsilon \boldsymbol{I}_{n_i} \right) \\ \boldsymbol{b}_{1i} \sim SN_q\left(-\sqrt{\frac{2}{\pi}}\delta_b, \Sigma_b, \Delta_b \right), \end{cases} \tag{4}$$

$$\begin{cases} h\left(t_i|\boldsymbol{x_{2i}}, \boldsymbol{z}_{2i}, \mathbf{b}_{2i}\right) = h_0(t)\exp\left\{\boldsymbol{x}_{2i}^T\boldsymbol{\beta}_2 + \boldsymbol{z}_{2i}^{\boldsymbol{T}}\boldsymbol{b}_{2i}\right\}, \\ \boldsymbol{b}_{2i} \sim SN_q\left(-\sqrt{\frac{2}{\pi}}\delta_b, \boldsymbol{\Sigma_b}, \boldsymbol{\Delta_b},\right), \end{cases}$$

where some components of $\mathbf{b}_i$ are "shared parameter" that serves to induce a correlation between both models through their joint dependence on $\mathbf{b}_i$. This implies that both models are conditionally independent given $\mathbf{b}_i$ may be interpreted as reflecting the belief that a common set of underlying characteristics of the individuals governs both outcomes processes. The main advantages of this approach is that both models do not have to be of the same type (see also Rizopoulos; 2012; Fitmaurice et al.; 1995; Bogaerts et al.; 2017, for overview). The key component behind the dropout/attrition mechanism considered in (4) is the random effects $\mathbf{b}_i$ since the survival and longitudinal submodels share the same $\mathbf{b}_i$.

We simplified the model by consider the stochastic representation of the skew-normal distribution for Markov Chain Monte Carlo (MCMC) approach in the Bayesian specification (Arellano-Valle et al.; 2007; Baghfalaki and Ganjali; 2015). Therefore, (4) can be written as:

$$\begin{cases} \boldsymbol{Y_i}|\mathbf{b}_{1i}, \boldsymbol{\beta_1}, \sigma_\epsilon^2, \mathbf{U}_i = \mathbf{u_i} \sim N_{n_i}\left(\mathbf{X}_{1i}\boldsymbol{\beta}_1 + \mathbf{Z}_{1i}\mathbf{b}_{1i} + \boldsymbol{\delta}_\epsilon u_i - \sqrt{\frac{2}{\pi}}\delta_\epsilon \mathbf{1}_{n_i}, \sigma_\epsilon^2\mathbf{I}_{n_i}\right), \\ \mathbf{U}_i \sim N_{n_i}\left(\mathbf{0}, \mathbf{I}_{n_i}\right) I(\mathbf{u}_i > 0), \end{cases}$$

where $\mathbf{u}_i$ is the observed value of $\mathbf{U}_i$ before dropout. Analysis of longitudinal data is often impeded by the presence of missing data which may be due to subject non-response, loss to follow-up or death of subject. This problem if not handle with appropriate statistical model may lead to biased estimates and loss of precision. The missing data can be classified as ignorable and non-ignorable. The reader should see Rubin (1976) for formal definitions. We considered a non-ignorable missingness mechanism in this article (see also Rubin; 1976, for formal definitions).

In Bayesian modelling, the prior distribution for unknown parameters $\boldsymbol{\Theta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \Sigma_b, \Delta_b, \delta_\epsilon, \delta_b, \sigma_b^2, \boldsymbol{\sigma}_\epsilon^2)$ should be defined. Variance components are often modelled with improper prior in hierarchical linear mixed model either due to lack of prior information or simply for convenience. Hobert and Casella (1996) observed that the posterior distributions for the hierarchical linear mixed model are rarely available in a closed form and they proposed proper but diffuse conditionally conjugate priors to ensure painless calculation of the Gibbs samplings. We assume that the components of $\boldsymbol{\Theta}$ are mutually independent and the prior distributions are given as

$$\begin{cases} \boldsymbol{\beta_1} \sim N_{p_1}\left(\beta_{01}, \Sigma_{01}\right), \boldsymbol{\beta_2} \sim N_{p_2}\left(\beta_{02}, \Sigma_{02}\right), \\ \boldsymbol{\Sigma}_{b_1 q_1}\left(\eta_{01}, \psi_{01}\right), \Sigma_{b_2} \sim IW_{q_2}\left(\eta_{02}, \psi_{02}\right), \\ \sigma_\epsilon^2 \sim IG\left(\alpha_{01}, \tau_{01}\right), \sigma_{\boldsymbol{b}}^2 \sim IG\left(\alpha_{02}, \tau_{02}\right), \\ \delta_b \sim N_q\left(\mu_b, \gamma_b\right) \boldsymbol{I}\left\{\delta_b > 0\right\}, \delta_\epsilon \sim N\left(\mu_{\delta\epsilon}, \sigma_{\delta_\epsilon}^2\right), \end{cases} \tag{5}$$

and these representations are important because they allow easy implementation in Bayesian software like BUGS. The full conditional distributions required to implement the Gibbs sampler are straightforward to derive and

sample from according to the specified equations (5). In case of no previous information, values for the hyperparameters can be chosen so that the prior distributions are vague. The integral for all this prior are equal to 1, making them proper prior and variance at talking to be large to express low previous information.

3 Bayesian model selection

There exists a wide variety of model selection/assessment measures within the Bayesian toolbox. To select the "best" model among the competing models, we used two versions of deviance information criteria (DIC). Spiegelhalter et al. (2002) introduced and developed the concept of DIC which begin with the definition of Bayesian measures of model complexity. The criterion is based on the *deviance*, defined as $D(\boldsymbol{\phi}) = -2\log p(\boldsymbol{y}|\boldsymbol{\phi}) + 2\log h(\boldsymbol{y})$, where $\boldsymbol{\phi}$ corresponds to the parameters in focus, $p(\boldsymbol{\phi}|\boldsymbol{y})$ is the likelihood function and $h(\boldsymbol{y})$ is an estimator of $\boldsymbol{\phi}$ that depends on data only. The effective number of parameters pD is given as

$$pD = -2\boldsymbol{E}_{\boldsymbol{\phi}}[\log p(\boldsymbol{y}|\boldsymbol{\phi})|\boldsymbol{y}] + 2\log p(\boldsymbol{y}|\hat{\boldsymbol{\phi}}),$$

$\hat{\boldsymbol{\phi}}$ is an estimate of $\boldsymbol{\phi}$ which is usually taken as the posterior mean or mode. The DIC may be viewed as a trade off between model adequacy and complexity. In models with latent variables, as explored by Celeux et al. (2006), there are several versions of DIC, specifically: the conditional DIC (cDIC) that incorporates the latent variables in the focus of the analysis and the marginal DIC (mDIC) which integrates them out. For example, suppose we add an additional vector of latent variables $\boldsymbol{\mu}$ with density $p(\boldsymbol{\mu}|\boldsymbol{\phi})$ to the model $p(\boldsymbol{y}|\boldsymbol{\phi})$. Then we have

$$p(\boldsymbol{y}|\boldsymbol{\phi}) = \int p(\boldsymbol{y}|\boldsymbol{\phi},\boldsymbol{\mu})p(\boldsymbol{\mu}|\boldsymbol{\phi})d\boldsymbol{\mu} = \int p(\boldsymbol{y},\boldsymbol{\mu}|\boldsymbol{\phi})d\boldsymbol{\mu}, \tag{6}$$

where $p(\boldsymbol{y}|\boldsymbol{\phi},\boldsymbol{\mu})$ is the conditional likelihood and $p(\boldsymbol{y}|\boldsymbol{\phi})$ is the integrated likelihood. The integrated likelihood in equation (6) yields marginalised DIC (mDIC). It naturally follows that the definition of mDIC from integrated likelihood is given as

$$mDIC = -4\boldsymbol{E}_{\boldsymbol{\phi}}[\log p(\boldsymbol{y}|\boldsymbol{\phi})|\boldsymbol{y}] + 2\log p(\boldsymbol{y}|\tilde{\boldsymbol{\phi}}), \tag{7}$$

where the estimate $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ is set of posterior mode $\tilde{\boldsymbol{\phi}}$. As noted by Chan and Grant (2016), it is clear from equation (7) that the mDIC depends on the prior only through its effect on the posterior distribution. Latent variables structure are often chosen so that the conditional likelihood $p(\boldsymbol{y}|\boldsymbol{\phi},\boldsymbol{\mu})$ is available in closed forms. Therefore, the alternative definition of DIC via conditional likelihood (cDIC) is given as

$$cDIC = -4\boldsymbol{E}_{\boldsymbol{\phi},\boldsymbol{\mu}}[\log p(\boldsymbol{y}|\boldsymbol{\phi},\boldsymbol{\mu})|\boldsymbol{y}] + 2\log p(\boldsymbol{y}|\tilde{\boldsymbol{\phi}},\tilde{\boldsymbol{\mu}}),$$

where $(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\mu}})$ is the joint maximum a posterior estimate of the pair $(\boldsymbol{\mu}, \boldsymbol{\phi})$ given the data $\boldsymbol{y}$ (see Celeux et al.; 2006). Computational difficulties are the major challenge of mDIC since the integral in equation (7) is generally intractable, and numerical analysis appears to be the solution. Although its computational difficulties notwithstanding, mDIC has been found to outperform its conditional counterpart in most specific conditions (Chan and Grant; 2016; Quintero and Lesaffre; 2017; Ariyo, Quintero, Muñoz, Verbeke and Lesaffre; 2019; Ariyo, Lesaffre, Verbeke and Quintero; 2019). In this work, we compute the mDIC using importance sampling to integrate out the latent variable $\boldsymbol{\mu}$. The details are available as supplementary materials to this article.

4 Simulation Study

We conducted a simulation study to illustrate the performance of the proposed method. This simulation study aims to investigate the impact on parameter inference when the assumption of normality is inappropriate, and when the dropout is related to the experiment. Also, we investigated whether the model selection measures, viz., the conditional/marginal DIC determines the best-fitting model to the simulated data. In this study, we generated 1000 samples data from the following joint model

$$y_{ij} = \beta_0 + \beta_1 k_{ij} + \beta_2 x_i + b_{1i} + b_{2i} k_{ij} + \epsilon_{ij}, i = 1, 2, \ldots, 1000, j = 1, 2, \ldots, 21. \quad (8)$$

We set the values of the parameters as: $k_{ij} = j$, $x_i \sim Ber(0.5)$, $\beta_0 = 5$, $\beta_1 = -2$, $\beta_2 = -1$, $\epsilon_{ij} \sim \left(-\sqrt{\frac{2}{\pi}}\delta_\epsilon, \sigma_\epsilon^2, \delta_\epsilon\right)$ where $\sigma_\epsilon = 2$ and $\delta_\epsilon = 3$. Also, we used a Cox proportional hazard model in a frailty structure with a Weibull baseline hazard (Vaida and Xu; 2000),

$$h(t) = h_0(t) exp\left\{\beta_{01} + \beta_{11} x_i + \rho_1 b_{1i} + \rho_2 b_{2i}\right\}. \quad (9)$$

The observed dropouts were simulated using a time-independent hazard dropout model in equation (9). In this simulation, we considered three rates of random dropout which were generated by using different values for $\beta_{01} = 3, -1$ and -2. The values of β_{01} were chosen to produce a proportion of missing equal to approximately 10%, 30% and 50%. The selection of these values is similar to that of Baghfalaki and Ganjali (2015); Todem et al. (2010) as motivated by Molenberghs et al. (1997); Diggle and Kenward (1994). The non-ignorable mechanism was taken to be $k_{ij} > T_i$ when i^{th} individual dropout of the study. Also, $\mathbf{b}_i \sim N_2(\mathbf{0}, \mathbf{G})$ where $\mathbf{G}$ to be 2×2 matrix with diagonal equal to 1 and off diagonal values were set to be 0.5. To check the effect of sample size, we chose $n = 100$ to represent moderate large sample size and $n = 500$, as large sample size. In the simulation study, 1000 Monte Carlo data sets were simulated from equation (8) and (9) to fit the data set using *rjags* and the following vague prior specifications

$$\beta_0 \sim N(0, 100), \beta_1 \sim N(0, 100), \beta_2 \sim N(0, 100), \sigma_\epsilon^2 \sim IG(0.01, 0.02),$$

$$\sigma_b^2 \sim IG(0.01, 0.01), \delta_b \sim N(0, 100)IG\left\{\delta_b > 0\right\},$$
$$\boldsymbol{G} \sim IW(100\boldsymbol{I}_2, 2), p_k \sim N(0, 100), k = 1, 2.$$

All model parameters in the simulation studies were estimated based on three chains of 70,000 iterations after discarding the first 30,000 iterations. The thinning factor was set at 7 to avoid correlation problems in the generated chains. When Brooks-Gelman-Rubin (BGR) statistic (Brooks and Gelman; 1998; Gelman et al.; 1992) was larger than 1.1, further sampling was performed until BGR < 1.1. We analysed the simulated data under different model assumption that we represented with alphabet A, B and C. The model assumption for estimated models are given below:

$$\begin{cases} A : b_i \sim N_q(0, G) \text{ and } \epsilon_i \sim N_{n_i}(0, \Psi), \\ B : b_i \sim SN_q\left(-\sqrt{\frac{2}{\pi}}\delta_b, \sigma_b, \delta_b\right) \text{and} \quad \epsilon_{ij} \sim N_{n_i}(0, \Psi) \quad \text{and} \\ C : b_i \sim SN_q\left(-\sqrt{\frac{2}{\pi}}\delta_b, \sigma_b, \delta_b\right) \text{and} \quad \epsilon_{ij} \sim SN_{n_i}\left(-\sqrt{\frac{2}{\pi}}\delta_\epsilon, \sigma_\epsilon^2, \delta_\epsilon\right) \end{cases} . \quad (10)$$

We compare the performance of the proposed model with other models using results of relative bias (Rel.Bias) and mean square error (MSE). The Rel.Bias and MSE of the parameter $\boldsymbol{\phi}$ for the considered models were calculated as defined as

$$\text{Rel.Bias}(\boldsymbol{\phi}) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{\boldsymbol{\phi}}_i}{\boldsymbol{\phi}} - 1\right), \quad \text{MSE}(\boldsymbol{\phi}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{\phi}}_i - \boldsymbol{\phi})^2,$$

where $\hat{\boldsymbol{\phi}}_i$ is the estimate of $\boldsymbol{\phi}$ for the i^{th} samples and $N = 1000$. To select the appropriate model among the competing models, we selected the model with the lowest value of marginal DIC (mDIC) and the conditional DIC (cDIC). We aimed to see the performance of these two versions of DIC in joint modelling of longitudinal and survival model.

4.1 Simulation results

Table 2 shows the results of this simulation study when the measurement and dropout data are generated from a shared normal random-effects model and estimated using three different model assumptions, "A", "B", and "C" described in Section 4. The higher dropout rates increase the bias and the MSEs for all the estimated models. However, we noticed that the differences fluctuate when the dropout parameter is closed to the true value. Besides, for a moderately large value of the dropout parameter compared with the true value, the normal effects give a larger percentage of bias and MSEs when the dropout rate is 50%.

Also, Table 3 shows the simulation results when the data is generated from a shared skew-normal random-effects model (Model "B") and estimated using the different model assumptions. As expected, higher dropout rates increase the bias as well as the MSEs for the considered models. The results also show

that some joint modelling parameters under normal assumption are estimated with bias. These parameters include the variance of the longitudinal model's error term, the variance of random effects, and the coefficients of the survival model's random effects. When the normality assumption of the random effects is violated, the higher percentage bias is observed regardless of the dropout parameter's true value. These results suggest that violation of the random effects distribution assumption yields biased estimates even at the true value of dropout parameter.

Likewise, Table 4 shows the simulation results when data is generated from shared skew-normal for random effects and measurement errors (Model C). As expected, the increase in dropout rates increases bias and MESs as well. Overall, based on the results in Tables 1-3, we can conclude that skew-normal distributions for the random effects lead to better inference and failure to choose the appropriate model introduces bias. Furthermore, the results (not shown here) showed that increasing the skew-normal scenario's sample size is an effective measure in decreasing standard errors, relative bias, and MSE of the parameters. As the sample size increases, the relative bias and MSE decrease when the data is estimated with Model "B" regardless of the measurement and dropout data are generated from a shared normal random-effects model.

In joint modelling of longitudinal and survival data, attention should be paid to the response's distribution. As such, this approach is useful when longitudinal data includes outliers with the assumption of skew-normal/independent distribution for responses. Therefore, the skew-normal assumption for random effects is essential, while such an assumption for random errors may be relaxed. These simulation studies strongly advocate for the sensitivity analysis in joint modelling as routine modelling is not appropriate in the dropout context.
Another aim of this paper is to evaluate the performance of the marginal and conditional DIC to select the "true" model, the simulation studies were replicated for the three distribution scenarios ("A", "B" and "C"), and the results are presented in Table 1. The discrepancy was observed among the models selected by cDIC, as it selected the most complicated model due to the influence of the latent variables. This agrees with previous studies that cDIC tends to select complex model (Chan and Grant; 2016; Ariyo, Quintero, Muñoz, Verbeke and Lesaffre; 2019; Ariyo, Lesaffre, Verbeke and Quintero; 2019). On the other hand, mDIC seems to perform better, based on the results obtained above. As such, the blatant use of DIC for a model with latent variables should be discouraged.

Table 1: Model ranking using estimated DICs for different model varying assumptions

Model	Sample	mDIC	Rank	cDIC	Rank
A	200	932.4	6	-164.1	6
	1000	947.7	5	-162.0	5
B	200	907.1	2	-170.1	4
	1000	902.4	1	-172.4	3
C	200	912.1	4	-194.2	1
	1000	909.1	3	-192.1	2

5 Application to Nigerian indigenous chickens(NIC) data set

The proposed Bayesian joint model was illustrated using a longitudinal data study of 313 chickens progenies produced from mating involving Nigerian indigenous chickens (normal-feathered, frizzle-feathered and naked neck) and an exotic broiler parent stock (Anak Titan). The details for the rationale and experimental design were reported by Adeleke et al. (2011). Genotypes were generated in a straight and reciprocal cross to evaluate the growth and survival performances of progenies produced from the same parent stock (pure breed) and their counterparts produced from different parents (cross breed). The chicken's body weight for each genotype group was measured every week from hatching up to twenty weeks. As the experiment progressed, some chicken from each group died and the mortality rate increased with time that is, there were a substantive number of dropouts in the data set after week 13. Figure 1 presents survival time comparing between the two groups and this figure suggested that pure breed had higher survival probability that the cross breed. The subject-specific profile of randomly selected individual chickens presented in Figure 2. This figure shows a sharply increasing degree of missing data due to mortality.

Non-normal characteristics such as skewness often appear in growth data and this gives bias estimates if neglected. Such characteristics are more problematic for mixed-effects models than for fixed-effects models, because both the within-subjects random effects and random error may jointly contribute to the skewness of the response in a longitudinal study (see Huang and Dagne; 2011). In the alternative, variables transformations are often suggested in practice but they often come with problems as pointed out by Azzalini and Capitanio (1999); the transformed variables are more difficult to interpret, especially when each variable is transformed by using a different function; the transformations are usually on each component separately, and achievement of joint normality is unrealistic. For this purpose, we used an alternative parametric class of multivariate distribution.

Figure 3 shows histogram and Q-Q plot of the response variable at randomly selected time points which shows that there is right skewness. In this model, y_{ij} is the j^{th} body weight measured on i^{th} chicken, $i = 1, 2, \ldots, 313$ and $j = 0, 2, \ldots, 20$. The linear mixed effect model with random intercept is

given as

$$y_{ij} = \beta_0 + \beta_1 age_{ij} + \beta_2 breed_i + b_{0i} + b_{1i} age_{ij} + \varepsilon_{ij}, \tag{11}$$

where breed_i is a chickens' breed indicator (0= pure, 1= cross). The Cox proportional hazard model, was used for the time to events process and this given as

$$h(\text{age}_i) = h_0(\text{age}_i) \exp\left\{\beta_{11} + \beta_{12}\text{breed}_i + \rho_1 b_{1i} + \rho_2 b_{2i}\right\}. \tag{12}$$

The vector $\boldsymbol{b_i} = (b_{1i}, b_{2i})^T$ is shared between model (11) and (12). We considered joint models (11)- (12) under three models assumptions, described in (10), differing in the error and random effects distribution. Prior distributions for the parameters involved were the same for the three models for comparison purposes. Fixed effects were given $N(0, 1000)$ independent prior distributions. For the scale parameter of the error distribution, $\sigma_\epsilon^2 \sim IG(0.01, 0.01)$, so that distribution mean is equal to 1. The hyperparameters of the prior distribution of the scale matrix of the random effect distribution are taken to be $\boldsymbol{\tau}_b$ and $\boldsymbol{T}_b = \boldsymbol{I}_b$, for the skewness parameter, we assigned independently truncated normal distributions with mean 0 and variance 100. For the piecewise baseline hazard function, we considered $\Gamma(1, 1)$ for each of the piecewise baseline function $(h_i, i = 1, 2)$. Hyperparameters are chosen such that the priors of the parameters tend to be weakly informative (Lesaffre and Lawson; 2012).

In Bayesian MCMC implementation, we use 70,000 iterations after discarding the first 30,000 iterations to make an inference. To avoid correlation problems in the generated chains the lag was set to 5. After checking Gelman-Rubin diagnosis test for convergence, the resulting parameter estimates are given in Table 5.

Estimates of β_s across each model with a different assumption for random effects and random error are significant. However, the estimates are larger when the skew-normal distribution is assumed, which means the wrong assumption leads to bias. In the survival model, ρ_1 and ρ_2 are significant, which shows that the two models are dependent, also the skewness parameters are significant. This implies that ignoring this parameter for the normal model leads to overestimating the variance of the error in longitudinal data (see Baghfalaki and Ganjali; 2015).

Further observed in the simulation studies, there are discrepancies in the different types of DIC in selection criteria. While mDIC favours model 'B', its conditional counterpart chose 'C' as the appropriate model. Model 'A' such that purely Gaussian was not selected by any of the criteria. This result is slightly different from the one obtained by Baghfalaki and Ganjali (2015) which using cDIC with other criteria selected the model that is purely skewed. This difference may be a consequence of over-parametrization of the model 'C' as DIC favours over-fitted models (see also Meng; 2009; Chan; 2016; Quintero and Lesaffre; 2017)

Table 2: Simulation results (Est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and MSE: mean square error) for 1000 samples when the measurement and dropout data were generated from a shared normal random-effects model and estimated using three different models assumptions "A", "B", and "C" described in section 4) with different dropout rates.

	Model		A			B			C		
Dropout rate	Parameters	Real	Est(S.E)	Rel.Bias	MSE	Est(S.E)	Rel.Bias	MSE	Est(S.E)	Rel.Bias	MSE
10%	β_0	5.00	4.966(0.31)	-0.002	0.074	5.015(0.29)	0.002	0.071	4.987(0.15)	0.002	0.066
	β_1	-2.00	-2.014 (0.20)	0.005	0.042	-2.000(0.20)	0.002	0.037	-2.321(0.06)	0.002	0.032
	β_2	-1.00	-0.992(0.54)	0.040	0.031	-0.998(0.51)	0.003	0.047	-1.261(0.21)	0.003	0.042
	β_{01}	3.00	2.999(0.54)	-0.031	0.241	3.001(0.51)	0.001	0.002	3.021(0.21)	0.001	0.001
	β_{11}	-3.00	-2.969(0.50)	-0.052	0.272	-3.006(0.49)	0.012	0.302	-3.000(0.19)	0.013	0.291
	$\sigma^2_{b_{11}}$	1.00	4.027(0.46)	4.132	15.231	1.010(0.42)	4.003	15.232	0.964(0.11)	4.001	15.200
	$\sigma^2_{b_{12}}$	0.50	0.241(0.23)	0.412	0.124	0.499(0.21)	0.342	0.201	0.492(0.95)	0.338	0.201
	$\sigma^2_{b_{22}}$	1.00	2.352(0.24)	2.314	4.521	1.002(0.21)	2.103	4.527	1.002(0.29)	2.100	4.341
	σ^2_ϵ	0.25	0.247(0.52)	3.240	10.213	0.246(0.48)	3.121	10.238	0.247(0.15)	2.998	10.210
	δ_ϵ	1.00	-	-	-	-		-	1.001(0.79)	0.005	0.023
	δ_b	1.00	-	-	-	1.002(0.27)	0.007	0.096	0.961(0.09)	0.005	0.092
	ρ_1	1.00	0.647(0.19)	0.352	-0.373	0.983(0.15)	0.006	0.124	1.010(0.00)	0.004	0.120
	ρ_2	2.00	2.214(0.20)	0.113	0.113	2.000(0.20)	0.206	0.212	1.075(0.00)	0.202	0.201
50%	β_0	5.00	5.034(0.23)	0.008	0.112	5.212(0.21)	0.100	0.107	5.021(0.21)	0.006	0.072
	β_1	-2.00	-2.231(0.39)	0.032	0.212	2.004(0.09)	0.092	0.041	2.012(0.18)	0.005	0.034
	β_2	-1.00	-0.996(0.78)	0.063	0.251	1.007(0.58)	0.017	0.501	1.041(0.34)	0.003	0.048
	β_{01}	-2.00	-2.001(0.69)	0.041	0.321	1.996(0.53)	0.131	0.002	1.998(0.29)	0.001	0.001
	β_{11}	-3.00	-2.879(0.55)	0.008	0.521	3.975(0.43)	0.019	0.304	2.997(0.37)	0.015	0.301
	$\sigma^2_{b_{11}}$	1.00	2.309(0.36)	4.231	16.021	0.893(0.69)	4.007	15.248	1.017(0.38)	4.021	15.246
	$\sigma^2_{b_{12}}$	0.50	0.421(0.32)	0.442	0.142	0.498(0.44)	0.401	0.224	0.502(0.62)	0.340	0.214
	$\sigma^2_{b_{22}}$	1.00	3.142(0.42)	2.601	4.732	1.621(0.57)	2.349	4.529	1.009(0.54)	2.110	4.443
	σ^2_ϵ	0.25	4.011(0.75)	3.252	10.302	0.679(0.67)	3.150	10.240	0.249(0.69)	3.000	10.301
	δ_ϵ	1.00	-	-	-	-	-	-	1.008(0.44)	0.005	0.023
	δ_b	1.00	-	-	-	0.938(0.64)	0.009	0.100	1.001(0.75)	0.006	0.094
	ρ_1	1.00	0.998(0.74)	0.443	0.123	1.007(0.27)	0.010	0.126	0.982(0.50)	0.005	0.122
	ρ_2	2.00	2.007(0.34)	0.116	0.172	1.897(0.25)	0.225	0.302	2.009(0.27)	0.204	0.212

Table 3: Simulation results (Est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and MSE: mean square error) for 1000 samples when the measurement and dropout data were generated from a shared skew-normal random-effects model and estimated using three different models assumptions "A", "B", and "C" described in section 4) with different dropout rates.

	Model		A			B			C		
Dropout rate	Parameters	Real	Est(S.E)	Rel.Bias	MSE	Est(S.E)	Rel.Bias	MSE	Est(S.E)	Rel.Bias	MSE
10%	β_0	5.00	5.427(0.52)	0.003	0.193	5.018(0.30)	0.001	0.070	4.972(0.16)	0.002	0.071
	β_1	-2.00	-2.242 (0.47)	0.021	0.083	-2.001(0.12)	0.001	0.024	-2.121(0.05)	0.002	0.041
	β_2	-1.00	-1.200(0.64)	0.071	0.049	-0.999(0.42)	0.001	0.001	-1.121(0.05)	0.002	0.045
	β_{01}	3.00	2.792(0.75)	0.092	0.349	3.000(0.44)	0.001	0.001	3.001(0.20)	0.001	0.001
	β_{11}	-3.00	-2.601(0.60)	-0.001	0.379	-3.001(0.50)	0.011	0.300	-3.000(0.17)	0.012	0.201
	$\sigma^2_{b_{11}}$	1.00	4.245(0.58)	5.071	16.000	1.010(0.31)	1.003	2.232	1.000(0.03)	1.001	2.420
	$\sigma^2_{b_{12}}$	0.500	0.363(0.69)	0.691	0.341	0.501(0.12)	0.211	0.201	0.421(0.85)	0.401	0.214
	$\sigma^2_{b_{22}}$	1.00	3.00(0.47)	2.791	4.822	1.002(0.20)	0.101	0.321	1.005(0.27)	0.104	0.201
	σ^2_{ϵ}	0.250	0.274(0.55)	3.946	2.413	0.256(0.32)	0.212	2.512	0.249(0.11)	0.001	0.002
	δ_{ϵ}	1.000	-	-	-	-		-	1.001(0.51)	0.001	0.021
	δ_b	1.000	-	-	-	1.002(0.24)	0.024	0.190	0.999(0.03)	0.005	0.092
	ρ_1	1.00	1.22(0.64)	0.872	-0.001	0.999(0.11)	0.001	0.124	1.000(0.00)	0.001	0.100
	ρ_2	2.00	2.714(0.68)	0.493	0.688	2.000(0.10)	0.178	0.199	1.021(0.001)	0.200	0.921
50%	β_0	5.00	5.824(0.76)	0.025	0.361	5.228(0.43)	0.100	0.073	5.001(0.16)	0.003	0.092
	β_1	-2.00	-2.04(0.58)	0.041	0.074	-2.001(0.13)	0.002	0.031	-2.124(0.05)	0.002	0.041
	β_2	-1.00	-1.230(0.75)	0.191	0.271	-1.000(0.52)	0.004	0.004	-1.131(0.13)	0.002	0.041
	β_{01}	-2.00	-2.501(0.74)	0.693	0.361	1.996(0.55)	0.131	0.002	1.998(0.29)	0.001	0.001
	β_{11}	-3.00	-2.345(0.83)	0.412	0.621	-3.00(0.42)	0.013	0.034	-3.00(0.22)	0.142	0.253
	$\sigma^2_{b_{11}}$	1.00	4.401(0.67)	0.041	16.01	1.009(0.32)	1.012	2.433	1.00(0.02)	1.062	2.752
	$\sigma^2_{b_{12}}$	0.500	0.474(0.74)	0.841	0.521	0.482(0.21)	0.511	0.400	0.392(1.00)	0.631	0.432
	$\sigma^2_{b_{22}}$	1.00	3.324(0.50)	2.943	4.721	1.004(0.42)	0.321	2.734	1.09(0.56)	0.321	0.534
	σ^2_{ϵ}	0.250	0.301(0.50)	2.79	4.841	1.004(0.32)	0.231	0.431	1.34(0.43)	0.231	0.431
	δ_{ϵ}	1.00	-	-	-	-	-	-	1.004(0.65)	0.06	0.060
	δ_b	1.00	-	-	-	1.043(0.23)	0.235	0.24	1.00(0.50)	0.23	0.120
	ρ_1	1.00	1.54(0.83)	0.732	0.201	1.00(0.32)	0.004	0.231	1.00(0.00)	0.030	0.004
	ρ_2	2.0	2.80(0.70)	0.505	0.62	2.00(0.11)	0.190	2.00	1.021(0.02)	0.321	0.942

Table 4: Simulation results (Est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and MSE: mean square error) for 1000 samples when the measurement and dropout data were generated from a shared skew-normal random-effects and measurement error model and estimated using three different models assumptions "A", "B", and "C" described in section 4) with different dropout rates.

	Model			A			B			C	
Dropout rate	Parameters	Real	Est(S.E)	Rel.Bias	MSE	Est(S.E)	Rel.Bias	MSE	Est(S.E)	Rel.Bias	MSE
10%	β_0	5.00	5.436(0.41)	0.003	0.193	5.072(0.32)	0.003	0.069	4.963(0.157)	0.002	0.062
	β_1	-2.00	-2.253 (0.47)	0.042	0.094	-2.001(0.22)	0.001	0.024	-2.131(0.05)	0.002	0.041
	β_2	-1.00	-1.201(0.75)	0.069	0.054	-0.999(0.54)	0.001	0.001	-1.131(0.07)	0.003	0.062
	β_{01}	3.00	2.789(0.83)	0.097	0.367	3.000(0.49)	0.001	0.001	3.001(0.20)	0.001	0.001
	β_{11}	-3.00	-2.643(0.63)	-0.001	0.384	-3.001(0.63)	0.016	0.300	-3.000(0.18)	0.012	0.212
	$\sigma^2_{b_{11}}$	1.00	4.545(0.59)	5.081	16.040	1.010(0.35)	1.003	2.432	1.000(0.03)	1.001	2.390
	$\sigma^2_{b_{12}}$	0.500	0.543(0.79)	0.673	0.352	0.521(0.12)	0.241	0.221	0.421(0.85)	0.401	0.214
	$\sigma^2_{b_{22}}$	1.00	3.00(0.65)	2.892	4.842	1.002(0.20)	0.151	0.331	1.005(0.37)	0.114	0.211
	σ^2_{ϵ}	0.250	0.286(0.65)	3.953	2.453	0.273(0.32)	0.212	2.542	0.252(0.11)	0.001	0.002
	δ_{ϵ}	1.000	-	-	-	-		-	1.021(0.56)	0.001	0.021
	δ_b	1.000	-	-	-	1.002(0.24)	0.042	0.190	0.999(0.03)	0.005	0.092
	ρ_1	1.00	1.32(0.64)	0.894	-0.001	0.999(0.11)	0.001	0.124	1.000(0.00)	0.001	0.100
	ρ_2	2.00	2.954(0.68)	0.510	0.699	2.000(0.10)	0.185	0.199	1.021(0.00)	0.210	0.951
50%	β_0	5.00	5.943(0.76)	0.025	0.1.384	5.678(0.46)	0.100	0.073	5.101(0.19)	0.009	0.092
	β_1	-2.00	-2.16(0.67)	0.041	0.089	-2.001(0.13)	0.002	0.031	-2.431(0.0.05)	0.002	0.041
	β_2	-1.00	-1.230(0.86)	0.191	0.361	-1.000(0.52)	0.008	0.009	-1.131(0.23)	0.002	0.084
	β_{01}	-2.00	-2.531(0.89)	0.745	0.361	1.986(0.55)	0.131	0.002	1.999(0.32)	0.001	0.001
	β_{11}	-3.00	-2.365(0.83)	0.452	0.661	-3.00(0.67)	0.016	0.034	-3.00(0.22)	0.142	0.253
	$\sigma^2_{b_{11}}$	1.00	4.401(0.69)	0.041	16.31	1.009(0.32)	1.012	2.633	1.00(0.02)	1.062	2.852
	$\sigma^2_{b_{12}}$	0.500	0.574(0.74)	0.891	0.521	0.492(0.21)	0.531	0.400	0.402(1.00)	0.631	0.492
	$\sigma^2_{b_{22}}$	1.00	3.544(0.50)	2.94	4.821	1.004(0.42)	0.421	2.834	1.09(0.58)	0.321	0.534
	σ^2_{ϵ}	0.250	0.341(0.50)	2.94	4.941	1.034(0.32)	0.231	0.451	1.34(0.43)	0.251	0.441
	δ_{ϵ}	1.00	-	-	-	-	-	-	1.004(0.67)	0.069	0.068
	δ_b	1.00	-	-	-	1.043(0.23)	0.235	0.24	1.00(0.50)	0.238	0.126
	ρ_1	1.00	1.63(0.94)	0.742	0.221	1.00(0.33)	0.004	0.251	1.00(0.00)	0.030	0.004
	ρ_2	2.0	2.90(0.73)	0.50	0.62	2.00(0.21)	0.190	2.00	1.021(0.021)	0.361	0.965

Table 5: Posterior means, standard deviation, 95 % credible intervals for the parameters under the three fitted models for analysing NLC data sets

	Model A				**Model B**				**Model C**			
Parameter	Mean	SD	2.5%	97.5 %	Mean	SD	2.5%	97.5 %	Mean	SD	2.5%	97.5 %
β_0	-4.4201	0.0515	-4.5208	-3.9112	-4.8210	0.0105	-4.9312	-4.4301	-5.8061	0.0107	-10.5154	-4.0187
β_1	1.2076	0.0091	1.2116	1.2434	1.2319	0.0195	1.2124	1.2019	1.9508	0.0163	1.9490	3.0282
β_2	1.0919	0.0438	1.1563	1.0263	1.1335	0.0335	1.0795	1.1933	1.1194	0.0342	1.1174	1.2539
β_{11}	-1.679	0.574	-3.315	-0.404	-1.639	0.702	-3.318	-0.931	-1.773	0.621	-2.934	-1.071
β_{12}	0.542	0.304	-0.098	1.340	0.480	0.292	0.501	0.932	0.393	0.294	-0.314	0.842
δ_{b_1}	-	-	-	-	0.0207	0.0011	0.0203	0.0759	0.0315	0.0012	0.0238	0.0887
δ_{b_2}	-	-	-	-	0.0101	0.0031	0.0104	0.1074	0.0089	0.0003	0.0083	0.0267
δ_ϵ	-	-	-	-	-	-	-	-	0.3075	0.0123	0.2830	0.3312
σ^2_ϵ	0.0476	0.0014	0.0468	0.0495	0.0413	0.0014	0.0413	0.0502	0.0128	0.0019	0.0092	0.0168
$\sigma^2_{b_1}$	0.0556	0.0084	0.0411	0.0741	0.0412	0.0056	0.0312	0.0532	1.5462	0.0440	2.2072	5.2060
$\sigma_{b_1}\sigma_{b_2}$	-0.0128	0.0023	-0.0179	-0.0086	-0.0114	0.0020	-0.0157	-0.0078	-1.1340	1.6372	-4.3519	-0.0101
$\sigma^2_{b_2}$	0.0112	0.0012	0.0091	0.0138	0.0109	0.0012	0.0089	0.0135	0.8754	1.2911	0.0090	3.6734
h_1	0.198	0.094	0.048	0.398	0.1812	0.082	0.034	0.0521	0.118	0.089	0.028	2.993
h_2	1.828	0.956	0.567	4.985	1.713	0.876	0.452	3.432	1.698	0.783	0.477	3.976
cDIC			-159.7			-170.32			-169.5			
mDIC			423.44			413.22			412.147			

6 Discussion

There are several research questions in animal improvement programme that are best addressed within the framework of a joint model. As such, this paper discusses Bayesian implementation of a robust alternative to joint modelling of longitudinal and survival data with skewness. Our focus on the scenario where the dropout (i.e. death of the animal) is related to the experiment. The main focus of the paper is on the right censoring. However, the extension of the method to other forms of complex longitudinal data involving double-, or interval-censoring following Sinha et al. (1999) and Yu (2010) is straightforward. We apply our methodology to a longitudinal data of Nigeria indigenous chicken data to illustrate how the procedure can be used to evaluate model assumptions, dropout in response and obtain robust parameter estimates. Besides, we illustrate the proposed method using a set simulation studies, and the results show a gain in efficiency and accuracy for parameter estimates as well as the superior performance of the marginal DIC to pick the best-fitting model, where the typical assumption of normality is unrealistic.

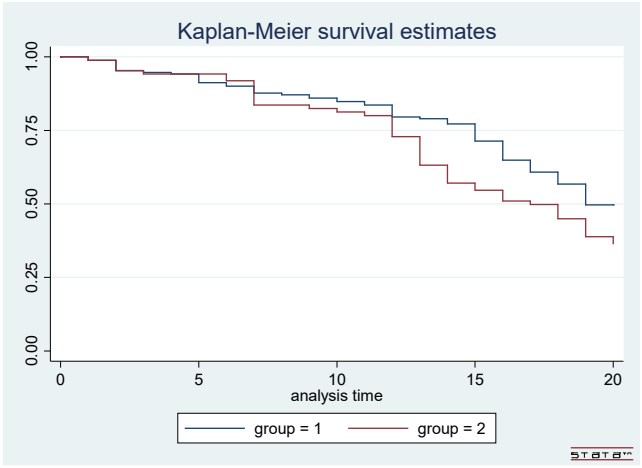


Fig. 1: Kaplan Meier estimates of the probability of survival for pure and cross breeds chickens

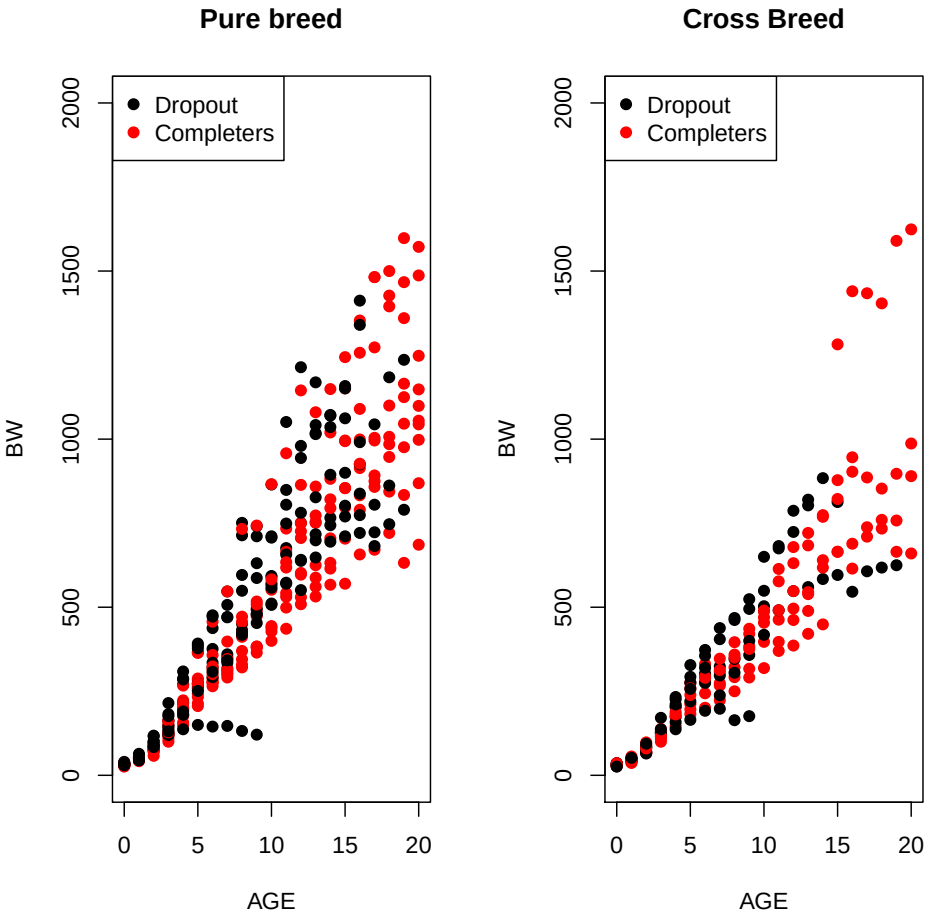
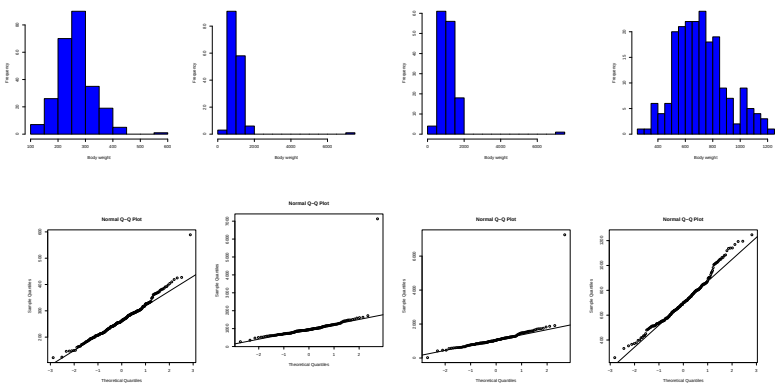


Fig. 2: The body weight for randomly selected chicken for both pure and cross breeds colour red indicate those who remain till the end of the study and colour black for dropout.



(a)

Fig. 3: Histogram and q-q plots of chickens' body weight at randomly selected time points.

References

Adeleke, M. A., Peters, S. O., Ozoje, M. O., Ikeobi, C. O. N., Bamgbose, A. M. and Adebambo, O. A. (2011). Growth performance of Nigerian local chickens in crosses involving an exotic broiler breeder, *Tropical Animal Health and Production* **43**(3): 643–650.
URL: *http://dx.doi.org/10.1007/s11250-010-9747-3*

Alsefri, M., Sudell, M., García-Fiñana, M. and Kolamunnage-Dona, R. (2020). Bayesian joint modelling of longitudinal and time to event data: a methodological review, *BMC Medical Research Methodology* **20**: 1–17.

Arellano-Valle, R., Bolfarine, H. and Lacho, V. (2007). Bayesian inference for skew-normal linear mixed models, *Journal of Applied Statistics* **34**(4): 663–682.

Ariyo, O., Lesaffre, E., Verbeke, G. and Quintero, A. (2019). Model selection for Bayesian linear mixed models with longitudinal data: Sensitivity to the choice of priors, *Communications in Statistics - Simulation and Computation* **0**(0): 1–25.
URL: *https://doi.org/10.1080/03610918.2019.1676439*

Ariyo, O., Quintero, A., Muñoz, J., Verbeke, G. and Lesaffre, E. (2019). Bayesian model selection in linear mixed models for longitudinal data, *Journal of Applied Statistics* **47**(5): 890–913.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**(3): 579–602.

Baghfalaki, T. and Ganjali, M. (2015). A Bayesian approach for joint modeling of skew-normal longitudinal measurement and time to event data, *REVSTAT-Statistical Journal* **13**(2): 169–191.

Bogaerts, K., Komarek, A. and Lesaffre, E. (2017). *Survival analysis with interval-censored data: A practical approach with examples in R, SAS, and BUGS*, CRC Press.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations, *Journal of Computational and Graphical Statistics* **7**(4): 434–455.

Celeux, G., Forbes, F., Robert, C. and Titterington, D. (2006). Deviance information criteria for missing data models, *Bayesian Analysis* **1**: 651–706.

Chan, J. (2016). Bayesian informative dropout for longitudinal binary data with random effects using conditional and joint modeling approaches, *Biometrical Journal* **58**: 549–569.

Chan, J. and Grant, A. (2016). On the observed-data deviance information criterion for volatility modeling, *J Financ Economet* **14**(4): 772–802.

Cox, D. R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society: Series B (Methodological)* **34**(2): 187–202.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **43**(1): 49–73.

Fitmaurice, G., Molenberghs, G. and Lipsitz, S. (1995). Regression models for longitudinal binary responses with informative drop-outs, *Journal of Royal Statistical Society* **57**(B): 691–704.

Gelman, A., Rubin, D. B. et al. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science* **7**(4): 457–472.

Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data, *Biostatistics* **1**(4): 465–480.
URL: *+ http://dx.doi.org/10.1093/biostatistics/1.4.465*

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**(1): 117.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models, *Journal of the American Statistical Association* **91**(436): 1461–1473.
URL: *http://www.jstor.org/stable/2291572*

Huang, X., Li, G. and Elashoff, R. M. (2010). A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements, *Statistics and its Interface* **3**(2): 185.

Huang, Y. and Dagne, G. (2011). A Bayesian approach to joint mixed-effects models with a skwe-normal distribution and measurement errors in covariates, *Biometrics* **67**: 260–269.

Ibrahim, J. G., Chu, H. and Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data, *Journal of Clinical Oncology* **28**(16): 2796.

Komarek, A. and Lesaffre, E. (2008). Generalized linear mixed model with a penalized gaussian mixture as a random effects distribution, *Computational Statistics Data Analysis* **52**(7): 3441 – 3458.
URL: *http://www.sciencedirect.com/science/article/pii/S016794730700429X*

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics* pp. 963–974.

Lesaffre, E. and Lawson, A. (2012). *Bayesian Biostatistics*, A John Wiley and Sons, Ltd Publication.

Li, N., Elashoff, R. M. and Li, G. (2009). Robust joint modeling of longitudinal measurements and competing risks failure time data, *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **51**(1): 19–30.

Li, Q. and Su, L. (2017). Accommodating informative dropout and death: a joint modelling approach for longitudinal and semi competing risks data, *Journal of Royal Statistical Society* **66**(Series C): 1–18.

Ma, J., Plesken, H., Treisman, J. E., Edelman-Novemsky, I. and Ren, M. (2004). Lightoid and Claret: a rab GTPase and its putative guanine nucleotide exchange factor in biogenesis of Drosophila eye pigment granules, *Proceedings of the National Academy of Sciences* **101**(32): 11652–11657.

Mchunu, N. N., Mwambi, H. G., Reddy, T., Yende-Zuma, N. and Naidoo, K. (2020). Joint modelling of longitudinal and time-to-event data: an illustration using CD4 count and mortality in a cohort of patients initiated on

antiretroviral therapy, *BMC Infectious Diseases* **20**: 1–9.

Meng, X. (2009). Discussion of Spiegelhalter et al., *J Roy Stat Soc B* **64**(4): 633–637.

Molenberghs, G., Kenward, M. G. and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop-out, *Biometrika* **84**(1): 33–44.

Plummer, M. et al. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, *Proceedings of the 3rd international workshop on distributed statistical computing*, Vol. 124, Vienna, Austria, pp. 1–10.

Quintero, A. and Lesaffre, E. (2017). Comparing latent variable models via the observed deviance information criterion, *Submitted* .

Rizopoulos, D. (2012). Joint models for longitudinal and time-to-event data: With applications in R.

Rubin, D. (1976). Inference and missing data, *Biometrics* **63**: 581–592.

Sahu, S., Dey, D. and Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models, *Canadian Journal of Statistics* **31**(2): 129–150.

Sinha, D., Chen, M.-H. and Ghosh, S. K. (1999). Bayesian analysis and model selection for interval-censored survival data, *Biometrics* **55**(2): 585–590.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society: Series B* **64**(4): 583–616.

Todem, D., Kim, K., Fine, J. and Peng, L. (2010). Semiparametric regression models and sensitivity analysis of longitudinal data with non-random dropouts, *Statistica Neerlandica* **64**(2): 133–156.

Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects, *Statistics In Medicine* **19**: 3309–3324.

Verbeke, G. and Lesaffre, E. (1997). A linear mixed-effect model with heterogeneity in the random-effects population, *Journal of the American Statistical Association* **91**(3): 217–221.

Wang, C., Douglas, J. and Anderson, S. (2002). Item response models for joint analysis of quality of life and survival, *Statistics in medicine* **21**(1): 129–142.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error, *Biometrics* pp. 330–339.

Yu, B. (2010). A Bayesian MCMC approach to survival analysis with doubly-censored data, *Computational Statistics & Data Analysis* **54**(8): 1921–1929.

Zhang, G. and Yuan, Y. (2012). Bayesian modeling longitudinal dydic data with nonignorable dropout,with application to a breast cancer study, *The Annals of Applied Statistics* **6**(2): 753–771.