

Bayesian Joint Modelling of Survival Time and Longitudinal CD4 Cell Counts Using Accelerated Failure Time and Generalized Error Distributions

Markos Abiso Erango^{1*}, Ayele Taye Goshu²

¹Department of Statistics, Arba Minch University, Arba Minch, Ethiopia

²Department of Mathematics, Kotebe Metropolitan University, Addis Ababa, Ethiopia

Email: *markos.erango73@gmail.com, ayele_taye@yahoo.com

How to cite this paper: Erango, M.A. and Goshu, A.T. (2019) Bayesian Joint Modelling of Survival Time and Longitudinal CD4 Cell Counts Using Accelerated Failure Time and Generalized Error Distributions. *Open Journal of Modelling and Simulation*, 7, 79-95.

<https://doi.org/10.4236/ojmsi.2019.71004>

Received: March 20, 2018

Accepted: January 18, 2019

Published: January 21, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Survival of HIV/AIDS patients is crucially dependent on comprehensive and targeted medical interventions such as supply of antiretroviral therapy and monitoring disease progression with CD4 T-cell counts. Statistical modelling approaches are helpful towards this goal. This study aims at developing Bayesian joint models with assumed generalized error distribution (GED) for the longitudinal CD4 data and two accelerated failure time distributions, Log-normal and loglogistic, for the survival time of HIV/AIDS patients. Data are obtained from patients under antiretroviral therapy follow-up at Shashemene referral hospital during January 2006-January 2012 and at Bale Robe general hospital during January 2008-March 2015. The Bayesian joint models are defined through latent variables and association parameters and with specified non-informative prior distributions for the model parameters. Simulations are conducted using Gibbs sampler algorithm implemented in the WinBUGS software. The results of the analyses of the two different data sets show that distributions of measurement errors of the longitudinal CD4 variable follow the generalized error distribution with fatter tails than the normal distribution. The Bayesian joint GED loglogistic models fit better to the data sets compared to the lognormal cases. Findings reveal that patients' health can be improved over time. Compared to the males, female patients gain more CD4 counts. Survival time of a patient is negatively affected by TB infection. Moreover, increase in number of opportunistic infection implies decline of CD4 counts. Patients' age negatively affects the disease marker with no effects on survival time. Improving weight may improve survival time of patients. Bayesian joint models with GED and AFT distributions are found to be useful

in modelling the longitudinal and survival processes. Thus we recommend the generalized error distributions for measurement errors of the longitudinal data under the Bayesian joint modelling. Further studies may investigate the models with various types of shared random effects and more covariates with predictions.

Keywords

Accelerated Failure Time, Bayesian Joint Model, CD4 Cell Count, Generalized Error Distribution, HIV/AIDS, Longitudinal, Survival Analysis

1. Introduction

Survival of HIV/AIDS patients is crucially dependent on comprehensive and targeted medical interventions. Health professionals monitor patients' health status using such disease markers as CD4 T-cells counts. The disease progression as indicated by the longitudinal CD4 measures may affect the time of an event of interest—death of a patient in this case. The main interest of inference is on the association between the longitudinal and survival processes. Joint models for longitudinal and time-to-event are based on the joint distribution of the two processes [1]. The joint analysis may be appropriate when the longitudinal variable is correlated with patient's health status and incorporate all information simultaneously so as to provide valid and efficient inferences [2].

The traditional approach in the analysis of survival data assumes a homogeneous population, where all individuals have same health risks. In practice, individual patients possibly differ in health risks such as their vulnerability to causes of death, responses to treatments, and influences of various risk factors. Joint modelling of the two data often assumes normal distributions in the linear mixed models [2] [3] [4]. It is interesting to look for alternative distributions that can accommodate data that may not be normally distributed.

The current study considers the longitudinal measure in terms of its rate of growth. The rate of growth is an important concept in studying changes. If the level of growth is viewed as the current status of a process at a specific time, the rate of growth measures how fast the process is changing at that time [4]. Studies by [4] extend the usual growth models using the generalized error distribution (GED) and estimate its parameters under Bayesian framework. The author studied such a general form of linear growth model $Y_{it} = X_t \beta_i + \epsilon_{it}$, where Y_{it} is growth observation for individual i at time t assuming the error term to have the generalized error distribution $\epsilon_{it} \sim \text{GED}(0, \sigma_\epsilon^2, \gamma_t)$ and independent of the random effects β_i . The GED has a shape parameter γ_t that may possibly vary across time, but here assumed constant. The model can handle data with both leptokurtic and platykurtic errors.

Markos *et al.* [5] studied joint modelling of survival time and longitudinal CD4 cell counts of HIV/AIDS patients using Bayesian methods. The authors

compared various Bayesian joint models involving the Weibull, lognormal and loglogistic AFT distributions and normality assumption for the longitudinal CD4 measure using two data sets. They recommended the Bayesian joint loglogistic model for one data set collected from Shashemene referral hospital and the Bayesian joint lognormal model for the second data set collected from Bale Robe general hospital. These models have same hazard rate functions as that of data sets. In the current study, we further analyze the same data sets with newly defined Bayesian joint models considering the generalized error distribution for the longitudinal measure instead of normal distribution.

2. Methodology

2.1. Description of Data

The study considers two data sets that are collected from two hospitals under similar settings and as considered by [5] [6] [7]. The data are extracted from patients' charts which contain epidemiological, laboratory and clinical information of the patients. Patients with ages less than 16 years old are also those who started ART before the defined study period were not included in this study.

Data 1: The first data set is obtained for 354 random samples of HIV/AIDS patients who had been under ART follow-up from January 2006 to December 2012 at the Shashemene referral hospital. There are two outcome variables in each data set. The longitudinal measure is the square root of the number of CD4 cell counts per mm^3 of blood repeatedly measured at approximately every 6 months interval. The survival outcome is the time in months of a patient to the associated death event. Explanatory variables for the longitudinal response are: visit time, square of visit time, sex, functional status, alcohol use, tobacco use, number of opportunistic infections. Predictors for the survival time are: TB infection status at baseline, awareness about ART, condom use, number of opportunistic infections, number of living rooms at home. **Data 2:** The second data set is obtained for 400 random samples of HIV/AIDS patients who had been under ART follow-up from January 2008 to March 2015 at the Bale Robe general hospital. Outcome variables are: longitudinal measure which is the square root of the number of CD4 cell counts per mm^3 of blood that repeatedly measured at approximately every 6 months interval, and survival time in months of a patient to death event. Explanatory variables for the longitudinal measure are: visit time, square of visit time, sex, age, weight, number of opportunistic infections, and those for the survival time are: age, weight, functional status, tobacco, condom. Description and codes of the explanatory variables are described in **Table 1**.

2.2. Linear Mixed Models

The longitudinal data, CD4 T-cell counts, are measurements on the response variable taken from same individuals over several observation times. Thus the set of observations on a subject tends to be inter-correlated [8] [9]. The two sources of variations expected are the within-patient and the between-patients variations.

Table 1. Explanatory variables with codes.

No.	Variable	Description and Codes
1	Observed time	Months (continuous)
2	Square of observed time	Months squared (continuous)
3	Gender	Female (0), Male (1)
4	Age	Years (continuous)
5	Weight	Kilograms (continuous)
6	Functional status	Working (0), Ambulatory (1), Bedridden (2)
7	Alcohol use	No (-1), No opinion (0), Yes (1)
8	Tobacco use	No (-1), No opinion (0), Yes (1)
9	Number of opportunistic infections	Counts
10	TB infection status at baseline	Negative (0), Positive (1)
11	Awareness about ART	No (0), Yes (1)
12	Condom use	Not always use (0), Use always (1)
13	Number of living rooms at home	Counts

Analysis of within-patient variation allows studying of changes of the CD4 counts over time, while analysis of between-patients variation allows understanding differences between patients.

Here we assume that the longitudinal CD4 measure has the generalized error distribution for instead of normal. For any variable Y that follows the generalized error distribution, its density function $GED(y; \gamma, \mu, \sigma^2)$ with three parameters as adapted by [4] from [10] is:

$$f(y) = a(\gamma) \frac{1}{\sigma} \exp \left[-c(\gamma) \left| \frac{y - \mu}{\sigma} \right|^{\frac{2}{1+\gamma}} \right] \quad (1)$$

$$\text{with: } a(\gamma) = \frac{\Gamma[3(1+\gamma)/2]^{\frac{1}{2}}}{(1+\gamma)\Gamma[(1+\gamma)/2]^{\frac{3}{2}}}, \quad c(\gamma) = \left\{ \frac{\Gamma[3(1+\gamma)/2]}{\Gamma[(1+\gamma)/2]} \right\}^{\frac{1}{1+\gamma}}$$

Here μ is location and σ^2 is scale parameters. And γ is the shape parameter of GED that is related to kurtosis of the distribution and characterizes non-normality of Y . The GED can model the error distribution more flexibly than the normal one [4] [10] [11].

The generalized error distribution generalizes the normal distribution. Normal distribution is a special case of GED when $\gamma = 0$ in which case $w(0) = 1/\sqrt{2\pi}$ and $c(0) = 1/2$. Other special cases include a Laplace (double exponential) distribution when $\gamma = 1$ and when becomes a uniform distribution γ approaches -1. It becomes leptokurtic distribution for $\gamma \in (0, 1)$, gets fatter tails when $\gamma > 0$, and gets thinner tails than the normal distribution when $\gamma < 0$. Choy and Smith [12] derived the GED as a scale mixture of normal distributions for $\gamma \in (0.5, 1]$.

The GED in Equation (1) is expressed in a simpler form by [13] as follows:

$$f(y) = \frac{\lambda s}{2\Gamma\left(\frac{1}{s}\right)} \exp\left(-\lambda^s |y - \mu|^s\right) \quad (2)$$

The normal distribution is a special case of this form of GED when $s = 2$ and so $\lambda = 1/\sigma\sqrt{2}$. But in many situations, data are assumed normal though normality may not be an appropriate assumption. The statistical package such as `fitdistrplus` developed by [14] can be used to see whether or not measurement errors of data at hand are normally distributed.

The generalized error distribution was first introduced by Subbotin [15] as class of symmetric distributions with variation in kurtosis. The distributions have many structural properties close to a normal distribution. Many researchers have studied GED including its applications but not in the joint models studied here. Nelson [16] developed linear regression and time series models with heavy tails assuming the underlying distribution to be the GED. It can be used in statistical modelling if the observation errors are not necessarily normally distributed. Zhang [4] proposed and studied linear mixed growth models for longitudinal data with the GED so as to handle leptokurtic and platykurtic errors. The author reported that such models fit better to data than the respective models with normality assumptions.

In our case, we first analyzed the CD4 counts data in `fitdistrplus` package [14] and found that measurement errors of the longitudinal CD4 data seem non-normally distributed. Then we define the linear mixed model using generalized error distribution. Let Y_{it} be the longitudinal CD4 measurement of the i^{th} patient $i = 1, 2, 3, \dots, n$ at times $t = t_1, \dots, t_T$. The linear mixed model for the longitudinal process with assumption of generalized error distribution for the error term is defined as:

$$Y_{it} = \mu_i(t_{it}) + W_{it}(t_{it}) + \epsilon_{it} \quad (3)$$

where $\mu_i(t_{it}) = \mathbf{X}_{it}\boldsymbol{\beta}_i$ is time dependent mean response of Y_{it} as a function of predictors \mathbf{X}_{it} and coefficients $\boldsymbol{\beta}_i$, $W_{it}(t_{it})$ is time dependent subject specific random effects having normal distribution with mean zero and variance σ^2 , and ϵ_{it} is a random error distributed as $\epsilon_{it} \sim \text{GED}(\gamma, 0, \sigma_\epsilon^2)$. The shape parameter γ is an important parameter to be studied here.

2.3. Survival Models

The survival time is random variable defined on non-negative real numbers. The observed time is taken as the minimum $T_i = \min(t_i, c_i)$ of the time to event t_i and time to censoring c_i . The time variable is modeled with two AFT distributions (lognormal and loglogistic) as considered in [5] [17] [18] [19].

We assume that the survival time follows lognormal distribution. Its probability density function $f(t)$, survival function $S(t)$ and hazard function $h(t)$ with parameters μ and σ can be expressed respectively as:

$$f(t) = \frac{1}{\sigma t \sqrt{2\pi}} \exp\left(-\frac{(\log(t) - \mu)^2}{2\sigma^2}\right),$$

$$S(t) = 1 - F\left(\frac{\log(t) - \mu}{\sigma}\right), \quad h(t) = \frac{f(t)}{S(t)}$$

The regression model linked with the covariates for each individual i is given as:

$$T_i \sim \text{lognormal}(\mu_i(t), \sigma^2) \quad \text{with} \quad \log(\mu_i(t)) = \mathbf{X}_{2i}^T \boldsymbol{\alpha}_{2i} + W_{2i}(t) \quad (4)$$

Assuming that the survival time follows loglogistic distribution, its probability density function $f(t)$, survival function $S(t)$ and hazard function $h(t)$ with parameters with parameters λ and ρ can be expressed respectively as:

$$f(t) = \frac{\lambda \rho t^{\rho-1}}{(1 + \lambda t^\rho)^2}, \quad S(t) = \frac{1}{1 + \lambda t^\rho}, \quad h(t) = \frac{\lambda \rho t^{\rho-1}}{1 + \lambda t^\rho}$$

The regression model is linked with the covariates for each individual i as follows:

$$T_i \sim \text{Loglogistic}(\rho, \lambda_i(t)) \quad \text{with} \quad \log(\lambda_i(t)) = -\rho(\mathbf{X}_{2i} \boldsymbol{\alpha}_i + W_{2i}(t)) \quad (5)$$

The AFT models allow the direct effects of covariates on survival time instead of hazard rate. Given a vector of predictors \mathbf{X}_{2i} , the log-linear form of the AFT model for survival time T_i of individual patient $i = 1, 2, 3, \dots, n$ can be written as:

$$\log(T_i | W_{2i}) = \mathbf{X}_{2i} \boldsymbol{\alpha}_i + W_{2i}(t) + \varepsilon_i \quad (6)$$

where $\boldsymbol{\alpha}_i$ is a vector of unknown coefficients of \mathbf{X}_{2i} , $W_{2i}(t)$ refers to subject specific random effects having normal distribution, ε_i is a sequence of mutually independent measurement errors that follows AFT distributions, in this case, lognormal and loglogistic distributions.

3. Bayesian Joint Models

3.1. Likelihood Model

The association between the longitudinal and survival processes is assumed to come through stochastic dependences denoted by $W_{1i}(t)$ and $W_{2i}(t)$. There are many ways of making the linkages [2]. Here we consider the links used in [5]. Thus the joint models that link the GED based model of longitudinal process in Equation (3) to the AFT based model of survival process in Equations (4)-(6) is given as follows:

$$W_{1i}(t) = U_{1i} + U_{2i} * t \quad (7)$$

$$W_{2i}(t) = r_1 U_{1i} + r_2 U_{2i} \quad (8)$$

where r_1, r_2 are association parameters. Note that U_{1i}, U_{2i} are latent variables that are independent subject-specific random effects having bivariate Gaussian distribution with mean zeros and constant variances. These effects are assumed

to be induced by the longitudinal process to the time-to-event process through the random intercept and random slope terms in the linear mixed model.

We assume Y and T are conditionally independent given the random effects $w = (w_1, w_2)$ and model parameters $\theta = (\theta_1, \theta_2)$. The two sets of parameters are one for the linear mixed model $\theta_1 = (\beta, \gamma, \sigma_\epsilon^2, \sigma_{u1}^{-2}, \sigma_{u2}^{-2})$ and those for the survival model $\theta_2 = (\alpha, \sigma_{LN}^{-2}, r_1, r_2)$ in the lognormal or $\theta_2 = (\alpha, \rho, r_1, r_2)$ in the loglogistic case. The joint likelihood function of the data from the two processes can be given as:

$$\begin{aligned} f(y, t, \delta | \theta_1, \theta_2, w_1, w_2) \\ = \int f(y | \theta_1, w_1) f(t, \delta | \theta_2, w_2) f(w_2 | w_1) f(w_1) dw_2 dw_1 \end{aligned} \quad (9)$$

where each δ is an indicator for a patient's survival with $\delta = 1$ if death event occurs and $\delta = 0$ if censored.

3.2. Prior Distributions

Non-informative joint prior distribution of the parameters $\pi(\theta, w_1, w_2)$ is considered. Individual parameters β 's and α 's are assumed to be independently and identically normally distributed with mean zero and large variance 1000. The association parameters r_1, r_2 are each assumed to have normal distribution with mean zero and variance 1000. The shape parameter γ of GED, the shape parameter of loglogistic distribution ρ , and precision parameters $\tau = \sigma^{-2}$ all are assumed to follow Gamma(2, 0.5).

3.3. Posterior Distribution

The Bayesian model [10] is defined by the posterior distribution $\pi(\theta, w | y, t, \delta)$ of the model parameters θ and random effects w given the data and is expressed by:

$$\pi(\theta, w | y, t, \delta) = \frac{f(y, t, \delta | \theta, w) \pi(\theta, w)}{\int f(y, t, \delta | \theta, w) \pi(\theta, w) d\theta dw} \quad (10)$$

where $f(y, t, \delta | \theta, w)$ is the likelihood function, $\pi(\theta, w)$ is the prior probability distribution, and $\int f(y, t, \delta | \theta, w) \pi(\theta, w) d\theta dw$ is the normalizing constant. The main challenge here is computational difficulty. The standard maximum likelihood method involves integrating out latent variables from the log likelihood function which is difficult when the parameters are of high dimensional. Simulation simplifies the computational challenges. Here the Bayesian model in Equation (10) is computed using Markov chain Monte Carlo methods with the Gibbs sampler algorithm that is based on full conditional distributions of the parameters [7] [13] [20]. The Gibbs sampler algorithm is implemented in the WinBUGS software version 1.4 [20]. The final inferences are made based on independent samples taken from the posterior distribution after convergence of the realizations. Time series plots, auto-correlations and Gelman-Rubin statistics are used to assess and confirm convergences.

4. Results and Discussion

The objective of this study is to model the longitudinal CD4 measurement and the associated time to death data using Bayesian joint modelling approach. The generalized error distribution is assumed for the square root of the CD4 T-cell counts, while lognormal and loglogistic distributions are assumed for the survival time. Two data sets collected from two hospitals are analyzed using four Bayesian joint models. The findings from the models are all interpreted as they are important in many ways.

4.1. Descriptive Analysis

For Data 1 taken from Shashemene referral hospital, the average baseline CD4 cell counts is estimated to be 156.9 with standard deviation of 92.5 per mm³ of blood sample. By the end of study period, percentage of death event is about 5.9%. The average survival time of the patients is estimated to be 48.7 with standard deviation of 21.3 in months.

For Data 2 taken from Bale Robe general hospital, the average number of baseline CD4 counts is about 177.6 with standard deviation of 104.8 per mm³ of blood sample. Among the sample of patients considered, percentage of death event is about 11.5%. The average survival time is about 55.1 with standard deviation of 21.8 in months.

The baseline CD4 counts reveal same variabilities in the two studies which is about 59% as measured by coefficient of variation. However, it seems that there a slight difference between variabilities for time to event data with 44% for Data 1 and 40% for Data 2.

To understand the relationship between the longitudinal measure and follow-up time, mean structures are plotted in **Figure 1**. The plots show that the average of square root of CD4 counts may have a quadratic relationship with patient's follow-up time. We thus include both observation time and its square in the linear mixed models as predictor variables.

4.2. Inferential Analysis in the Case of Data 1

In the analysis of Data 1, twenty one parameters are estimated using the two defined Bayesian joint models based on GED-lognormal and GED-loglogistic distributions. The results of analysis are displayed in **Table 2** & **Table 3**.

Table 2. Parameter estimations for the Bayesian Joint GED Lognormal Model in the case of Data 1.

Parameter	Mean	SD	MC error	95% CI
β_c	13.650	0.294	0.030	(13.08, 14.22)*
β_{obt}	2.134	0.084	0.010	(1.978, 2.318)*
β_{obt^2}	-0.118	0.010	0.001	(-0.141, -0.10)*
$\beta_{sexMale}$	-0.891	0.364	0.033	(-1.609, -0.137)*
β_{fun}	-0.292	0.329	0.027	(-0.923, 0.363)

Continued

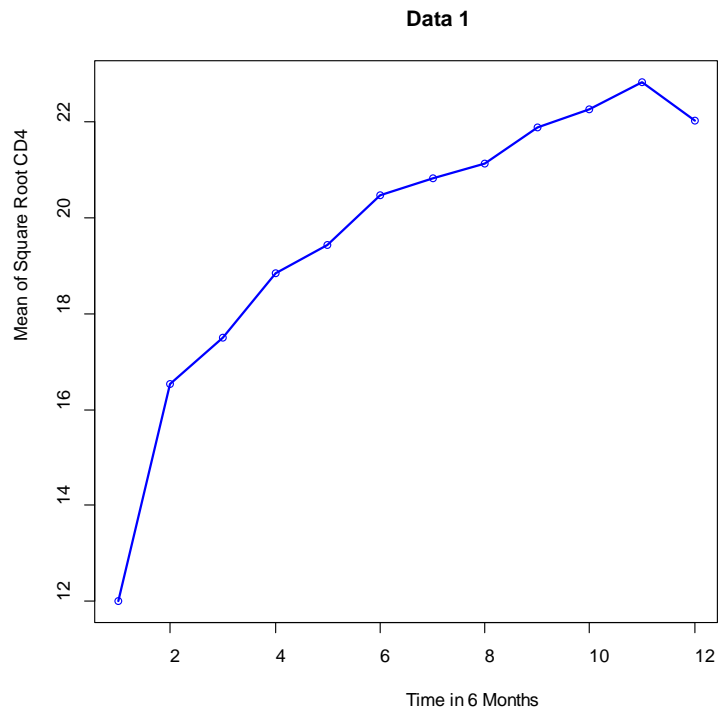
β_{alc}	-0.712	0.546	0.064	(-1.721, 0.452)
β_{tobac}	0.894	0.509	0.059	(-0.164, 1.811)
β_{ois}	0.056	0.085	0.007	(-0.110, 0.222)
γ	1.021	0.102	0.003	(0.8291, 1.227)*
σ_{ϵ}	3.051	0.074	0.002	(2.911, 3.206)*
τ_{u1}	0.097	0.010	0.000	(0.079, 0.117)*
τ_{u2}	2.363	0.339	0.026	(1.796, 3.133)*
α_0	1.368	0.014	0.001	(1.340, 1.394)*
α_{tb}	-0.072	0.015	0.000	(-0.1008, -0.044)*
α_{arrk}	0.080	0.007	0.000	(0.067, 0.095)*
α_{cond}	-0.091	0.012	0.000	(-0.114, -0.067)*
α_{ois}	0.003	0.002	0.000	(-0.002, 0.007)
α_{room}	0.003	0.004	0.000	(-0.0054, 0.0124)
τ_{LN}	7.869	0.595	0.013	(6.743, 9.081)*
r_1	0.001	0.002	0.000	(-0.003, 0.004)
r_2	0.003	0.011	0.001	(-0.018, 0.025)

*Significant at 5% significant level; SD: Standard deviation; CI: Credible interval.

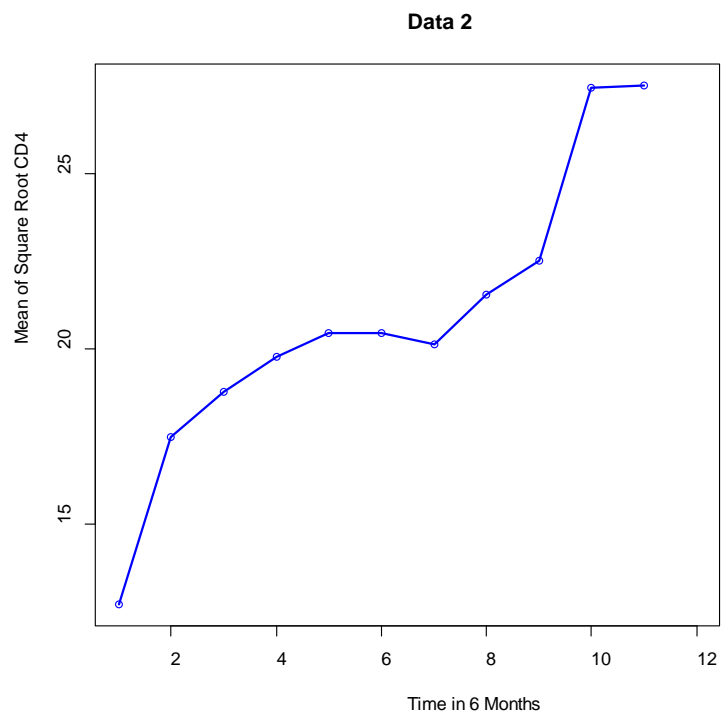
Table 3. Parameter estimations for the Bayesian Joint GED Loglogistic Model in the case of Data 1.

Parameter	Mean	SD	MC error	95% CI
β_o	13.720	0.320	0.019	(13.11, 14.38)*
β_{obt}	2.125	0.077	0.005	(1.97, 2.268)*
β_{obt^2}	-0.117	0.009	0.001	(-0.134, -0.099)*
$\beta_{sexMale}$	-0.935	0.377	0.020	(-1.64, -0.183)*
β_{fun}	-0.259	0.357	0.018	(-0.96, 0.4647)
β_{alc}	-0.771	0.582	0.040	(-1.946, 0.354)
β_{tobac}	0.938	0.574	0.039	(-0.226, 2.074)
β_{ois}	0.033	0.094	0.005	(-0.151, 0.217)
γ	1.017	0.103	0.002	(0.823, 1.229)*
σ_{ϵ}	3.047	0.074	0.001	(2.910, 3.198)*
τ_{u1}	0.097	0.010	0.000	(0.0795, 0.117)*
τ_{u2}	2.326	0.318	0.012	(1.775, 3.016)*
α_0	3.986	0.038	0.001	(3.911, 4.062)*
α_{tb}	-0.241	0.039	0.000	(-0.319, -0.164)*
α_{arrk}	0.282	0.021	0.000	(0.241, 0.323)*
α_{cond}	-0.361	0.036	0.001	(-0.4297, -0.291)*
α_{ois}	0.006	0.007	0.000	(-0.006, 0.020)
α_{room}	0.006	0.013	0.000	(-0.018, 0.031)
ρ	6.450	0.295	0.003	(5.878, 7.034)*
r_1	0.002	0.005	0.000	(-0.007, 0.012)
r_2	0.001	0.029	0.001	(-0.055, 0.058)

*Significant at 5% significant level; SD: Standard deviation; MC: Monte Carlo; CI: Credible interval.



(a)



(b)

Figure 1. Plots of mean of square roots of CD4 Counts over observed time for the two studies: (a) Data 1 and (b) Data 2.

4.2.1. Bayesian GED Lognormal Analysis

The results of analysis are displayed in **Table 2**. They reveal that the shape parameter of the GED of the longitudinal measure is significantly different from

zero ($\hat{\gamma} = 1.021, \text{CI}(0.8291, 1.227)$) at 5% level of significance. This is because the 95% credible interval does not contain zero. The standard deviation of the GED is as well significant ($\hat{\sigma}_\epsilon = 3.051, (2.911, 3.206)$).

The subject-specific random effects U_1 and U_2 are found to be significant since their respective precision parameters are significant ($\hat{\tau}_{u1} = 0.097, (0.079, 0.117)$) and ($\hat{\tau}_{u2} = 2.363, (1.796, 3.133)$). For AFT model, the precision parameter in the lognormal distribution is significant ($\hat{\tau}_{LN} = 7.869, (6.743, 9.081)$). However, the association parameters r_1 and r_2 in the joint model are not significant ($\hat{r}_1 = 0.001(-0.003, 0.004)$) and ($\hat{r}_2 = 0.003(-0.018, 0.025)$). Thus the Bayesian joint GED lognormal model is not significant.

4.2.2. Bayesian GED Loglogistic Analysis

Analysis results of the Bayesian GED loglogistic model in the case of Data 1 are displayed in **Table 3**. The parameters of the generalized error distribution are both significant at significant 5% level based on estimations of its shape parameter ($\hat{\gamma} = 1.017, (0.823, 1.229)$) and scale parameter ($\hat{\sigma}_\epsilon = 3.047, (2.910, 3.198)$). The random effects U_1 and U_2 are significant as well since their respective precision parameters are significant ($\hat{\tau}_{u1} = 0.097, (0.0795, 0.117)$) and ($\hat{\tau}_{u2} = 2.326, (1.775, 3.016)$). In the survival sub-model, the shape parameter of the loglogistic distribution is significant ($\hat{\rho} = 6.450, (5.878, 7.034)$). Similar to the lognormal case, the Bayesian joint GED loglogistic model is not significant since the association parameters r_1 and r_2 are both insignificant ($\hat{r}_1 = 0.002, (-0.007, 0.012)$) and ($\hat{r}_2 = 0.001, (-0.055, 0.058)$).

Comparing their total DIC values, the Bayesian GED loglogistic model fits better to Data 1 than the lognormal case. From the analysis of Bayesian GED loglogistic model, the covariates that are found to affect the longitudinal CD4 measure are: observed time, squared observed time and gender. They imply that the disease marker improves over time but later reaches maximum and then declines. Female patients gain more CD4 counts as compared to the males. Effects of the functional status, alcohol use, tobacco use, and opportunistic infection status of patients are not statistically significant. From the survival sub-model, the results show that TB infection status at baseline, awareness about ART and condom use have significant effects on the survival time of a patient.

4.3. Inferential Analysis in the Case of Data 2

Here we analyze Data 2 and estimate twenty parameters of the same two defined models: Bayesian joint GED-lognormal and Bayesian GED-loglogistic models. The results are given in **Table 4** & **Table 5**.

4.3.1. Bayesian GED Lognormal Analysis

The results of analysis in **Table 4** show that the shape parameter of the generalized growth distribution is significantly different from zero ($\hat{\gamma} = 0.05244, (0.00209, 0.1442)$). The standard deviation of the GED is signifi-

cant ($\hat{\sigma}_\epsilon = 5.006, (4.846, 5.176)$) and relatively larger than those obtained from Data 1 case. The subject-specific random effects U_1 and U_2 are significant due to the fact that their respective precision parameters are significant ($\hat{\tau}_{u1} = 4.764, (0.547, 15.47)$) and ($\hat{\tau}_{u2} = 5.231, (0.866, 12.12)$).

For the survival sub-model, the precision of the lognormal distribution is significant ($\hat{\tau}_{LN} = 7.43, (5.678, 10.53)$). Again, the association parameters r_1 and r_2 are not significant ($\hat{r}_1 = (-0.0047, (-0.1297, 0.101))$ and ($\hat{r}_2 = -0.010, (-0.144, 0.178)$) and hence the Bayesian joint GED lognormal model is not significant in the case of Data 2.

The covariates observed time, square of observed time, sex, age, weight and number of opportunistic infection have statistically significant effects on CD4 counts of the HIV/AIDS patients. For survival part, age, functional status, tobacco use, and condom use have effects on survival time of the patients. Patient's weight is not significant in this case.

4.3.2. Bayesian GED Loglogistic Analysis

The results are displayed in Table 5. The shape parameter of the generalized

Table 4. Parameter estimations for the Bayesian Joint GED Lognormal Model in the case of Data 2.

Parameter	Mean	SD	MC error	95% CI
β_o	14.31	0.2427	0.02007	(13.84, 14.77)*
β_{obr}	2.897	0.1616	0.01875	(2.573, 3.196)*
β_{obr^2}	-0.2712	0.02531	0.003028	(-0.321, -0.225)*
$\beta_{sexMale}$	-1.627	0.2513	0.007866	(-2.123, -1.124)*
β_{age}	-0.6367	0.1199	0.003315	(-0.867, -0.405)*
β_{wt}	0.7092	0.1243	0.00304	(0.462, 0.949)*
β_{ois}	-0.4647	0.1139	0.002709	(-0.68, -0.241)*
γ	0.05244	0.03823	0.00091	(0.002, 0.144)*
σ_ϵ	5.006	0.08456	0.0018	(4.846, 5.176)*
τ_{u1}	4.764	3.845	0.3743	(0.547, 15.47)*
τ_{u2}	5.231	3.04	0.3509	(0.8663, 12.12)*
α_0	1.27	0.01391	0.00063	(1.243, 1.298)*
α_{age}	-0.01079	0.005444	0.00011	(-0.022, -4.1E-05)*
α_{wt}	0.00933	0.005371	0.00012	(-0.0014, 0.020)
α_{fun}	0.0194	0.006951	0.00028	(0.005, 0.033)*
α_{tobac}	-0.03173	0.006202	0.00012	(-0.044, -0.020)*
α_{cond}	0.09406	0.0109	0.00037	(0.072, 0.115)*
τ_{LN}	7.43	1.271	0.09489	(5.678, 10.53)*
r_1	-0.004651	0.05618	0.0051	(-0.130, 0.101)
r_2	-0.01015	0.08735	0.01037	(-0.144, 0.178)

*Significant at 5% significant level; SD: Standard deviation; MC: Monte Carlo; CI: Credible interval.

Table 5. Parameter estimations for the Bayesian Joint GED Logistic Model in the case of Data 2.

Parameter	Mean	SD	MC error	95% CI
β_o	14.270	0.2212	0.01681	(14.280, 14.690)*
β_{obt}	2.939	0.1494	0.01712	(2.930, 3.259)*
β_{obt^2}	-0.299	0.0243	0.00293	(-0.302, -0.243)*
$\beta_{sexMale}$	-1.540	0.2518	0.00999	(-1.538, -1.020)*
β_{age}	-0.603	0.1197	0.00336	(-0.603, -0.362)*
β_{wt}	0.708	0.1258	0.00428	(0.710, 0.955)*
β_{ois}	-0.486	0.1152	0.00351	(-0.488, -0.260)*
γ	0.093	0.0455	0.00121	(0.089, 0.190)*
σ_ϵ	4.778	0.0860	0.00262	(4.778, 4.95)*
τ_{u1}	4.705	3.6040	0.41920	(3.939, 15.05)*
τ_{u2}	4.605	2.7750	0.34590	(4.356, 10.15)*
α_0	3.586	0.0584	0.00402	(3.587, 3.695)*
α_{age}	-0.041	0.0229	0.00089	(-0.041, 0.005)
α_{wt}	0.046	0.0233	0.00097	(0.046, 0.093)*
α_{jun}	0.061	0.0292	0.00183	(0.060, 0.120)*
α_{tobac}	-0.111	0.0248	0.00098	(-0.111, 0.063)
α_{cond}	0.371	0.0442	0.00210	(0.372, 0.458)*
ρ	5.559	0.4328	0.02653	(5.526, 6.539)*
r_1	-0.021	0.1464	0.01219	(-0.005, 0.238)
r_2	0.627	0.2127	0.02693	(0.643, 1.005)*

*Significant at 5% significant level; SD: Standard deviation; MC: Monte Carlo; CI: Credible interval.

growth distribution is significant ($\hat{\gamma} = 0.093, (0.089, 0.190)$) and its standard deviation is also significant ($\hat{\sigma}_\epsilon = 4.778, (4.778, 4.95)$). The subject-specific random effects U_1 and U_2 are significant because their respective precision parameters are so ($\hat{\tau}_{u1} = 4.705, (3.939, 15.05)$) and ($\hat{\tau}_{u2} = 4.605, (4.356, 10.15)$).

For the survival sub-model, the precision parameter of the logistic distribution is significant ($\hat{\rho} = 5.559, (5.526, 6.539)$). Only the slope term r_2 of the association parameters is significant ($\hat{r}_2 = 0.627, (0.643, 1.005)$) but not the intercept ($\hat{r}_1 = -0.021, (-0.005, 0.238)$). The Bayesian joint GED loglogistic model is significant implying that the joint model is important for this data set.

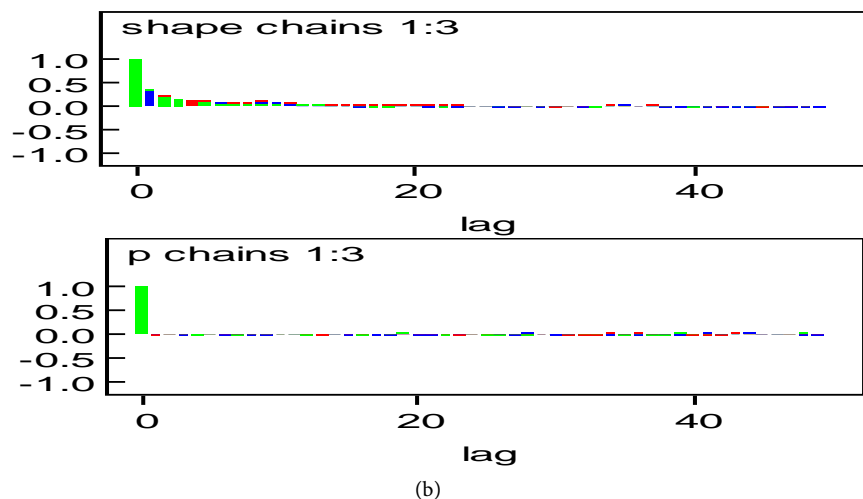
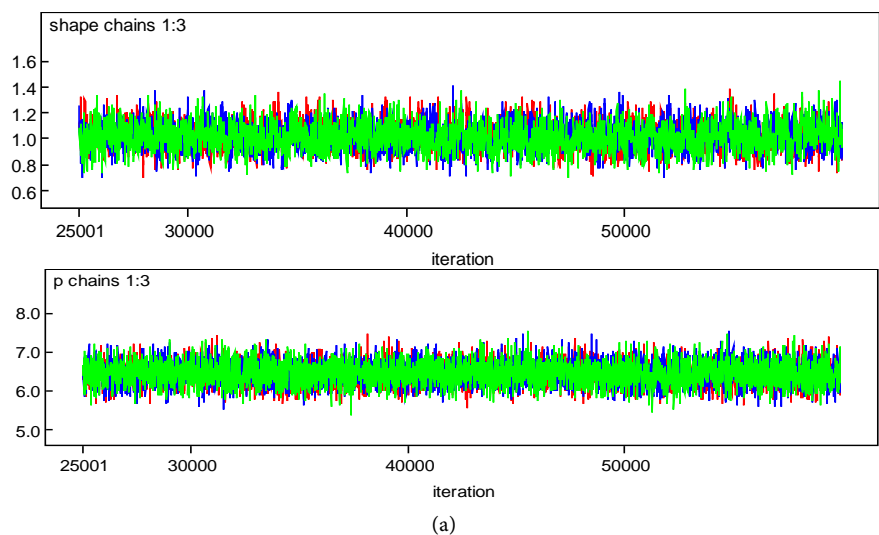
The Bayesian joint GED logistic model fits better to the data than the log-normal case based on the total DIC values. From analysis of Data 2 with the selected loglogistic model, the covariates that are found to significantly affect longitudinal CD4 measure of an HIV/AIDS patient are: observed time, square of observed time, sex, age, weight and number of opportunistic infection. With increasing ART follow-up time, CD4 counts of a patient can get parabolic mean growth with reaching maximal point. Female patients achieve higher CD4

counts on average than males. Moreover, the mean CD4 counts of a patient declines as patient's age increases, weight decreases and number of opportunistic infection increases.

For survival sub-model, the results reveal that improving a patient's weight improves her/his survival time. Note that weight is an important variable in explaining both the longitudinal and survival processes of a patient. Healthy functional status and condom use during sexual intercourse have also positive effects on survival time. Age and tobacco use are not significant for this data case.

4.4. Assessment of Convergence

For each of the Bayesian models, three parallel sampling chains of 60000 iterations with different starting values are generated. Some plots are given in **Figure 2**. Inferences are made based on samples of the posterior distributions that are taken with thinning of 10 after burn-in of 25000. Time series plots of the history of the simulations show a reasonable degree of randomness and they may converge to same values. Auto-correlations and Gelman-Rubin statistics are also used to assess convergences. Finally independent samples are taken from the



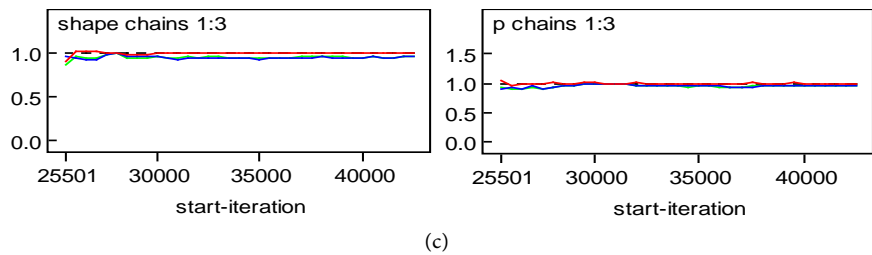


Figure 2. Plots from analysis of Data 1 using the Bayesian GED Loglogistic. (a) Time Series plots of simulations of shape parameter γ of GED and shape parameter ρ of loglogistic distribution; (b) Autocorrelation plots; (c) Plots of Gelman-Rubin statistics.

posterior distribution after convergence of the realization with specified burn-in and thinning values, and then all inferences are made using those samples.

Assessment plots are displayed in **Figure 2** from the analysis of Data 1 using the Bayesian GED loglogistic model. Results of two parameters: shape parameter γ of GED and shape parameter ρ of loglogistic distribution are illustrated. They show that the simulations converge.

The findings reveal that the generalized error distribution for both data sets has positive estimate of the shape parameter and so it is of fatter tails than normal distribution. There is higher variation on the CD4 T-cell counts for the data from Bale Robe general hospital (about 49%) than that obtained from Shashemene referral hospital (about 10%).

The posterior distributions estimated under the selected models, Bayesian GED loglogistic models, are the solutions required in this analysis. Though the association parameter in the joint model is significant for one data but not for the other data case, both fitted models are still important to consider as they are newly defined models implementing the generalized error distribution. Thus the respective results are used to report findings on how the longitudinal CD4 counts and survival time of a patient are related and parameter estimations.

The findings from this study are fairly consistent with the studies by [5] except for the type of models selected. They suggested different models for the two data sets while only one is recommend here. This may be expected as the GED is involved in this study and as indicated by [4] the GED model can gain more insight on the error distributions than the normal growth curve models.

5. Conclusions

The current study focuses on developing Bayesian joint models with the assumption of generalized error distribution for the longitudinal CD4 observations and of two AFT distributions for the survival time of HIV/AIDS patients. Analyses of two different data sets show that measurement errors of the longitudinal CD4 variable are not normally distributed and so are modelled by the generalized error distribution. The distributions have fatter tails than the normal distribution. The Bayesian joint GED loglogistic models are found to be important models in fitting to the data sets. Fairly consistent estimates of parameters and

more insights of the data are obtained from the models. In one of the data sets, it is found that survival time of a patient is affected by the latent variables generated from the longitudinal CD4 T-cell counts.

Covariates with significant effects are identified from analysis of the Bayesian GED loglogistic models. The findings reveal that under ART follow-up, patients' health can be improved over time. Female patients gain more CD4 counts as compared to the males. Survival time of a patient is negatively affected by TB infection. Awareness about ART is an important factor as well.

Increase in number of opportunistic infection implies decline of CD4 counts. Age of a patient negatively affects the disease marker but no effects on the survival time. Weight loss is related with decline of CD4 counts and shortening of survival time on average. Improving a patient's weight may improve her/his survival time. Condom use is positively related with survival time of a patient.

Bayesian joint models with GED and AFT distributions are found to be useful in modelling the longitudinal and survival processes. Bayesian computations of such complex models are well handled in the WinBUGS software by providing simulations and estimations of the posterior distributions. We recommend Bayesian joint models with generalized error distributions when measurement errors of the longitudinal data are not necessarily normally distributed. Further studies may investigate the models with various types of shared random effects, more covariates and prediction of future observations.

Acknowledgements

The authors would like to sincerely thank the Shashemene Referral and Bale Robe General Hospitals for providing us the data sets used in this study. We acknowledge the anonymous reviewers for their detailed comments and suggestions.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Sousa, I. (2011) A Review on Joint Modelling of Longitudinal Measurements and Time-to-Event. *Statistical Journal*, **9**, 57-81.
- [2] Henderson R., Diggle, P. and Dobson, A. (2000) Joint Modeling of Longitudinal Measurements and Event Time Data. *Biostatistics*, **4**, 465-480.
<https://doi.org/10.1093/biostatistics/1.4.465>
- [3] Hatfield, A., Hodges, S. and Carlin, P. (2012) Combining Longitudinal and Survival Information in Bayesian Joint Models: When Are Treatment Estimates Improved? *Biostatistics*, 1-31.
- [4] Zhang, Z.Y. (2013) Bayesian Growth Curve Models with the Generalized Error Distribution. *Journal of Applied Statistics*, **40**, 1779-1795.
<https://doi.org/10.1080/02664763.2013.796348>

- [5] Erango, M.A., Goshu, A.T., Buta, G.B. and Dessiso, A.H. (2017) Bayesian Joint Modeling of Survival of HIV/AIDS Patients Using AFT Data and Longitudinal CD4 Cell Counts. *British Journal of Medicine & Medical Research*, 20, 1-12. <https://doi.org/10.9734/BJMMR/2017/32123>
- [6] Buta, G.B., Goshu, A.T. and Worku, H.M. (2014) Bayesian Joint Modelling of Disease Progression Marker and Time to Death Event of HIV/AIDS Patients under ART Follow-Up. *British Journal of Medicine & Medical Research*, 5, 1034-1043. <https://doi.org/10.9734/BJMMR/2015/12907>
- [7] Dessiso, A.H. and Goshu, A.T. (2017) Bayesian Joint Modelling of Longitudinal and Survival Data of HIV/AIDS Patients: A Case Study at Bale Robe General Hospital, Ethiopia. *American Journal of Theoretical and Applied Statistics*, 6, 182-190. <https://doi.org/10.11648/j.ajtas.20170604.13>
- [8] Wu, L. (2009) Mixed Effects Models for Complex Data. CRC Press, Boca Raton. <https://doi.org/10.1201/9781420074086>
- [9] Viviani, S. (2012) Mixed Effect Joint Models for Longitudinal Responses with Dropout: Estimation and Sensitivity Issues. Thesis, Department of Statistics, Sapienza—University of Rome, Rome, 1-83.
- [10] Box, G.E.P. and Tiao, G.C. (1973) Bayesian Inference in Statistical Analysis. Wiley, New York.
- [11] Nadarajah, S. (2005) A Generalized Normal Distribution. *Journal of Applied Statistics*, 32, 685-694. <https://doi.org/10.1080/02664760500079464>
- [12] Choy, S.T.B. and Smith, A.F.M. (1997) On Robust Analysis of a Normal Location Parameter. *Journal of the Royal Statistical Society. Series B*, 59, 463-474. <https://doi.org/10.1111/1467-9868.00079>
- [13] Purczynski, J. (2003) The Use of Computer Simulations in the Estimation of Selected Econometric and Statistical Models. Publishing Series Rozprawy I Studia (DXXV) 451, University of Szczecin, Szczecin.
- [14] Delignette-Muller, M.L. and Dutang, C. (2015) Fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64, 1-34. <https://doi.org/10.18637/jss.v064.i04>
- [15] Subbotin, M.T. (1923) On the Law of Frequency of Error. *Matematicheski Sbornik*, 31, 296-301.
- [16] Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59, 347-370. <https://doi.org/10.2307/2938260>
- [17] Klein, J.P. and Moeschberger, M.L. (2003) Survival Analysis Techniques for Censored and Truncated Data. Second Edition, Statistics for Biology and Health, Springer, Berlin.
- [18] Guo, X. and Carlin, B.P. (2004) Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages. *The American Statistician*, 58, 16-24. <https://doi.org/10.1198/0003130042854>
- [19] Tseng, Y., Hsieh, F. and Wang, J.L. (2005) Joint Modelling of Accelerated Failure Time and Longitudinal Data. *Biometrika*, 92, 587-603. <https://doi.org/10.1093/biomet/92.3.587>
- [20] Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003) WinBUGS Version 1.4 User Manual. <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf>