

Bayesian estimation of generalized gamma shared frailty model

Sukhmani Sidhu¹ · Kanchan Jain¹ ·
Suresh Kumar Sharma¹

Received: 8 July 2016 / Accepted: 20 April 2017 / Published online: 3 May 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract Multivariate survival analysis comprises of event times that are generally grouped together in clusters. Observations in each of these clusters relate to data belonging to the same individual or individuals with a common factor. Frailty models can be used when there is unaccounted association between survival times of a cluster. The frailty variable describes the heterogeneity in the data caused by unknown covariates or randomness in the data. In this article, we use the generalized gamma distribution to describe the frailty variable and discuss the Bayesian method of estimation for the parameters of the model. The baseline hazard function is assumed to follow the two parameter Weibull distribution. Data is simulated from the given model and the Metropolis–Hastings MCMC algorithm is used to obtain parameter estimates. It is shown that increasing the size of the dataset improves estimates. It is also shown that high heterogeneity within clusters does not affect the estimates of treatment effects significantly. The model is also applied to a real life dataset.

Keywords Generalized gamma distribution · Gauss–Laguerre quadrature · Metropolis–Hastings algorithm · MCMC algorithm · Credible intervals

✉ Kanchan Jain
jaink14@gmail.com
Sukhmani Sidhu
sukhmani.15@gmail.com
Suresh Kumar Sharma
ssharma643@yahoo.co.in

¹ Department of Statistics, Panjab University, Chandigarh 160014, India

1 Introduction

In survival studies, we often come across data that occurs in clusters or groups that arise due to some common factor associated with each group. For example, a cluster may refer to a set of event times for the same individual or event times belonging to related individuals (parent–child, siblings or twins). Sometimes there may be a set of event times belonging to individuals suffering from a particular disease in the same community or being treated in the same hospital under similar conditions like exposure to the same environment or treatment. In all of these cases, there is reason to group together certain event times. This stratification helps to analyse time to event data in a more effective way.

Proportional Hazards models introduced by [Cox \(1972\)](#) assumed that the study population is homogeneous given some observed covariates. However, this may not be true in general and individuals may differ in many aspects. For instance, two individuals at the same stage of disease progression, receiving same treatment, having no significant difference in the covariates may respond to the same treatment in a significantly different manner. This can be attributed to different individuals having different frailties. In general, more frail individuals in a population die earlier than those who are less frail. Hence, heterogeneity due to unknown factors can be attributed to an individual's frailty and is incorporated in the model as a randomness factor. It modifies the hazard function for each individual and hence increases the precision with which we study the effect of the known covariates.

In particular, shared frailty models assume that the value of frailty is same for each observation in a cluster and frailty levels of different clusters are independent of each other. In other words, there is dependence between event times in a cluster but there is independence between observations from different clusters.

Frailty (named as such by [Vaupel et al. 1979](#)) was studied by [Greenwood and Yule \(1920\)](#) for recurrent data and by [Clayton \(1978\)](#) for bivariate data. Shared frailty models have been extensively studied with different choices of distribution viz: gamma distribution ([Clayton 1978](#); [Vaupel et al. 1979](#); [Klein 1992](#); [Yashin et al. 1995](#)), log-normal distribution ([McGilchrist and Aisbett 1991](#); [Xue and Brookmeyer 1996](#); [Yau and McGilchrist 1997](#); [Ripatti and Palmgren 2000](#); [Vaida and Xu 2000](#)), and positive stable distribution ([Hougaard 1986a,b](#); [Oakes 1994](#); [Hougaard 1995](#); [Lam and Kuk 1997](#)).

In order to obtain efficient estimators for the regression coefficients and also to recognize covariates which are significantly affecting the event times, it is important to make the most appropriate choice for the frailty distribution. For this reason, flexible distributions have been used to describe the frailty variable. Some of the flexible frailty distributions that have been studied in literature are the positive stable distribution ([Hougaard 1986a](#)), generalized gamma distribution ([Balakrishnan and Peng 2006](#); [Pengcheng et al. 2013](#)), compound Poisson distribution ([Aalen 1992](#); [Wienke et al. 2003](#)) and log-skew normal distribution ([Callegaro and Iacobelli 2012](#)).

In this paper, we consider the generalized gamma distribution (GGD) to describe the frailty term effect. This distribution has many sub models as its special cases. Four popular sub-models of the GGD are the Weibull, exponential, log-normal and gamma distribution. Fitting a flexible distribution helps in improving the accuracy of

the estimates obtained and also saves time and cost incurred in fitting survival data to each of these distributions independently to come up with the most efficient model. We apply a Monte Carlo Markov Chain (MCMC) procedure to obtain Bayesian estimates of the parameters of the model. We perform a simulation study to observe the effect of sample size, cluster size and frailty variance on the efficiency of the estimates obtained in terms of their biases and mean squared errors.

Similar methodology has been successfully applied to survival studies previously (Clayton 1991; Zeger and Karim 1991; Dellaportas and Smith 1993; Ducrocq and Casella 1996; Golightly and Wilkinson 2008). Bayesian estimation methods have also found use in the study of random effects model in many other fields. Khoshravesh et al. (2015) apply Bayesian regression to study evapotranspiration in three arid environments. In econometric studies, similar methodology has been applied to longitudinal data by Koop et al. (1997) to make inference about firm specific inefficiencies and by Allenby and Rossi (1998) to model consumer heterogeneity. Bayesian estimation procedures have also been applied in meta analysis wherein multiple studies have been combined (Bodnar et al. 2017; Higgins et al. 2009; Pullenayegum 2011; Turner et al. 2015).

Earlier work done on the generalized gamma shared frailty model (Balakrishnan and Peng 2006; Pengcheng et al. 2013) makes use of the Newton Raphson algorithm or the EM algorithm to obtain Maximum Likelihood estimates. These methods are computationally lengthy especially when the number of treatment effects is large and convergence for such methods may take a long time if initial parameter values in the algorithm are far from the true parameter values. Also, it is never certain whether convergence will happen. MCMC methods like Metropolis–Hastings algorithm can be run for shorter durations and problems like multiple modes and non-convergence etc. can be diagnosed earlier. Bayesian estimation also allows us to incorporate any prior information that may be available about treatment effects from previous studies. Hence, we use MCMC algorithm to obtain Bayesian estimates of the model parameters.

In Sect. 2, the general shared frailty model is defined. Some basic properties of the GGD and its special cases are presented in Sect. 3. In Sect. 4, the generalized gamma shared frailty model is proposed and an estimation method formulated in Sect. 5. Simulations have been carried out in Sect. 6 followed by a real life illustration in Sect. 7 wherein we fit the model to exercise times data set of coronary heart patients (Danahy et al. 1977).

2 Shared frailty model with censoring

Consider a case where continuous random variable T is the lifetime of an individual and random variable C is the censoring time. Let random variable Z be the frailty variable, $\mathbf{X} = (X_1, X_2, \dots, X_r)'$ denote a vector of 'r' known covariates and δ be the censoring indicator.

Suppose there are M clusters each having n_i observations ($i = 1, 2, \dots, M$) and associated with each cluster is the unobserved frailty z_i . The clusters can represent hospitals, schools or cities and n_i survival times in the i th cluster can represent time to event (death or recovery) of patients suffering from a particular disease, time to event

(completing a test or passing a grade) for a student, or time to demographical events (marriage or divorce) for the urban population.

The survival data are right censored such that $t_{ij} = \min \{ T_{ij}, C_{ij} \}$ denotes the survival time and

$$\delta_{ij} = \begin{cases} 1 & \text{if } T_{ij} \leq C_{ij}, \text{ that is, event is not censored} \\ 0 & \text{if } T_{ij} > C_{ij}, \text{ that is, event is censored.} \end{cases}$$

denotes the censoring indicator.

The shared frailty model assumes that all the event times in a cluster are conditionally independent given the frailties and event times from different clusters are independent. In case of censoring and for given frailty $Z = z_i$, the conditional hazard function for i th cluster at time $t_{ij} > 0$ is written as

$$h(t_{ij}|z_i, \mathbf{X}_{ij}) = z_i h_0(t_{ij}) e^{\mathbf{X}'_{ij}\boldsymbol{\beta}} \tag{1}$$

$j = 1, 2, \dots, n_i$ and $i = 1, 2, \dots, M$

where,

$h_0(t_{ij})$ is the baseline hazard function;

\mathbf{X}_{ij} are $r \times 1$ column vectors of known covariates of the j th subject in the i th cluster;

$\boldsymbol{\beta}$ is a $r \times 1$ column vector of regression coefficients.

The conditional survival function for a given frailty at time $t_{ij} > 0$ is

$$\begin{aligned} S(t_{ij}|z_i, \mathbf{X}_{ij}) &= e^{-\int_0^{t_{ij}} h(u|z_i, \mathbf{X}_{ij}) du} \\ &= e^{-\int_0^{t_{ij}} z_i h_0(u) e^{\mathbf{X}'_{ij}\boldsymbol{\beta}} du} \\ &= e^{-z_i e^{\mathbf{X}'_{ij}\boldsymbol{\beta}} \int_0^{t_{ij}} h_0(u) du} \\ &= e^{-z_i H_0(t_{ij}) e^{\mathbf{X}'_{ij}\boldsymbol{\beta}}} \end{aligned} \tag{2}$$

where $H_0(t_{ij})$ is the cumulative hazard function.

3 Generalized gamma distribution

A random variable $Z > 0$ is said to follow generalized gamma distribution with parameters b, d , and k (GGD(b, d, k)) if its probability density function is given by

$$f(z : b, d, k) = \frac{d}{\Gamma(k) b} \left(\frac{z}{b}\right)^{dk-1} e^{-\left(\frac{z}{b}\right)^d} \tag{3}$$

where $d > 0, k > 0$ are the shape parameters and $b > 0$ is the scale parameter.

Although the GGD is complex to handle in its likelihood form, it could be quite useful since it includes other distributions as special cases (Khodabin and Ahmadabadi

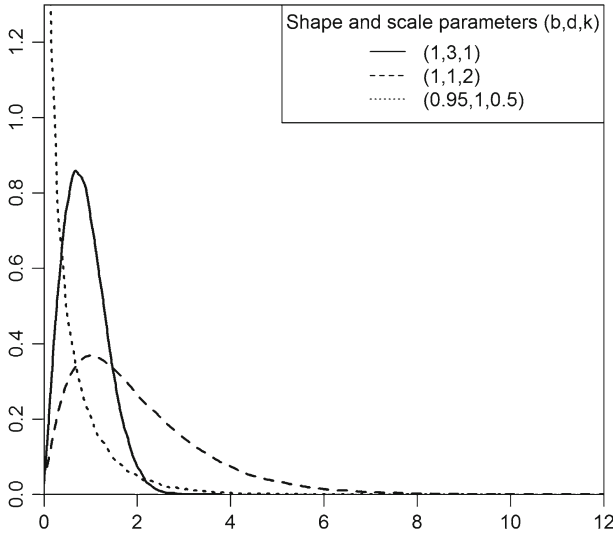


Fig. 1 Density functions of GGD for different parametric combinations

2010) based on the values of the parameters. For example, the distribution reduces to exponential distribution when $d = k = 1$, gamma distribution when $d = 1$, Weibull distribution when $k = 1$, log-normal distribution as a limiting case when $k \rightarrow \infty$. Also by setting $d = 2$, we can obtain another sub-family known as the generalized normal distribution which itself being a flexible distribution includes the half-normal ($k = 1/2, b^2 = 2\sigma^2$), Rayleigh ($k = 1, b^2 = 2\sigma^2$), Maxwell–Boltzmann ($k = 3/2$), and chi square ($k = n/2, n = 1, 2, \dots$).

Figure 1 shows the shapes of the probability density function of GGD for different parametric values. The distribution is suitable for a variety of cases. As Hougaard (1995, 2000) shows, the tails of the frailty distribution can determine the type of dependence among correlated observations. A strong early (late) dependence is indicated by a long right (left) tail. As indicated by Fig. 1, generalized gamma distribution when chosen as a distribution to model frailty, can handle varying degrees of early dependencies.

The mean and variance of the GGD(b,d,k) are

$$E(z) = \frac{b\Gamma(k + \frac{1}{d})}{\Gamma(k)}$$

$$V(z) = \frac{b^2\Gamma(k + \frac{2}{d})}{\Gamma(k)} - \left(\frac{b\Gamma(k + \frac{1}{d})}{\Gamma(k)}\right)^2.$$

In order to make the parameters of the model identifiable, we need to set the mean of the frailty distribution to one. Hence,

$$E(z) = 1 \Rightarrow b = \frac{\Gamma(k)}{\Gamma(k + \frac{1}{d})}.$$

The variance of the frailty distribution obtained after replacing the value of b is given by

$$V(z) = \frac{\Gamma(k) \Gamma(k + \frac{2}{d})}{\Gamma(k + \frac{1}{d})^2} - 1.$$

4 The generalized gamma shared frailty model

4.1 Baseline distribution

Weibull distribution is one of the most widely used distributions in life time data analysis. Its versatility comes from the various shapes of the hazard function for different values of the shape parameter ψ . The hazard rate of the Weibull distribution decreases with time (early life or infantile high mortality) when shape parameter is less than 1. It is fairly constant (useful or mid life) for shape parameter equal to or close to 1 and increases with time (wear out or old age) for shape parameter greater than 1. These three sections comprise the “bath tub curve” and are indicative of human survival rates from birth to death. Its probability density function is defined by

$$f(t) = \psi \omega t^{\psi-1} e^{-\omega t^\psi},$$

where $\psi > 0$ is the shape parameter and $\omega > 0$ is the scale parameter. The baseline hazard function is given by

$$h_0(t) = \omega \psi t^{\psi-1}. \quad (4)$$

The cumulative baseline hazard function is given by

$$H_0(t) = \omega t^\psi. \quad (5)$$

4.2 Likelihood function

The full conditional likelihood function for the generalized gamma frailty model is given by the following equation

$$L_c = \prod_{i=1}^M \prod_{j=1}^{n_i} [h(t_{ij}|z_i, \mathbf{X}_{ij})]^{\delta_{ij}} S(t_{ij}|z_i, \mathbf{X}_{ij}), \quad (6)$$

where the baseline hazard function and the conditional survival function are given by (1) and (2) respectively.

The full unconditional likelihood function obtained by integrating the likelihood function over the entire range of the random variable Z , is given by

$$L = \int_0^\infty \prod_{i=1}^M \prod_{j=1}^{n_i} [h(t_{ij}|z_i, \mathbf{X}_{ij})]^\delta S(t_{ij}|z_i, \mathbf{X}_{ij}) f(z_i) dz_i.$$

Using Eqs. (1) and (2) and writing $\eta_{ij} = e^{\mathbf{X}'_{ij}\beta}$, we have

$$\begin{aligned} L &= \prod_{i=1}^M \left\{ \int_0^\infty \left(\prod_{j=1}^{n_i} [z_i h_0(t_{ij}) e^{\mathbf{X}'_{ij}\beta}]^\delta e^{-z_i H_0(t_{ij}) e^{\mathbf{X}'_{ij}\beta}} \right) f(z_i) dz_i \right\} \\ &= \prod_{i=1}^M \left\{ \left(\prod_{j=1}^{n_i} [h_0(t_{ij}) \eta_{ij}]^\delta \right) \int_0^\infty \left(\prod_{j=1}^{n_i} z_i^\delta e^{-z_i H_0(t_{ij}) \eta_{ij}} \right) f(z_i) dz_i \right\} \\ &= \prod_{i=1}^M \left\{ \left(\prod_{j=1}^{n_i} [h_0(t_{ij}) \eta_{ij}]^\delta \right) \int_0^\infty z_i^{\sum_{j=1}^{n_i} \delta} e^{-z_i \sum_{j=1}^{n_i} H_0(t_{ij}) \eta_{ij}} f(z_i) dz_i \right\} \\ &= \prod_{i=1}^M \left\{ \left(\prod_{j=1}^{n_i} [h_0(t_{ij}) \eta_{ij}]^\delta \right) \int_0^\infty z_i^{D_i} e^{-z_i A_i} f(z_i) dz_i \right\}, \end{aligned} \tag{7}$$

where $D_i = \sum_{j=1}^{n_i} \delta_{ij}$ and $A_i = \sum_{j=1}^{n_i} H_0(t_{ij}) \eta_{ij}$.

Writing the integral in (7) as I_i and putting $t_i = z_i A_i$, we have

$$\begin{aligned} I_i &= \int_0^\infty \left(\frac{t_i}{A_i} \right)^{D_i} e^{-t_i} f\left(\frac{t_i}{A_i}\right) \frac{1}{A_i} dt_i, \\ &= \frac{1}{A_i^{D_i+1}} \int_0^\infty t_i^{D_i} e^{-t_i} f\left(\frac{t_i}{A_i}\right) dt_i \\ &\approx \frac{1}{A_i^{D_i+1}} \sum_n w_n u_n^{D_i} f\left(\frac{u_n}{A_i}\right), \end{aligned} \tag{8}$$

using the Gauss-Laguerre quadrature rule, where w_n and u_n are the n weights and nodes of the Laguerre polynomial.

Hence the full unconditional likelihood function reduces to

$$L = \prod_{i=1}^M \left\{ \left(\prod_{j=1}^{n_i} [h_0(t_{ij}) \eta_{ij}]^\delta \right) I_i \right\}, \tag{9}$$

where I_i is as given in (8).

5 Estimation procedure

Often methods like Maximum Likelihood (ML) and Expectation–Maximization (EM) have been used in frailty models. However, in our case, the integral is not in a closed form and this complicates the estimation procedure. Earlier, authors have approximated the integral using Monte Carlo simulations and used Newton Raphson algorithm to obtain ML estimates (Balakrishnan and Peng 2006) or used EM algorithm for estimation with MCMC Metropolis algorithm to calculate conditional expectations (Pengcheng et al. 2013). In this paper, the Random Walk Metropolis–Hastings algorithm, one of the MCMC methods, is used to obtain Bayesian estimates of the parameters of the model.

In the Bayesian approach, the parameters of the model $\theta = (\psi, \omega, \beta_1, \dots, \beta_r, d, k)$ are treated as random variables. A prior distribution $p(\theta)$ is used to describe the currently available information about the parameter of interest, θ . Data collected from a study independent of this information is then incorporated into the model to draw conclusions about θ . Ideally the prior distribution should cover all reasonable values that θ can take, but the distribution need not play a significant role in the posterior distribution. This is because information from the data collected in the study will far outweigh any prior information available or guesses made about θ , in the form of prior probabilities specified. Hence, we use “reference priors”, that can be described as vague, diffused, or flat, non-informative priors (Gelman et al. 2013). The rationale for this is to let the data play a more significant role so that the inferences made are driven by the data.

A gamma prior, $G(\phi, \phi)$ is used for ψ , ω , d and k whereas a normal prior $N(0, \epsilon^2)$ is used for regression coefficients β_i . For simulations, we set the hyperparameters $\phi = 0.001$ and $\epsilon^2 = 1000$, such that the variance is very large. Similar types of prior distributions have been used earlier (Ibrahim et al. 2001; Sahu et al. 1997; Santos and Achcar 2010; Hanagal and Dabade 2013).

Hence, using the full likelihood function $L(\theta|x)$ with the prior distribution, we obtain the updated posterior distribution of the parameters as

$$\pi(\theta|x) \propto L(\theta|x)p(\theta). \quad (10)$$

Under the assumption that all the parameters of the model, ψ , ω , β_1, \dots, β_r , d , and k are independently distributed, the joint posterior density is

$$\pi(\psi, \omega, \beta, d, k) \propto L(\psi, \omega, \beta, d, k)g_1(\psi)g_2(\omega)g_3(d)g_4(k) \prod_{i=1}^r p_i(\beta_i), \quad (11)$$

where $g(\cdot)$ is the prior density function of ψ , ω , d and k with known hyperparameters. $p_i(\cdot)$ is the prior density function with known hyperparameters of β_i . The likelihood function is as given by Eq. (9).

Given the likelihood and priors, we can obtain the full conditional distribution of these parameters by considering only those terms which include that parameter. For

example, for ψ we have

$$\pi(\psi|\omega, \boldsymbol{\beta}, d, k) \propto L(\psi, \omega, \boldsymbol{\beta}, d, k)g_1(\psi). \tag{12}$$

Using Eq. (9),

$$\pi(\psi|\omega, \boldsymbol{\beta}, d, k) \propto \prod_{i=1}^M \left\{ \prod_{j=1}^{n_i} [h_0(t_{ij})\eta_{ij}]^{\delta_{ij}} \right\} I_i g_1(\psi).$$

Hence, ignoring the terms that do not involve ψ we have,

$$\pi(\psi|\omega, \boldsymbol{\beta}, d, k) \propto \prod_{i=1}^M \left\{ \prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} \right\} I_i g_1(\psi).$$

Similarly

$$\begin{aligned} \pi(\omega|\psi, \boldsymbol{\beta}, d, k) &\propto \prod_{i=1}^M \left\{ \prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} \right\} I_i g_2(\omega), \\ \pi(d|\psi, \omega, \boldsymbol{\beta}, k) &\propto \prod_{i=1}^M I_i g_3(d), \\ \pi(k|\psi, \omega, \boldsymbol{\beta}, d) &\propto \prod_{i=1}^M I_i g_4(k), \\ \pi(\beta_i|\psi, \omega, \boldsymbol{\beta}_i, d, k) &\propto \prod_{i=1}^M \left\{ \prod_{j=1}^{n_i} \eta_{ij}^{\delta_{ij}} \right\} I_i p_i(\beta_i) \quad \text{for } i = 1, 2, \dots, r. \end{aligned} \tag{13}$$

where $\boldsymbol{\beta}_i = (\beta_1, \beta_2, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_r)$ represents a vector of regression coefficients except β_i .

Using the Random Walk Metropolis–Hastings Algorithm, a sample is drawn from each of the populations with the conditional posterior distributions given by (13). The algorithm was first described by [Metropolis and Ulam \(1949\)](#) and [Metropolis et al. \(1953\)](#) and later generalized by [Hastings \(1970\)](#). The algorithm can be summarized as follows:

1. Start with an initial set of values for the parameters $\boldsymbol{\theta}^{(0)} = (\psi^{(0)}, \omega^{(0)}, \boldsymbol{\beta}^{(0)}, d^{(0)}, k^{(0)})$.
2. Set iteration counter at $i = 1$.
3. Generate a proposal value θ^* from the normal transition kernel. For example, ψ^* is generated from $N(\psi^{(i-1)}, c)$.
4. Calculate the acceptance ratio, say for ψ^* we have $\alpha = \min \left\{ 1, \frac{\pi(\psi^*|\omega, \boldsymbol{\beta}, d, k)}{\pi(\psi^{(i-1)}|\omega, \boldsymbol{\beta}, d, k)} \right\}$

5. Generate a random number r from $U(0,1)$.
6. Accept ψ^* as $\psi^{(i)}$ if $r \leq \alpha$ else $\psi^{(i)} = \psi^{(i-1)}$
7. Repeat steps 3–6 for remaining parameters ω , β , d and k each time updating $\theta^{(i)}$ with the previous point accepted in the chain. These steps are repeated N times to get N samples from $\pi(\theta)$.

The standard deviation of the transition kernel ‘ c ’ is chosen independently for each parameter (after a few trial runs) such that the chain has an acceptance ratio of 25 % (Gelman et al. 1996) i.e. 25% of the new proposed values are accepted and others are rejected.

In implementations of this type, it is common to run the chain for an initial period of time (the burn-in period) and to store every l th iteration ($l > 1$) thereafter until some pre-specified stopping time. This is referred to as thinning the chain and it reduces dependence among the sample points from the posterior distribution. The remaining iterations are used to obtain the posterior summaries. We use the squared error loss function for which the Bayes estimator is the posterior mean. Convergence and stationarity of the Markov chain is studied using the Geweke (Geweke 1992) and the Heidelberger–Welch tests (Heidelberger and Welch 1983). More information about convergence tests in MCMC algorithms can be found in Cowles and Carlin (1996). We also study the Inefficiency Factor (proposed by Kim et al. 1998) which is the ratio of the variance of the sample mean of the correlated draws to the variance of the sample mean of hypothetical i.i.d. draws, to give us a measure of the factor by which the number of iterations needs to be increased (Kohn et al. 2001; Chib and Kang 2016).

Our main interest lies in finding an estimate of variance of the frailty distribution and this is done by using the estimates of the parameters d and k at each iteration of the chain. The estimates of variance of the frailty distribution form another chain, which is subjected to testing for convergence and stationarity.

6 Simulation study

The survival times are simulated from the model given in (1). To observe the effect of cluster size, sample size, and frailty variance on the performance of the model and the estimation procedure, nine datasets consisting of 25, 50 and 100 clusters, each with cluster sizes 2, 4 and 8 have been considered. Clusters of equal size have been chosen for computational convenience although the model can be used for unequal cluster sizes as well.

The survival times for the same cluster are considered to be correlated, having the same frailty described by z_i . In set-up I, the z_i 's follow log-normal distribution with mean 1 and variance 0.5. Set-up II and III assume gamma distribution for z_i with mean 1 and variance 2 and 4 respectively. Weibull distribution with both shape and scale parameters equal to 2 is chosen to be the baseline distribution. Two treatment effects are considered i.e. the number of covariates r equals 2 with $X = (X_1, X_2)'$ generated from $N_2(0, I_2)$ and $\beta = (1.5, 1.25)'$. For censoring the survival times, it is assumed that the censoring variable follows $U(0,a)$ where a is chosen so as to obtain a censoring of around 10%.

Table 1 Percentage of the chains converging as per the Geweke Test

M	n_i	Set-up I				Set-up II				Set-up III			
		ψ	β_1	β_2	σ_{FR}^2	ψ	β_1	β_2	σ_{FR}^2	ψ	β_1	β_2	σ_{FR}^2
25	2	89.4	90.8	91.2	89.6	89.2	92.6	92.0	85.2	90.8	90.6	92.0	71.2
	4	91.8	92.2	92.4	91.0	92.2	90.6	91.6	78.0	89.8	92.0	90.6	77.6
	8	89.8	89.6	90.8	90.4	91.6	90.6	91.4	79.6	90.6	91.8	91.6	78.4
50	2	90.8	89.6	89.6	87.2	89.0	91.6	89.2	77.2	88.2	88.0	89.4	75.2
	4	91.2	90.4	91.8	89.4	91.0	91.4	93.0	76.4	89.2	90.2	91.6	72.4
	8	91.4	93.4	90.4	88.8	93.8	90.8	89.6	76.4	91.4	88.8	91.2	80.0
100	2	89.6	91.6	88.8	89.2	83.8	85.4	88.4	74.2	85.0	89.6	89.2	69.8
	4	90.6	90.8	91.6	90.4	92.2	89.8	87.8	75.4	89.6	91.8	91.2	74.8
	8	90.6	89.2	90.4	91.4	89.6	92.4	92.6	76.8	91.6	91.4	89.8	79.8

500 datasets are generated under the different scenarios with $M = 100, 50, 25$ and $n = 2, 4, 8$. These datasets are then fit to the generalized gamma shared frailty model with Weibull baseline distribution using the estimation method described in Sect. 5. The algorithm described in the previous section is used to estimate the parameters. The integral (8) is computed using 32 point Gauss-Laguerre quadrature rule. The computations are performed using the R statistical software.

The convergence of the method is first monitored graphically using cumulative mean plots. Considering that the state of the chain for the first few iterations is dependent on the starting point, these plots are used to decide the number of iterations that need to be discarded as burn-in.

The trace plots of the chains are studied to determine whether the chain is exploring the parametric space well.

Auto Correlation Function (ACF) plots study the autocorrelation in the data. Since adjacent points in the Markov chain are correlated, it is expected that the autocorrelation will decrease as the lag increases. After discarding the first few iterations as burn-in, every l th value is chosen to arrive at a sample of draws from the chain.

Since it was not possible to see the plots individually for all 500 datasets under each scenario, the burn-in and lag were fixed after studying 50 plots under each setting. Geweke and Heidelberger–Welch (H–W) tests are used to study the convergence and stationarity of the chains. Tables 1 and 2 report the percentage of the datasets for which the chains converged (p -value of the Geweke test statistic > 0.05) and stationarity was achieved (p -value of H–W test statistic > 0.05) and Table 3 reports inefficiency factor for each parameter. These tests help us to decide a suitable length of the Markov chain, that is, the number of iterations required. This number varies for different models and although the same model is used for estimation in our simulation study, the data sizes and frailty variances differ. Hence, we use a standard length of 45,000 iterations with a burn-in of 10,000 and a lag of 150 for all estimations (though these numbers can be changed for quicker estimation in less complex models).

The resulting samples are then used to obtain posterior summaries. The squared error loss function is the mean square error (MSE) in our case since the posterior mean

Table 2 Percentage of the chains achieving stationarity as per the H–W Test

M	n_i	Set-up I				Set-up II				Set-up III			
		ψ	β_1	β_2	σ_{FR}^2	ψ	β_1	β_2	σ_{FR}^2	ψ	β_1	β_2	σ_{FR}^2
25	2	99.0	98.0	98.2	97.8	98.6	99.2	98.0	97.4	98.2	98.8	98.6	95.6
	4	99.8	98.8	99.0	98.6	99.0	98.6	98.6	96.2	99.2	98.2	99.8	97.4
	8	98.8	99.0	99.6	98.2	99.0	98.0	99.2	97.2	99.8	98.6	98.8	93.6
50	2	98.8	99.6	99.0	98.4	98.2	98.6	98.8	97.0	98.2	98.8	98.0	93.6
	4	98.8	99.2	99.4	98.4	99.0	98.8	98.6	96.8	99.2	99.0	100.0	95.0
	8	98.8	99.4	98.8	98.6	98.8	99.4	99.0	96.0	98.8	99.4	99.4	96.4
100	2	98.6	98.8	99.0	98.2	96.8	98.6	98.2	95.8	96.4	98.8	98.4	95.2
	4	98.8	98.4	98.2	98.8	98.4	99.2	98.6	96.8	97.8	98.8	98.2	96.0
	8	99.4	98.8	99.0	99.0	99.2	99.4	98.2	94.2	98.6	98.6	99.2	96.2

Table 3 Inefficiency factors for the Markov chains for each parameter

M	n_i	Set-up I				Set-up II				Set-up III			
		ψ	β_1	β_2	σ_{FR}^2	ψ	β_1	β_2	σ_{FR}^2	ψ	β_1	β_2	σ_{FR}^2
25	2	52.8	40.8	36.9	54.3	51.5	37.2	33.5	80.8	46.1	33.7	28.4	85.4
	4	37.6	31.2	28.1	34.9	34.8	26.1	23.0	71.3	31.8	25.1	22.0	88.9
	8	29.5	24.0	21.6	28.1	28.9	24.4	21.7	72.3	27.5	22.6	20.1	96.5
50	2	58.2	44.4	39.7	61.2	44.4	31.6	28.0	86.3	42.6	30.1	25.7	95.3
	4	33.9	27.0	24.1	46.4	30.7	24.7	21.9	96.6	27.3	21.9	19.7	96.4
	8	27.9	22.8	20.6	34.4	29.2	23.9	21.3	96.2	27.4	22.9	20.4	99.2
100	2	52.8	40.0	35.7	62.7	44.1	30.7	27.1	101.0	40.4	27.3	23.6	105.4
	4	35.5	28.1	25.1	49.2	29.1	23.3	20.8	101.1	29.0	23.4	20.9	116.5
	8	27.6	22.3	19.9	31.7	26.2	22.3	19.9	94.9	27.5	23.5	20.9	107.9

is the Bayes estimator. The MSEs, biases and Monte Carlo Standard Errors (MCSE) are displayed in Tables 4.

On the basis of Tables 1, 2, 3 and 4, it can be concluded that

1. as the cluster size or the number of clusters increases, the percentage of converging or stationary chains is not affected significantly.
2. as the frailty variance of the generated datasets increases, the convergence and stationarity of the Markov chains for the baseline shape parameter ψ and treatment effects β do not get affected significantly. However, for the estimate of frailty variance, the percentage of chains converging and achieving stationarity reduces.
3. decreasing inefficiency factors as cluster size increases indicates that more iterations are required when the cluster size is small. Also, to estimate frailty variance with better precision, more iterations would be required if the frailty variance is high.

Table 4 Biases and MSEs obtained from thinned sample and corresponding Monte Carlo Standard Errors for each parameter

M	n_i	ψ				β_1				β_2				$\sigma_{\tilde{F}_R}^2$			
		Bias	MSE	MCSE	Bias	MSE	MCSE	Bias	MSE	MCSE	Bias	MSE	MCSE	Bias	MSE	MCSE	
Set-up I																	
25	2	0.303	0.219	0.014	0.243	0.270	0.012	0.185	0.179	0.010	0.524	1.888	0.043				
	4	0.048	0.056	0.007	0.045	0.057	0.006	0.033	0.043	0.005	-0.081	0.123	0.013				
	8	0.014	0.022	0.004	0.014	0.022	0.004	0.012	0.017	0.003	-0.103	0.052	0.007				
50	2	0.065	0.082	0.010	0.058	0.072	0.008	0.030	0.052	0.007	0.009	0.168	0.020				
	4	0.015	0.025	0.005	0.018	0.027	0.004	0.013	0.020	0.003	-0.140	0.058	0.008				
	8	0.011	0.011	0.003	0.010	0.010	0.002	0.008	0.009	0.002	-0.118	0.036	0.005				
100	2	0.009	0.035	0.007	0.010	0.031	0.005	0.003	0.023	0.005	-0.054	0.084	0.012				
	4	-0.002	0.013	0.003	-0.003	0.012	0.003	0.003	0.010	0.003	-0.161	0.046	0.005				
	8	0.005	0.005	0.002	0.001	0.261	0.002	0.004	0.068	0.001	-0.123	0.027	0.003				
Set-up II																	
25	2	0.175	0.219	0.017	0.139	0.205	0.014	0.121	0.188	0.012	2.713	23.263	0.338				
	4	0.065	0.056	0.007	0.050	0.051	0.006	0.041	0.049	0.006	1.757	10.158	0.191				
	8	0.038	0.022	0.004	0.033	0.021	0.004	0.023	0.017	0.003	1.336	5.975	0.148				
50	2	0.046	0.082	0.010	0.043	0.088	0.009	0.035	0.069	0.007	1.761	9.840	0.193				
	4	0.026	0.025	0.005	0.027	0.026	0.004	0.019	0.023	0.004	1.199	5.224	0.132				
	8	0.017	0.011	0.003	0.014	0.010	0.003	0.008	0.009	0.002	0.972	3.680	0.107				
100	2	-0.012	0.035	0.007	-0.018	0.035	0.006	-0.004	0.027	0.005	1.286	4.976	0.134				
	4	0.010	0.013	0.003	0.005	0.012	0.003	0.008	0.010	0.003	0.764	2.381	0.087				
	8	0.003	0.005	0.002	0.002	0.005	0.002	0.001	0.004	0.002	0.496	1.357	0.063				

Table 4 continued

M	n_i	ψ			β_1			β_2			σ_{FR}^2			
		Bias	MSE	MCSE	Bias	MSE	MCSE	Bias	MSE	MCSE	Bias	MSE	MCSE	
Set-up III														
25	2	0.096	0.172	0.016	0.081	0.184	0.014	0.033	0.171	0.012	0.012	0.012	0.012	0.843
	4	0.061	0.059	0.007	0.047	0.058	0.006	0.030	0.052	0.006	0.006	0.006	0.006	0.679
	8	0.041	0.026	0.004	0.025	0.024	0.004	0.024	0.022	0.003	0.003	0.003	0.003	0.620
50	2	-0.048	0.069	0.010	-0.042	0.074	0.009	-0.049	0.074	0.007	0.007	0.007	0.007	0.594
	4	0.008	0.029	0.004	0.002	0.027	0.004	0.001	0.024	0.003	0.003	0.003	0.003	0.450
	8	0.016	0.012	0.003	0.009	0.011	0.003	0.007	0.010	0.002	0.002	0.002	0.002	0.438
100	2	-0.092	0.038	0.007	-0.086	0.040	0.006	-0.073	0.038	0.005	0.005	0.005	0.005	0.451
	4	-0.012	0.013	0.003	-0.015	0.014	0.003	-0.009	0.011	0.003	0.003	0.003	0.003	0.395
	8	0.001	0.006	0.002	-0.004	0.005	0.002	0.000	0.005	0.002	0.002	0.002	0.002	0.292

4. since MCSEs reduce with increasing cluster size, the efficiency of the algorithm to simulate draws from each posterior distribution improves as cluster size increases.
5. in general, the bias and mean squared error (MSE) reduces as the number of clusters increases or the cluster size increases. This is due to the increase in the number of data points in the study. Hence, larger studies would give better results.
6. the method is sensitive to the frailty variance. As the variance of the frailty term increases, the MSEs also increase. This indicates that increasing heterogeneity in event times within a cluster can affect the estimates of frailty variance. However, estimates of covariates are unaffected with heterogeneity in the clusters.

Since, the cost of data collection is an important factor that decides the sample size, the researchers should look at the trade-off between improving results or reducing costs of data collection. Also, in case the model indicates high heterogeneity in the data (high frailty variance), it may be worthwhile to look for covariates that have not been considered in the model or consider multiple populations with vastly different hazard rates.

7 Real life data study

Danahy et al. (1977) studied the effect of high dose oral isosorbide dinitrate (ISDN) on heart rate, blood pressure and exercise time until angina pectoris on 21 patients with coronary atherosclerotic heart disease. The patients were studied for 5 days in the hospital where they were given exercise tests lasting until the onset of angina pectoris. The exercise times were recorded for each patient after sublingual nitroglycerin (SLTNG) and sublingual placebo (SLP), at 1, 3 and 5 hours after oral placebo and its control period (OP1, OP2, OP3 and OPC) and at 1, 3 and 5 hours after oral ISDN and its control period (OI1, OI3, OI5 and OIC) as well.

The data set has 10 exercise times for each of the 21 patients with some observations considered to be censored if the patient was too exhausted to complete the test. The data can be analysed using a shared frailty model considering that the observations from a single cluster are coming from the same patient and are hence correlated. The generalized gamma frailty model as described in the previous sections is applied to the exercise data set with baseline distribution as Weibull distribution.

The nine treatment effects along with the baseline and frailty distribution parameters are to be estimated. The algorithm is run for fewer iterations (10,000 to 20,000) to set the standard deviations of the transition kernels in order to have an acceptance rate of approximately 25% for each parameter. Subsequently, the algorithm is run twice for 1,20,000 iterations using two distinct set of initial values.

The trace plots in Fig. 2 indicate whether the chain is exploring the parametric space well.

Observing the burn-in period of the chains from the Cumulative Mean (CM) plots given in Fig. 3, the initial 20,000 iterations for each parameter are discarded. The CM plot also shows the two chains obtained as a result of using two different starting points. This indicates that the estimates are independent of the initial values given to the algorithm. Use of distinct starting values can also check for multiple modes of the likelihood function.

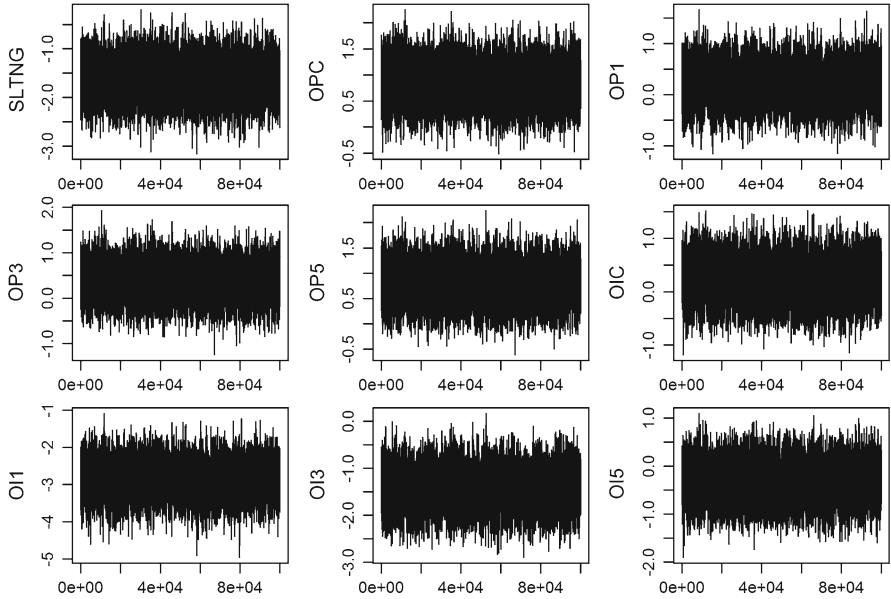


Fig. 2 Trace plots of Markov chains for treatment effects

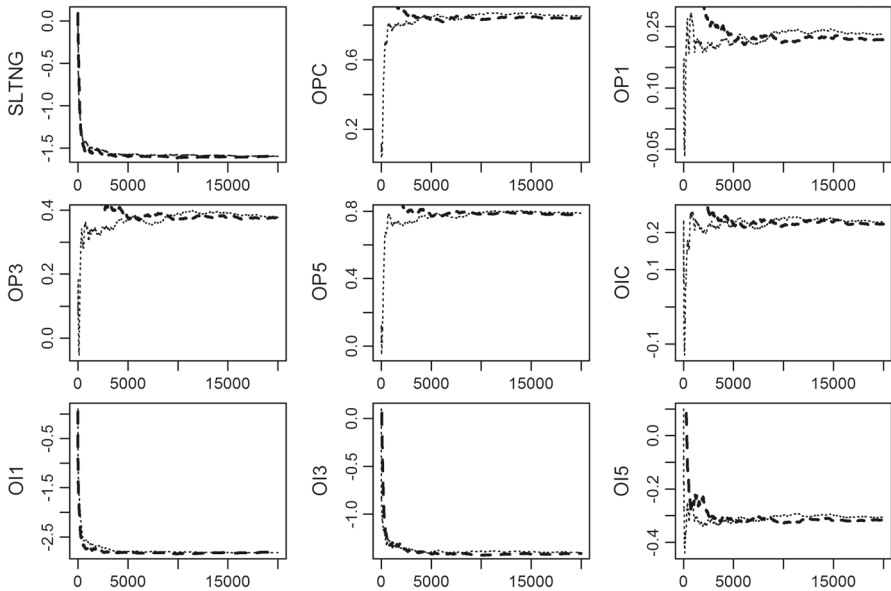


Fig. 3 Cumulative mean plots of two chains for the initial 20,000 iterations

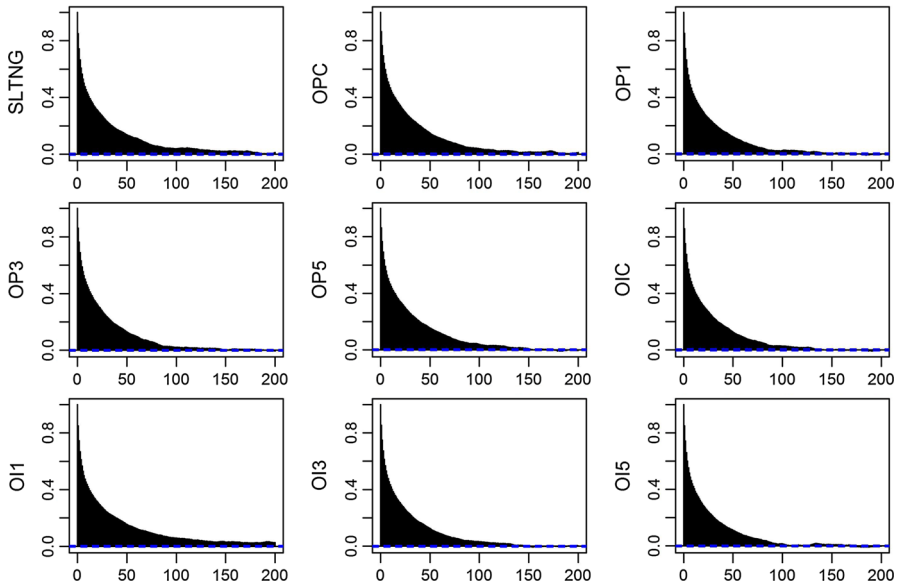


Fig. 4 The ACF plots

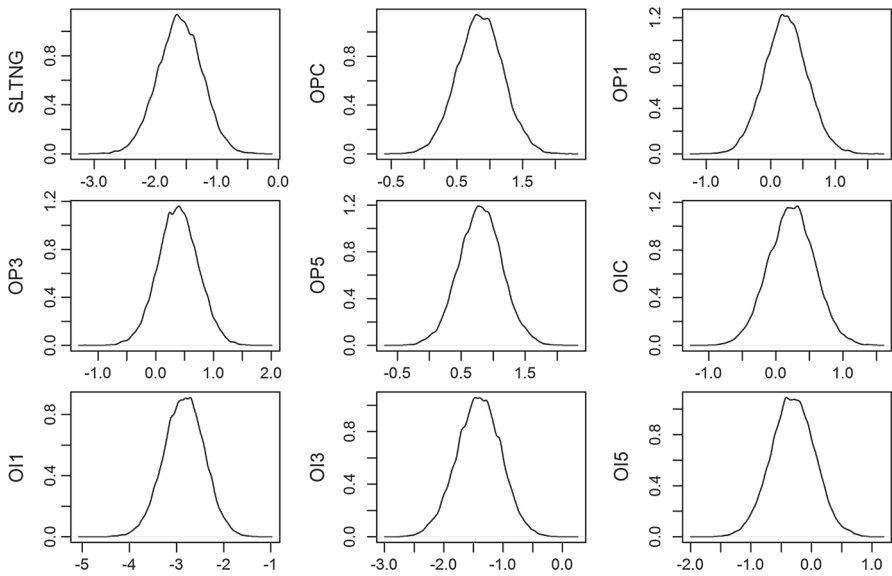


Fig. 5 The density plots for treatment effects

The ACF plots shown in Fig. 4 indicate a sufficient reduction in auto-correlation (<0.1) at a lag of 150.

The density plots of the posterior sample (Fig. 5) obtained for the treatment effects indicate a symmetric sample. Hence, we take the mean of the thinned sample to obtain an estimate for each treatment effect.

Table 5 Estimated treatment effects when compared with SLP

Treatment	$\hat{\beta}$	Geweke test	H–W test	Credible intervals	
		<i>p</i> -value	<i>p</i> -value	Lower	Upper
SLTNG	−1.5820	0.44	0.91	−2.3443	−0.7842
OPC	0.8496	0.78	0.90	0.1814	1.5170
OP1	0.2351	0.99	0.89	−0.4282	0.9559
OP3	0.4007	0.53	0.52	−0.2893	1.1280
OP5	0.7926	0.60	0.50	0.1263	1.5351
OIC	0.2302	0.69	0.57	−0.4177	0.8584
OI1	−2.8340	0.84	0.66	−3.7322	−1.9999
OI3	−1.4120	0.67	0.54	−2.1882	−0.7419
OI5	−0.3117	0.87	0.54	−1.0295	0.4274
ψ	5.397	0.77	0.94	4.750	6.085
Frailty variance	9.045	0.90	0.28	2.6725	25.7461

Table 5 gives the results based on the estimation procedure. The values in Table 5 lead us to the following conclusions.

1. Treatments SLTNG and ISDN (OI1, OI3, OI5) are beneficial to the patients as they reduce the risk of angina pectoris ($\hat{\beta} < 0$).
2. Use of an oral placebo increases the risk of heart disease as the corresponding $\hat{\beta}$ is positive.
3. Treatment effects SLTNG, OI1, and OI3 are significant as the credible intervals do not contain zero. Hence, there is a significant reduction in the risk of angina pectoris with the use of SLTNG and ISDN with the most benefit at one hour and three hours after the ISDN treatment.
4. Credible intervals of treatment effects OPC and OP5 also do not contain zero indicating a significant increase in risk of heart disease on using an oral placebo specially at the time of treatment and five hours after treatment.
5. The estimate of the shape parameter of the baseline distribution is greater than one, indicating an increasing hazard function with respect to time.
6. The frailty variance is estimated to be around 9.045 which indicates heterogeneity among clusters.

8 Conclusions

Bayesian estimation procedure is applied to the generalized gamma shared frailty model. The Random Walk Metropolis–Hastings algorithm (a Markov Chain Monte Carlo method) has been used to estimate the parameters. Each simulated data set of 100 clusters of size 8 each took approximately 20 minutes on a Dell Vostro 360, Intel i5-2400S processor with 4GB RAM.

In simulation study, it is observed that the Bayesian method performs well with generalized gamma frailty distribution with Weibull baseline distribution. However,

the model is sensitive to the variance of the frailty distribution. As the heterogeneity in the data increases, the squared error loss also increases. The estimation method is also applied to the exercise data set given by Danahy et al. (1977) and Bayesian estimates of the treatment effects are obtained. Convergence and stationarity of the Monte Carlo Markov Chains is studied using graphical methods viz. cumulative mean plots and trace plots, and through diagnostic tests such as Geweke and Heidelberger–Welch tests. The credible intervals obtained for each treatment effect are used to identify significant treatment effects.

In this model, one of the basic assumptions is that all treatment effects are independent. However, this may not be true in real life data. One possible way to handle multicollinearity problems is to check for correlation between the covariates prior to analysis. To this end, use of factor analysis method to arrive at a list of the more important independent covariates, may be explored as a part of future research. Use of other baseline distributions will also be explored. There is scope for improving computation times involved in calculations for which the performance of block updation of MCMC algorithm will be looked into.

Acknowledgements The first author is grateful to University Grants Commission, Govt. of India for providing financial support for carrying out this work. The authors are also thankful to Department of Science and Technology (DST), Govt. of India for providing support under PURSE grant and are grateful to the referees for their constructive suggestions which have helped immensely in improving the quality of the paper.

References

- Aalen O (1992) Modelling heterogeneity in survival analysis by the compound poisson distribution. *Ann Appl Probab* 2:951–972
- Allenby GM, Rossi PE (1998) Marketing models of consumer heterogeneity. *J Econom* 89(12):57–78. doi:[10.1016/S0304-4076\(98\)00055-4](https://doi.org/10.1016/S0304-4076(98)00055-4)
- Balakrishnan N, Peng Y (2006) Generalized gamma frailty model. *Stat Med* 25(16):2797–2816. doi:[10.1002/sim.2375](https://doi.org/10.1002/sim.2375)
- Bodnar O, Link A, Arendack B, Possolo A, Elster C (2017) Bayesian estimation in random effects meta-analysis using a non-informative prior. *Stat Med* 36(2):378–399. doi:[10.1002/sim.7156](https://doi.org/10.1002/sim.7156), [sim.7156](https://doi.org/10.1002/sim.7156)
- Callegaro A, Iacobelli I (2012) The cox shared frailty model with log-skew-normal frailties. *Stat Model* 12(5):399–418
- Chib S, Kang KH (2016) Efficient posterior sampling in Gaussian affine term structure models. <http://apps.olin.wustl.edu/faculty/chib/chibkang2016.pdf>. Accessed 14 Feb 2017
- Clayton D (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65(1):141–151. doi:[10.1093/biomet/65.1.141](https://doi.org/10.1093/biomet/65.1.141)
- Clayton DG (1991) A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* 47(2):467–485. doi:[10.2307/2532139](https://doi.org/10.2307/2532139)
- Cowles MK, Carlin BP (1996) Markov chain monte carlo convergence diagnostics: a comparative review. *J Am Stat Assoc* 91(434):883–904
- Cox D (1972) Regression models and life-tables. *J Roy Stat Soc Ser B Methodol* 34(2):187–220
- Danahy DT, Burwell DT, Aronow WS, Prakash R (1977) Sustained hemodynamic and antianginal effect of high dose oral isosorbide dinitrate. *Circulation* 55(2):381–387
- Dellaportas P, Smith AF (1993) Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *J Roy Stat Soc Ser C (Appl Stat)* 42(3): 443–459. doi:[10.2307/2986324](https://doi.org/10.2307/2986324)
- Ducrocq V, Casella G (1996) A bayesian analysis of mixed survival models. *Genet Sel Evol* 28(6):505. doi:[10.1186/1297-9686-28-6-505](https://doi.org/10.1186/1297-9686-28-6-505)

- Gelman A, Roberts GO, Gilks WR (1996) Efficient metropolis jumping rules. In: Bayesian statistics, vol 5 (Alicante, 1994). Oxford University Press, pp 599–607
- Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D (2013) Bayesian data analysis, 3rd edn. Chapman and Hall, New York
- Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Stat* 4:169–188
- Golightly A, Wilkinson DJ (2008) Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput Stat Data Anal* 52(3):1674–1693
- Greenwood M, Yule G (1920) An enquiry into the nature of frequency distributions representative of multiple happenings with particular reference of multiple attacks of disease or of repeated accidents. *J Roy Stat Soc* 83:255–279
- Hanagal DD, Dabade AD (2013) Bayesian estimation of parameters and comparison of shared gamma frailty models. *Commun Stat Simul Comput* 42(4):910–931
- Hastings W (1970) Monte carlo sampling methods using markov chains and their applications. *Biometrika* 57:97–109. doi:[10.1093/biomet/57.1.97](https://doi.org/10.1093/biomet/57.1.97)
- Heidelberger P, Welch PD (1983) Simulation run length control in the presence of an initial transient. *Oper Res* 31(6):1109–1144
- Higgins JPT, Thompson SG, Spiegelhalter DJ (2009) A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 172(1):137–159. doi:[10.1111/j.1467-985X.2008.00552.x](https://doi.org/10.1111/j.1467-985X.2008.00552.x)
- Hougaard P (1986a) A class of multivariate failure time distributions. *Biometrika* 73:671–678
- Hougaard P (1986b) Survival models for heterogeneous populations derived from stable distributions. *Biometrika* 73(2):387–396. doi:[10.1093/biomet/73.2.387](https://doi.org/10.1093/biomet/73.2.387)
- Hougaard P (1995) Frailty models for survival data. *Lifetime Data Anal* 1(3):255–273. doi:[10.1007/BF00985760](https://doi.org/10.1007/BF00985760)
- Hougaard P (2000) Analysis of multivariate survival data. Springer, New York. doi:[10.1007/978-1-4612-1304-8](https://doi.org/10.1007/978-1-4612-1304-8)
- Ibrahim JG, Chen MH, Sinha D (2001) Bayesian survival analysis. Springer, Berlin
- Khodabin M, Ahmadabadi A (2010) Some properties of generalized gamma distribution. *Math Sci* 4(1):9–28
- Khoshravesh M, Sefidkouhi MAG, Valipour M (2015) Estimation of reference evapotranspiration using multivariate fractional polynomial, Bayesian regression, and robust regression models in three arid environments. *Appl Water Sci*. doi:[10.1007/s13201-015-0368-x](https://doi.org/10.1007/s13201-015-0368-x)
- Kim S, Shephard N, Chib S (1998) Stochastic volatility: likelihood inference and comparison with arch models. *Rev Econ Stud* 65(3):361–393
- Klein JP (1992) Semiparametric estimation of random effects using the cox model based on the EM algorithm. *Biometrics* 48(3):795–806
- Kohn R, Smith M, Chan D (2001) Nonparametric regression using linear combinations of basis functions. *Stat Comput* 11(4):313–322
- Koop G, Osiewalski J, Steel MF (1997) Bayesian efficiency analysis through individual effects: hospital cost frontiers. *J Econom* 76(12):77–105. doi:[10.1016/0304-4076\(95\)01783-6](https://doi.org/10.1016/0304-4076(95)01783-6)
- Lam KF, Kuk AY (1997) A marginal likelihood approach to estimation in frailty models. *J Am Stat Assoc* 92(439): 985–990
- McGilchrist CA, Aisbett CW (1991) Regression with frailty in survival analysis. *Biometrics* 47(2):461–466
- Metropolis N, Ulam S (1949) The Monte Carlo method. *J Am Stat Assoc* 44(247):335–341. doi:[10.2307/2280232](https://doi.org/10.2307/2280232)
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092
- Oakes D (1994) Use of frailty models for multivariate survival data. In: Proceedings of the XVIIth international biometrics conference, Hamilton, Ontario, Canada, pp 275–286
- Pengcheng C, Jiajia Z, Riquan Z (2013) Estimation of the accelerated failure time frailty model under generalized gamma frailty. *Comput Stat Data Anal* 62:171–180
- Pullenayegum EM (2011) An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes. *Stat Med* 30(26):3082–3094. doi:[10.1002/sim.4326](https://doi.org/10.1002/sim.4326)
- Ripatti S, Palmgren J (2000) Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* 56(4):1016–1022
- Sahu SK, Dey DK, Aslanidou H, Sinha D (1997) A weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Anal* 3:123–137

- Santos CA, Achcar JA (2010) A Bayesian analysis for multivariate survival data in the presence of covariates. *J Stat Theory Appl* 9:233–253
- Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JPT (2015) Predictive distributions for between-study heterogeneity and simple methods for their application in bayesian meta-analysis. *Stat Med* 34(6):984–998. doi:[10.1002/sim.6381](https://doi.org/10.1002/sim.6381)
- Vaida F, Xu R (2000) Proportional hazards model with random effects. *Stat Med* 19:3309–3324
- Vaupel J, Manton K, Stallard E (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3):439–454. doi:[10.2307/2061224](https://doi.org/10.2307/2061224)
- Wienke A, Lichtenstein P, Yashin AI (2003) A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. *Biometrics* 59(4):1178–1183. doi:[10.1111/j.0006-341x.2003.00135.x](https://doi.org/10.1111/j.0006-341x.2003.00135.x)
- Xue X, Brookmeyer R (1996) Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Anal* 2(3):277–289. doi:[10.1007/bf00128978](https://doi.org/10.1007/bf00128978)
- Yashin A, Vaupel J, Iachine I (1995) Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Math Popul Stud* 5(2):145–159. doi:[10.1080/08898489509525394](https://doi.org/10.1080/08898489509525394)
- Yau K, McGilchrist C (1997) Use of generalised linear mixed models for the analysis of clustered survival data. *Biom J* 39:3–11
- Zeger SL, Karim MR (1991) Generalized linear models with random effects: a gibbs sampling approach. *J Am Stat Assoc* 86(413):79–86. doi:[10.1080/01621459.1991.10475006](https://doi.org/10.1080/01621459.1991.10475006)