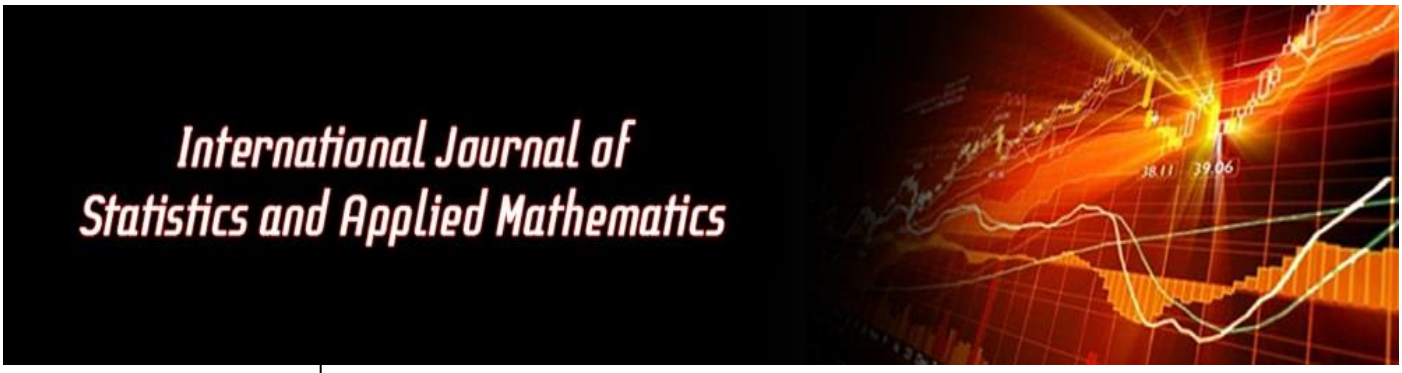
International Journal of Statistics and Applied Mathematics 2018; 3(2): 511-517



ISSN: 2456-1452
Maths 2018; 3(2): 511-517

www.mathsjournal.com
Received: 15-01-2018
Accepted: 16-02-2018

Arvind Pandey
Department of Statistics, Central University of Rajasthan, Ajmer, Rajasthan, India

Lalpawimawha
Department of Statistics, Pachhunga University Collage, Mizoram, India

Correspondence
Lalpawimawha
Department of Statistics, Pachhunga University Collge, Mizoram, India

Comparison of shared frailty models: A Bayesian approach

Arvind Pandey and Lalpawimawha

Abstract
In this article, we propose gamma and inverse Gaussian frailty shared models with Akash distribution as baseline to analyze the bivariate survival data set of Mc Gilchrist and Aisbett (1991). Bayesian approach of Markov Chain Monte Carlo technique was employed to estimate the parameters involves in the models. The better model also suggested for the data.

Keywords: Bayesian comparison, gamma frailty, inverse gaussian frailty, MCMC, Shared frailty

1. Introduction
In survival analysis Cox proportional hazard model by Cox (1972) [2] is the most commonly used method. It is assumed that the study population is homogeneous. But In many applications, consider populace cannot be expected to be homogeneous but must be considered as a heterogeneous sample, i.e. a blend of people with different hazards. For illustration, in many cases it is possible to include all the covariates related to the illness of interest due to economical reasons or sometimes the significance of a few covariates is still obscure. The missing or unobserved covariate is termed as "frailty" which account for heterogeneity in the population. The notion of frailty provides a convenient way to introduce random effects, dependence and unobserved heterogeneity into models for survival data. The term frailty itself was introduced by Vaupel *et al*. (1979) [14] in univariate survival models and the model was substantially promoted by its application to multivariate survival data in a seminal paper by Clayton (1978) [1] (without using the notion "frailty") on chronic disease incidence in families.
The frailty approach could be a statistical modeling concept which points to account for heterogeneity, caused by unmeasured covariates. In statistical terms, a frailty model may be a random effect model for time-to-event data, where the random effect (the slightness) features a multiplicative impact on the baseline hazard function. Hanagal (2007) [3] proposed gamma frailty regression models in mixture distribution. Hangal and Shama (2013) [7] also suggested gamma shared frailty for modeling heterogeneity in bivariate survival data. Hanagal and Pandey (2017) [6] considered inverse Gaussian distribution as frailty parameter and compared the models by using kidney infection data.
In this manuscript, we compare shared gamma frailty model, shared inverse Gaussian frailty model and the model without frailty i.e. Cox proportional hazard model by using Bayesian method of comparison.

2. General shared frailty model
For the shared frailty model, it is assumed that survival times are conditionally independent, for given shared frailty. That means, dependence between survival times is only due to unobservable covariates or frailty. When there is no variability in the distribution of frailty variable it has a degenerate distribution and when the distribution of it is not degenerate the dependence is positive.
Suppose there are n clusters and that cluster i has ni observations and associates with the unobserved frailty M_i ($1 \leq i \leq n$). The vector X_{ij} ($1 \leq i \leq n$, $1 \leq j \leq n_i$) contains the covariate information of the event time T_{ij} of the j^{th} observation in the i^{th} cluster. Conditional on the frailty term M_i, the survival times in cluster i ($1 \leq i \leq n$) are assumed to be independent and

their hazard functions and survival functions to be of the form

$$h(t_{ij} \mid X_{ij}, M_i) = M_i h_0(t_{ij}) e^{\beta' X_{ij}} \text{ and } S(t_{ij} \mid m_j, X_j) = e^{-m_j H_0(t) e^{\beta' X_{ij}}} \quad (2.1)$$

where $h_0(t_{ij})$ denotes the baseline hazard function, $H_0(t)$ is cumulative baseline hazard function at time t > 0,and β is a vector of regression coefficients. The frailties M_i (i = 1... n) are assumed to be independently and identically distributed random variables with density function *f (m).* The frailty density depends on unknown parameters to be estimated. The main assumption of a shared frailty model is that all individuals in cluster i share the same value of frailty M_i (i = 1 . . . n), and this is why the model is called the shared frailty model.

Under the assumption of conditional independence, bivariate conditional survival function for given frailty $M_j=m_j$ at times t_{1j} >0 and t_{2j} >0 is,

$$S(t_{1j,}t_{2j} \mid m_j, X_j) = S(t_{1j} \mid m_j, X_j) S(t_{2j} \mid m_j, X_{ij}) = e^{-m_j (H_{01}(t_{1j}) + H_{02}(t_{2j})) e^{\beta' X_{ij}}} \quad (2.2)$$

Unconditional bivariate survival function at times t_{1j} >0 and t_{2j} >0 can be obtained by integrating over frailty variable M_j having the probability function f (m_j), for j^{th} individual.

$$S(t_{1j}, t_{2j} \mid X_{ij}) = L_{Z_j}[(H_{01}(t_{1j}) + H_{02}(t_{2j})) e^{\beta' X_{ij}}] \quad (2.3)$$

Where $L_{M_j}(.)$ is Laplace transform of distribution of frailty variable M_j for j^{th} individual. Here onwards we represent S (t_{1j}, t_{2j} | X_{ij}) as S (t_{1j}, t_{2j}).

Now we consider frailty distributions one by one. First, gamma distribution is considered because the gamma distribution fits well to failure data from a computational and analytical point of view. It is easy to derive the closed form expression of survival and hazard function. For identifiability, we assume M has expected value equal to one. Under this restriction, Laplace transform of a gamma distribution is given by,

$$L_M(s) = (1 + \xi s)^{\frac{-1}{\xi}} \quad (2.4)$$

With variance of M is ξ.

Replacing Laplace transform in equation (2.4), we get the unconditional bivariate survival function for j^{th} individual at times t_{1j} >0 and t_{2j} >0 as,

$$S(t_{1j}, t_{2j}) = [1 + \xi((H_{01}(t_{1j}) + H_{02}(t_{2j})) e^{\beta' X_{ij}}]^{\frac{-1}{\xi}} \quad (2.5)$$

The gamma distribution is most commonly used frailty distribution because of its mathematical convenience. However, it has drawbacks (see Kheiri *et al*. 2007) [10] for example, it may weaken the effect of covariates. Alternative to the gamma distribution Hougaard (1984) [8] introduced inverse Gaussian as a frailty distribution. It provides much flexibility in modeling.

Under the identifiability condition, Laplace transform of inverse Gaussian distribution is,

$$L_M(s) = \exp\left[\frac{1 - (1 + 2\xi s)^{\frac{1}{2}}}{\xi}\right] \quad (2.6)$$

With variance of M is ξ.

Replacing Laplace transform in equation (2.6), the unconditional bivariate survival
Function for j th individual at times t_{1j} >0 and t_{2j} >0 is,

$$S(t_{1j}, t_{2j}) = \exp\left\{\frac{1 - [1 + 2\xi(H_{01}(t_{1j}) + H_{02}(t_{2j})) e^{\beta' X_{ij}}]^{\frac{1}{2}}}{\xi}\right\} \quad (2.7)$$

The bivariate survival function in the case when the frailty variable is degenerate is given by

$$S(t_{1j},t_{2j}) = e^{-[e^{\beta' X_{ij}}(H_{01}(t_{1j})+H_{02}(t_{2j}))]} \tag{2.8}$$

3. Baseline distribution

The baseline distribution used here is Akash Distribution proposed by Shanker (2015) [13], which is the modification of Lindley distribution (Lindley, 1958) [11]. Akash distribution is able to fit in data obtained from medical sciences and engineering.

A continuous random variable T is said to follow the Akash distribution with the parameter α if its survival function is,

$$S(t) = \left[1+\frac{\alpha t(\alpha t+2)}{\alpha^2+2}\right]e^{\alpha t};\quad t>0,\alpha>0 \tag{3.1}$$

And the hazard function and the cumulative hazard function as

$$h(t) = \frac{\alpha^2(1+t)}{(\alpha+1)+\alpha t}$$

$$H(t) = \alpha t - \log\left[1+\frac{\alpha t(\alpha t+2)}{\alpha^2+2}\right];\quad t>0,\alpha>0 \tag{3.2}$$

The hazard function is increasing and more flexible than Lindley distribution and exponential distribution. That is why we choose as baseline distribution.

4. Proposed Models

Substituting the cumulative hazard function in equations (2.5), (2.7) and (2.8), we get the unconditional survival function of bivariate random variable (T_{1j}, T_{2j}), then

$$S(t_{1j},t_{2j}) = \left[1+\xi\left(\alpha_1 t_{1j} - \log\left(1+\frac{\alpha_1 t_{1j}(\alpha_1 t_{1j}+2)}{\alpha_1{}^2+2}\right)+\left(\alpha_2 t_{2j} - \log\left(1+\frac{\alpha_2 t_{2j}(\alpha_2 t_{2j}+2)}{\alpha_2{}^2+2}\right)\right)\right)e^{\beta' X_{ij}}\right]^{\frac{-1}{\xi}} \tag{4.1}$$

$$S(t_{1j},t_{2j}) = \exp\left\{\frac{1-\left[1+2\xi\left(\alpha_1 t_{1j} - \log\left(1+\frac{\alpha_1 t_{1j}(\alpha_1 t_{1j}+2)}{\alpha_1{}^2+2}\right)+\left(\alpha_2 t_{2j} - \log\left(1+\frac{\alpha_2 t_{2j}(\alpha_2 t_{2j}+2)}{\alpha_2{}^2+2}\right)\right)\right)e^{\beta' X_{ij}}\right]^{\frac{1}{2}}}{\xi}\right\} \tag{4.2}$$

$$S(t_{1j},t_{2j}) = \exp\left\{-\left(\alpha_1 t_{1j} - \log\left(1+\frac{\alpha_1 t_{1j}(\alpha_1 t_{1j}+2)}{\alpha_1{}^2+2}\right)+\left(\alpha_2 t_{2j} - \log\left(1+\frac{\alpha_2 t_{2j}(\alpha_2 t_{2j}+2)}{\alpha_2{}^2+2}\right)\right)\right)e^{\beta' X_{ij}}\right\} \tag{4.3}$$

Equations (4.1) and (4.2) are gamma shared frailty and inverse Gaussian shared frailty models with baseline Akash distribution and called as model-I and model-II, equation (4.3) is without frailty with baseline Akash distribution and called as model-III.

5. Methodology

The likelihood function associated with the failures times and censoring variables ε_{ij} (i =1,2; j = 1,2,…n) based on the survival function is given as

$$L(\underline{\varphi},\underline{\beta},\xi) = \left(\prod_{j=1}^{n_1} f_1(t_{1j},t_{2j})\right)\left(\prod_{j=1}^{n_2} f_2(t_{1j},\varepsilon_{2j})\right)\left(\prod_{j=1}^{n_3} f_3(\varepsilon_{1j},t_{2j})\right)\left(\prod_{j=1}^{n_4} f_4(\varepsilon_{1j},\varepsilon_{2j})\right) \tag{5.1}$$

Where $\underline{\varphi}$, $\underline{\beta}$ and ξ are the vector of baseline parameter, the vector of regression coefficients, and the frailty parameter. For without frailty model likelihood function is

$$L(\underline{\varphi},\underline{\beta})=\left(\prod_{j=1}^{n_1} f_1(t_{1j},t_{2j})\right)\left(\prod_{j=1}^{n_2} f_2(t_{1j},\varepsilon_{2j})\right)\left(\prod_{j=1}^{n_3} f_3(\varepsilon_{1j},t_{2j})\right)\left(\prod_{j=1}^{n_4} f_4(\varepsilon_{1j},\varepsilon_{2j})\right) \tag{5.2}$$

Where n_1, n_2, n_3, and n_4 are the number of observations observed to falls in the range $t_{1j} \le \varepsilon_j$, $t_{2j} \le \varepsilon_j$; $t_{1j} \le \varepsilon_j$, $t_{2j} > \varepsilon_j$; $t_{2j} > \varepsilon_j$, $t_{1j} \le \varepsilon_j$; and $t_{1j} > \varepsilon_j$ j, $t_{2j} > \varepsilon_j$, respectively and the contribution of the jth individual in the likelihood as

$$\begin{aligned} f_1(t_{1j},t_{2j}) &= \frac{\partial^2(t_{1j},t_{2j})}{\partial t_{1j}\partial t_{2j}} \\ f\,2(t_{1j},\varepsilon_{2j}) &= -\frac{\partial(t_{1j},\varepsilon_{2j})}{\partial t_{1j}} \\ f_3(\varepsilon_{1j},t_{2j}) &= \frac{\partial(\varepsilon_{1j},t_{2j})}{\partial t_{2j}} \\ and \quad f_4(\varepsilon_{1j},\varepsilon_{2j}) &= S(\varepsilon_{1j},\varepsilon_{2j}) \end{aligned} \tag{5.3}$$

Substituting distribution function S (t_{1j}, t_{2j}) and by differentiating we get the likelihood function given by equation (5.1). Similarly we get the likelihood function for without frailty model.

The likelihood function obtained in equation (5.1) is not easy to solve since it involves a large number of parameters in the model by using Newton-Raphson method, the MLEs is not converge. Thus, we move to the Bayesian approach, which does not suffer any kind of such difficulties.

The joint posterior density function of parameters for given failure times is obtained as,

$$\pi(\alpha_1,\alpha_2,\xi,\beta) \propto L(\alpha_1,\alpha_2,\xi,\beta)\times g_1(\alpha_1)g_2(\alpha_2)g_3(\xi)\prod_{i=1}^{5} p_i(\beta_i)$$

Where $g_i(.)$ (i = 1, 2, 3) indicates the prior density function with known hyper parameters of corresponding arguments for baseline parameters and frailty variance; $p_i(.)$ is prior density function for regression coefficient β_i and likelihood function L (.) is given by equation (5.1) or (5.2). Here we assume that all the parameters are independently distributed.

The non-informative prior was used for frailty parameters and regression coefficients. Since we do not have prior information on baseline parameters, it is assumed to be flat. The prior distribution for the parameters used is as follows-

For frailty parameter ($\underline{\xi}$) ~ Γ (0.0001, 0.0001)

For regression coefficients ($\underline{\varphi}$) ~ Normal (0, 1000)

For baseline parameters ($\underline{\beta}$) ~ Γ (1, 0.0001) and U (0, 100)

We have fitted the Bayesian model with the above prior density functions and likelihood function (5.1) using the MCMC methods such as, the Metropolis-Hastings algorithm. We have monitored convergence of Markov chain to a stationary distribution by Gelman-Rubin convergence statistic and Geweke test. Trace plots, coupling from the past plots and sample autocorrelation function plots have been used to check the behavior of the chain, to decide burn-in period and sample autocorrelation lag respectively.

In order to compare the proposed models we have used several Bayesian model selection criteria such as, Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC) and Deviance Information Criteria (DIC). Also we have used the Bayes factor B_{jk} for comparison of the modelsM_j against M_k. To compute the Bayes factor we have considered MCMC approach given in Kass and Raftery (1995) [9].

6. Data Analysis

Kidney infection data of McGrilchrist and Aisbett (1991) [12] was utilized demonstrate the Bayesian estimation technique. The data are comprises of infection recurrence times at the point of insertion of the catheter for 38 kidney patients using portable dialysis equipment. For each patient, the occurrences of infection for the first and second time were recorded. It may be removed other than infection and regarded as censoring. So the first or second infection time or censoring time may be the survival times. To cure the first infection sufficient time was permitted before the catheter was inserted for the second time. So the first and second recurrence times are taken to be independent apart from the common frailty component. The data of the risk variables are

age, sex and disease type GN, AN and PKD where GN, AN and PKD are brief forms of Glomerulo Neptiritis, Acute Neptiritis and Polycystic Kidney Disease.

Let T_1 and T_2 be the first and the second recurrence time to infection. Five covariates age, sex and presence or absence of disease type GN, AN and PKD are represented by X_1, X_2, X_3, X_4, and X_5.

First we check goodness of fit to the data for frailty models and then apply the Bayesian estimation procedure. Kolmogorov-Smirnov test was considered to check goodness of fit for the kidney infection data set test and the p-values also obtained for T_1 and T_2 separately. Corresponding p-values are displayed in Table 1.

Table 1: p-values of K-S Statistics for goodness of fit test for Kidney Infection data set

Model	Recurrences times	
	First	Second
Gamma frailty	0.9024	0.3720
Inverse Gaussian Frailty	0.4490	0.1229

The p-values are large enough to say there is no statistical evidence to reject the hypothesis that first and second recurrence times follow one of the distributions with survival functions given in Eqs. (4.1) and (4.2).

The posterior summary is displayed in Tables 2-4. In these tables second and third columns represent estimate (posterior mean) and standard error whereas last three columns represent Gelman-Rbin values, Geweke values and p-values.

Table 2: Posterior summary for model-I

Parameter	Estimate	SE	Lower Credible Limit	Upper Credible Limit	GR values	Geweke values	p-values
Burn-in period = 6700 ; autocorrelation lag = 125							
α_1	0.0711	0.0031	0.0648	0.0775	1.0000	-0.0022	0.4991
α_2	0.0519	0.0049	0.0417	0.0595	1.0001	-0.0031	0.4987
ξ	1.2647	0.0506	1.1774	1.3606	1.0037	0.0060	0.5024
β_1	0.0301	0.0051	0.0207	0.0391	1.0030	-0.0079	0.4968
β_2	-2.7443	0.3646	-3.4104	-1.9932	1.0015	0.0110	0.4968
β_3	0.4500	0.0520	0.3569	0.5434	1.0021	-0.0006	0.4997
β_4	0.6125	0.0520	0.5180	0.7020	1.0003	0.0116	0.5046
β_5	-0.3989	0.1191	-0.6002	-0.1667	1.0125	-0.0048	0.4980

Table 3: Posterior summary for model-II

Parameter	Estimate	SE	Lower Credible Limit	Upper Credible Limit	GR values	Geweke values	p-values
Burn-in period = 6500 ; autocorrelation lag = 170							
α_1	0.0904	0.0032	0.0841	0.0970	1.0006	0.0047	0.5019
α_2	0.0612	0.0053	0.0508	0.0697	1.0002	-0.0018	0.4992
ξ	1.9016	0.4891	1.0972	2.9179	1.0001	0.0009	0.5003
β_1	0.0197	0.0049	0.0108	0.0289	0.9999	0.0089	0.5035
β_2	-2.4555	0.3558	-3.1231	-1.7951	1.0004	-0.0061	0.5035
β_3	0.4473	0.0504	0.3570	0.5440	1.0001	0.0115	0.5045
β_4	0.6135	0.0527	0.5176	0.7012	1.0001	0.0008	0.5003
β_5	-1.4045	0.4975	-2.2576	-0.4637	1.0004	-0.0075	0.4969

Table 3: Posterior summary for model-III

Parameter	Estimate	SE	Lower Credible Limit	Upper Credible Limit	GR values	Geweke values	p-values
Burn-in period = 6200; autocorrelation lag = 240							
α_1	0.0623	0.0029	0.0561	0.0678	1.0009	0.0079	0.5031
α_2	0.0429	0.0040	0.0338	0.0494	1.0008	-0.0152	0.4939
β_1	0.0164	0.0039	0.0104	0.0250	1.0005	-0.0044	0.4982
β_2	-2.5814	0.2383	-3.0742	-2.1532	1.0002	0.0077	0.4982
β_3	0.0318	0.0050	0.0226	0.0411	1.0017	0.0094	0.5037

β_4	0.8290	0.0494	0.7399	0.9182	1.0040	0.0023	0.5009
β_5	-2.2982	0.3937	-3.1220	-1.5989	1.0000	0.0103	0.5041

From Tables 2-4, the value zero is not a credible value for all credible intervals for all the models. The negative value of regression coefficient β_2 indicates that female has lower risk of infection than male.

Table 5: AIC, BIC and DIC values for all models

Baseline distribution	Model	AIC	BIC	DIC	Log-likelihood
Akash Distribution	Gamma Frailty	689.0755	702.1761	676.6887	-336.5377
	Inverse Gaussian Frailty	682.4093	695.5100	671.0884	-333.2046
	Without Frailty	706.0754	717.5385	697.8843	-346.0377

The AIC, BIC and DIC values for model-I and model-II are comparatively much smaller than for model-III, so frailty models are better than without frailty model. The difference between AIC, BIC and DIC values for model-I and model-II is very small, so AIC, BIC and DIC values are not enough to take the decision between the model-I and model-II. Same thing hold for log likelihood. Now Bayes factor B_{jk} was considered for comparing the models j and k. $D_{jk} = 2 \ln (B_{jk})$ for the model-II against model-I is 5.4129; for model-II against model-III is 27.2213 and for model-I against model-III is 21.8083 given in Table (6).

Table 6: Bayes factor for all models

Numerator model against denominator model	$D_{jk}=2\log_e(B_{ij})$	range	Evidence against model in denominator
II against I	5.4129	≥ 2 and ≤ 6	Positive
II against III	27.2213	≥ 10	Very strong positive
I against III	21.8083	≥ 10	Very strong positive

Thus from all the model comparison criteria we can say model-II is better than model-I and model-III for modeling kidney infection data.

7. Results and Discussion

In the present paper, we proposed new model of gamma frailty and inverse Gaussian frailty with Akash distribution as baseline distribution.

We have used the Metropolis–Hastings algorithm to fit all the models. We analyzed kidney infection data using our proposed models and the best model is suggested. We have used self-written programs in R statistical software to perform analysis.

For all the models, the value zero was not a credible value for the all the credible intervals. So, all covariates are significant. Negative value of regression coefficient of X_2 indicating that female has lower risk of infection than male. All the model comparison criteria suggested that inverse Gaussian frailty model with Akash baseline distribution is better for modeling of kidney infection data than other models. The estimates of frailty variance are high in the frailty models which are 1.2647 and 1.9016 for gamma and inverse frailty models respectively. This indicates that there is a strong evidence of high degree of heterogeneity in the population of patients.

The proposed model suggested that the frailty variance in two proposed models are very high as compare to the earlier frailty model proposed in Mc Grilchrist and Aisbett (1991) [12] on log-normal frailty, Hanagal and Dabade (2012) [4] on compound Poisson frailty and Hanagal and Dabade (2013) [5] on gamma frailty models.

8. Conclusion

Shared gamma frailty model and shared inverse Gaussian frailty model are better than without frailty to analyze Kidney infection data.

9. References

1. Clayton DG. A model for association in bivariate life tables and its application in epidemologicalstudies of familial tendency in chronic diease incidence, Biometrika. 1978; 65:141-151.
2. Cox DR. Regression models and life tables (with discussion), Journal of the Royal Statistical Society B. 1972; 34:187-220.
3. Hanagal DD. Gamma frailty regression models in mixture distributions. Economic Quality Control. 2007; 22(2):295-302.
4. Hanagal DD, Dabade AD. Modelling heterogeneity in bivariate survival data by compound Poisson distribution using Bayesian approach. International Journal of Statistics and Management Systems. 2012; 7(1):36-84.
5. Hanagal DD, Dabade AD. Bayesian estimation of parameters and comparison of shared gamma frailty models. Communications in Statistics-Simulation and Computation. 2013; 42 (4):910-931.
6. Hanagal DD, Pandey A. Shared inverse Gaussian frailty models based on additive hazards. Communications in Statistics-Theory and Methods. 2017; 46(22):11143-11162.
7. Hanagal DD, Sharma R. Modeling heterogeneity for bivariate survival data by shared gamma frailty regression model. Model Assisted Statistics and Applications. 2013; 8(2):85-102.
8. Hougaard P. Life table methods for heterogeneous populations. Biometrika. 1984; 71(1):75-83.
9. Kass RE, Raftery AE, Bayes factor. Journal of the American Statistical Association. 1995; 90(430):773-795.
10. Kheiri S, Kimber A, Meshkani MR. Bayesian analysis of an inverse Gaussian correlated frailty model. Computational Statistics and Data Analysis. 2007; 51:5317-5326.

11. Lindley DV. Fiducial distributions and Bayes' theorem, Journal of the Royal Statistical Society, Series B. 1958; 20:102-107.
12. McGilchrist CA, Aisbett CW. Regression with frailty in survival analysis. Biometrics. 1991; 47:461-466.
13. Shanker R. Akash distribution and its applications. International journal of probability and Statistics. 2015; 4(3):67-75.
14. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailtyon the dynamics of mortality. Demography. 1979; 16:439-454.