STATISTICS IN TRANSITION new series, December 2022
Vol. 23, No. 4, pp. 161–176, DOI 10.2478/stattrans-2022-0048
Received – 22.02.2021; accepted – 10.01.2022

Generalised Lindley shared additive frailty regression model for bivariate survival data

Arvind Pandey [1], **David D. Hanagal** [2], **Shikhar Tyagi**[3]

ABSTRACT

Frailty models are the possible choice to counter the problem of the unobserved heterogeneity in individual risks of disease and death. Based on earlier studies, shared frailty models can be utilised in the analysis of bivariate data related to survival times (e.g. matched pairs experiments, twin or family data). In this article, we assume that frailty acts additively to the hazard rate. A new class of shared frailty models based on generalised Lindley distribution is established. By assuming generalised Weibull and generalised log-logistic baseline distributions, we propose a new class of shared frailty models based on the additive hazard rate. We estimate the parameters in these frailty models and use the Bayesian paradigm of the Markov Chain Monte Carlo (MCMC) technique. Model selection criteria have been applied for the comparison of models. We analyse kidney infection data and suggest the best model.

Key words: Bayesian estimation, frailty, generalised Lindley frailty, generalised log-logistic distribution, generalised Weibull distribution, hazard rate, MCMC, random censoring.

1. Introduction

To analyse the survival data in biological, epidemiological, and medical studies, a common approach is that subjects are supposed to have the same risk of occurrence of an event of interest, which acts multiplicatively. However, this assumption rarely occurs because neither all the covariates can be measured nor can be included in the study due to technical difficulties, time limitations, or financial implications. In real-life situations risk (hazard rate) changes from one family to another family, one group to another group, one cluster to another cluster. Heterogeneity in the population exists, because of the mixture of groups of individuals with different risk factors. This heterogeneity is called frailty. Ignoring frailty may have adverse consequences. A random impact that is an unobservable risk shared by the subject is characterized as frailty, which was introduced by Vaupel et al. (1979). To handle such kinds of problems, many models have been derived in survival analysis. Since the establishment of the proportional hazard model given by Cox (1972), the survival function has been dominated by hazard rate models. The reason behind the popularity of this model is the significance of known covariates that can be tested, also a relationship between lifetimes and covariates can be incorporated.

[1]Department of Statistics, Central University of Rajasthan, Rajasthan, India. E-mail: arvindmzu@gmail.com. ORCID: https://orcid.org/0000-0003-1324-6985.
[2]Department of Statistics, Savitribai Phule Pune University, Pune-411007, India. E-mail: david.hanagal@gmail.com. ORCID: https://orcid.org/0000-0002-8918-6934.
[3]Department of Statistics, Central University of Rajasthan, Rajasthan, India. E-mail: shikhar1093tyagi@gmail.com. ORCID: https://orcid.org/0000-0003-1606-0844.

Research on the bivariate survival models has grown rapidly over the past few years. Clayton's (1978) random effect model of the bivariate survival was a key innovation. He introduced the notion of the shared relative risk. This model was further developed by Oakes (1982) to analyze the association between two non-negative random variables. Hougaard (1985, 1991, 2000) discussed the different aspects of frailty on a broad scale. In the last decade, frailty regression models in mixture distribution were discussed by Hanagal (2008). Modelling kidney infection data for inverse Gaussian shared frailty was done by Hanagal and Pandey (2014a). Gamma frailty models for bivariate survival data were given by Hanagal and Pandey (2015a). Hanagal and Pandey (2017a) used the shared inverse Gaussian frailty models based on additive hazard. Hanagal (2019) gave an extensive literature review on different shared frailty models. Pandey et al. (2020a) presented shared inverse Gaussian frailty models for bivariate findings. Pandey et al. (2020b) looked at generalised inverse Gaussian shared frailty models based on reversed hazard rates. Pandey et al. (2021a, 2022) and Tyagi et al. (2021a) developed distinct Generalised Lindley (GL) shared frailty models based on the reversed hazard rate. Tyagi et al. (2021b, 2022a, 2022b), Gupta et al. (2022), Pandey et al. (2021b), and Pandey and Tyagi (2021) developed inverse weighted Lindley, and GL shared frailty models, respectively. In this article, we assume that frailty acts additively to the hazard rate. The additive hazard models characterize a different facet of the association between covariates and the failure time than the proportional hazard model and are more plausible than the latter for many applications (Lin and Ying, 1994; Bin, 2010). The additive hazard models can be authentically a better alternative to proportional hazard or other nonlinear hazard regression models to narrate the consequences of covariates on survival time (Hosmer and Royston, 2002). When the absolute change in risk, instead of the risk ratio, is of primary interest or when the proportional hazard assumption for the Cox proportional hazard model is violated, an additive hazard regression model may be more appropriate (Xie et al., 2013). Let a continuous random variable T be a lifetime of an individual and the random variable W be frailty variable. The conditional hazard function for a given frailty variable, $W = w$ at time $t \in \mathbb{R}^+$ is

$$\phi(t \mid w) = \phi_0(t) + e^{\underline{K}\underline{\beta}+\underline{V}\underline{\beta}_1} = \phi_0(t) + we^{\underline{K}\underline{\beta}}, w \in \mathbb{R}^+, V \in \mathbb{R}, \tag{1}$$

where $w = e^{\underline{V}\underline{\beta}_1}$ and $\phi_0(t)$ is a baseline hazard function at time $t \in \mathbb{R}^+$, $\underline{K}$ is a row vector of covariates, and $\underline{\beta}$ is a column vector of regression coefficients. The cumulative hazard rate function is given by

$$\Phi(t \mid z) \quad = \quad \Phi_0(t) + wte^{\underline{K}\underline{\beta}}. \tag{2}$$

The conditional survival function for given frailty at time $t \in \mathbb{R}^+$ is

$$S(t \mid w) \quad = \quad e^{-\int_0^t \phi(x|w)dx} = e^{-\left[\Phi_0(t)+wte^{\underline{K}\underline{\beta}}\right]}, \tag{3}$$

where $\Phi_0(t)$ is the cumulative baseline hazard function at time $t \in \mathbb{R}^+$. Integrating over the

range of frailty variable W having density $f(w)$, we get the marginal survival function as

$$S(t) = \int_{w\in\mathbb{R}^+} S(t \mid w) f(w) dw = \int_{w\in\mathbb{R}^+} e^{-\left[\Phi_0(t)+wte^{\underline{K}\underline{\beta}}\right]} f(w) dw, = S_0(t) L_w(te^{\underline{K}\underline{\beta}}), \quad (4)$$

where $L_Z(.)$ is the Laplace transformation of the distribution of Z and $S_0(t)$ is the baseline survival function of T. Once we get the survival function at time $t \in \mathbb{R}^+$, of lifetime random variable for an individual, we can obtain the probability structure and make its inferences based on it.

The main objective of this article is threefold. First, generalised Lindley (GL) shared frailty models for additive hazard rate with generalised Weibull and generalised log-logistic as baseline distributions have been introduced. Second, the Bayesian approach of estimation has been employed to estimate the unknown parameters under random censoring. Third, a simulation study and data analysis have been done for the Kidney infection data set.

2. General Shared Frailty Model

The shared frailty models are applicable to event time of the related individuals, similar organs, and repeated measurements. In this model individuals from a group shares common covariates. It has been considered that survival times are conditionally independent, for a given shared frailty. Shared frailty indicates dependence between survival times is only because of unobservable covariates (frailty). Frailty variable W has a degenerate distribution in the absence of variability. If the dependence is positive, the distribution of W is not degenerate.

Assume n individuals are considered under the study. Bivariate random variables (T_{1j}, T_{2j}) are postulated as the first and the second survival times of the j^{th} individual $(j = 1,2,3,\ldots,n)$. Also m known covariates are supposed to be collected in a vector $\underline{K}_j = (K_{1j},\ldots,K_{mj})$ for the j^{th} individual where K_{aj} $(a = 1,2,3,\ldots,m)$ represents the value of the a^{th} observed covariate for the j^{th} individual. Under shared frailty model, it has been presupposed that both survival times for everyone share the similar value of the covariates. Let W_j be shared frailty for the j^{th} individual. Assuming that the frailties are acting additively on the baseline hazard function and both the survival times of individuals are conditionally independent for given frailty, the conditional hazard function for the j^{th} individual at the i^{th} $(i = 1,2)$ survival time $t_{ij} \in \mathbb{R}^+$ for given frailty $W_j = w_j$ has the form

$$\phi(t_{ij} \mid W_j, \underline{K}_j) \quad = \quad \phi_0(t_{ij}) + w_j e^{\underline{K}_j\underline{\beta}},$$

where $\phi_0(t_{ij})$ is the baseline hazard at time $t_{ij} \in \mathbb{R}^+$ and $\underline{\beta}$ is a vector of order m, of the regression coefficients. The conditional cumulative hazard function for the j^{th} individual at the i^{th} survival time $t_{ij} \in \mathbb{R}^+$ for a given frailty $W_j = w_j$ is

$$\Phi(t_{ij} \mid w_j, \underline{X}_j) \quad = \quad \Phi_0(t_{ij}) + w_j t_{ij} \rho_j,$$

where $\rho_j = e^{\underline{K}_j\underline{\beta}}$ and $\Phi_0(t_{ij})$ is the cumulative baseline hazard function at time $t_{ij} \in \mathbb{R}^+$. The conditional survival function for the j^{th} individual at the i^{th} survival time $t_{ij} \in \mathbb{R}^+$ for

a given frailty $W_j = w_j$ is

$$S(t_{ij} \mid w_j, \underline{K}_j) = e^{-\Phi(t_{ij}|w_j,\underline{K}_j)} = e^{-\left[\Phi_0(t_{ij})+w_j t_{ij}\rho_j\right]}.$$

Under the assumption of independence, the bivariate conditional survival function for a given frailty, $W_j = w_j$ at time $t_{1j} \in \mathbb{R}^+$ and $t_{2j} \in \mathbb{R}^+$ is

$$S(t_{1j},t_{2j} \mid w_j, \underline{K}_j) = S(t_{1j} \mid w_j, \underline{K}_j) S(t_{2j} \mid w_j, \underline{K}_j) = e^{-\left[(\Phi_{01}(t_{1j})+\Phi_{02}(t_{2j}))+w_j(t_{1j}+t_{2j})\rho_j\right]}.$$

The unconditional bivariate survival function at time $t_{1j} \in \mathbb{R}^+$ and $t_{2j} \in \mathbb{R}^+$ can be obtained by integrating over the frailty variable W_j having the probability function $f_W(w_j)$, for the j^{th} individual

$$S(t_{1j},t_{2j} \mid \underline{K}_j) = \int_{W_j \in \mathbb{R}^+} S(t_{1j},t_{2j} \mid w_j) f_W(w_j) dw_j = e^{-(\Phi_{01}(t_{1j})+\Phi_{02}(t_{2j}))} L_{W_j}[(t_{1j}+t_{2j})\rho_j], \tag{5}$$

where $L_{Z_j}(.)$ is the Laplace transform of the frailty variable of W_j for the j^{th} individual. Here onwards we represent $S(t_{1j},t_{2j} \mid \underline{K}_j)$ as $S(t_{1j},t_{2j})$.

3. Generalised Lindley Frailty Model

Lindley (1958) proposed a distribution with one parameter. Because of having only one parameter, the Lindley distribution does not provide enough flexibility for modelling purposes. It will be useful to consider further alternatives of this distribution. For a frailty distribution, a new generalised Lindley distribution has been considered in this paper. This distribution is the mixture of two gamma distributions G(θ,μ) and G(θ,η) with mixing coefficient $\theta/(\theta+1)$ (Elbatal, et al. (2013)). Because of the mixture of two gamma densities, a slight suppleness can be seen during analysis of time to event data. That is the reason why the GL frailty model is more adaptable in comparison with the gamma frailty model. the probability density function of GL distribution has been specified below:

$$f_W(w) = \begin{cases} \frac{1}{(1+\theta)}\left[\frac{\theta^{\mu+1}w^{\mu-1}}{\Gamma\mu} + \frac{\theta^{\eta}w^{\eta-1}}{\Gamma\eta}\right]e^{-\theta w} & ;w \in \mathbb{R}^+, \mu,\eta,\theta \in \mathbb{R}^+ \\ 0 & ;otherwise, \end{cases}$$

with mean $E[W] = \frac{1}{1+\theta}\left[\mu + \frac{\eta}{\theta}\right]$. And corresponding variance is,

$$V(W) = \frac{1}{(1+\theta)}\left[\left(\mu^2 + \frac{\eta^2}{\theta}\right)\left(\frac{1}{\theta(1+\theta)}\right) + \left(\frac{\mu+\eta}{\theta}\right) - \left(\frac{2\mu\eta}{\theta(1+\theta)}\right)\right],$$

after applying identifiability property, i.e., $E[W] = 1$ we get a relation between parameters $\eta = \theta(1+\theta-\mu)$. Consequently, the density function, the Laplace transformation and variance for GL are reduced to

$$f_W(w) = \begin{cases} \frac{1}{(1+\theta)}\left[\frac{\theta^{\mu+1}w^{\mu-1}}{\Gamma\mu} + \frac{\theta^{\theta(1+\theta-\mu)}w^{\theta(1+\theta-\mu)-1}}{\Gamma\theta(1+\theta-\mu)}\right]e^{-\theta w} & ;\ w,\theta \in \mathbb{R}^+, \mu \in (0,1+\theta) \\ 0 & ;\ otherwise. \end{cases}$$

$$L_W(s) = \frac{1}{(1+\theta)}\left[\frac{\theta^{\mu+1}}{(s+\theta)^{\mu}} + \frac{\theta^{\theta(1+\theta-\mu)}}{(s+\theta)^{\theta(1+\theta-\mu)}}\right], \tag{6}$$

$$V(W) = \frac{\theta^4 - \theta^3\mu + 3\theta^2(1+\theta) - 4\theta^2\mu + 3\theta\mu(\mu-1) + \mu^2}{\theta(1+\theta)^2}. \tag{7}$$

Let n be the number of observations under study. Let (T_{1j}, T_{2j}) be the first and second survival times of pairs of components of j^{th} $(1,2,...,n)$ objects. The unconditional bivariate survival function at time $t_{1j} \in \mathbb{R}^+$ and $t_{2j} \in \mathbb{R}^+$ using equations (5) and (6) can be written as

$$S(t_{1j}, t_{2j}) = \frac{e^{-(\Phi_{01}(t_{1j})+\Phi_{02}(t_{2j}))}}{(1+\theta)}\left[\frac{\theta^{\mu+1}}{(\theta+\rho(t_{1j}+t_{2j}))^{\mu}} + \frac{\theta^{\theta(1+\theta-\mu)}}{(\theta+\rho(t_{1j}+t_{2j}))^{\theta(1+\theta-\mu)}}\right], \tag{8}$$

where $\Phi_{01}(t_{1j})$, $\Phi_{02}(t_{2j})$ are the cumulative baseline hazard rate functions of the lifetime T_{1j} and T_{2j}, respectively. One can have different baseline distributions for T_1 and T_2. After substituting different cumulative hazard functions in (8), we get different generalised Lindley frailty distributions.

4. Dependence Measure

Sometimes due to complex form of frailty models, it is difficult to compare the degree of dependence between different frailty models. Kendall's τ can be used to quantify dependence because it is independent of transformations on the time scale and the frailty model used. It is a rank-based dependence measure.

$$\tau = \int_{s\in\mathbb{R}^+} 4sL_W^{"}(s)L_W(s)ds - 1. \tag{9}$$

After using equation (8) and (9), we get,

$$\tau = \int_{s\in\mathbb{R}^+} R(s \mid \theta,\mu)ds - 1, \tag{10}$$

where $R(s \mid \theta,\mu) = \frac{4\theta s\left(\theta^{\mu+1}A^{-\mu}+\theta^{\theta B}A^{\theta(\mu-\theta-1)}\right)\left(\mu(\mu+1)\theta^{\mu}A^{-\mu}+\theta^{\theta B}B\left(-\mu\theta+\theta^2+\theta+1\right)A^{\theta(\mu-\theta-1)}\right)}{(\theta+1)^2A^2}$.
$A = (\theta+s)$, $B = (1+\theta-\mu)$.
Kendall's τ cannot be found in closed form for GL frailty. Some numerical approaches can be utilized to obtain Kendall's τ dependence measure.

5. Baseline Distributions

5.1. Generalised Weibull Distribution

Here, the generalised Weibull distribution has been postulated as a baseline distribution. If a continuous random variable $T \in \mathbb{R}^+$ follows the generalised Weibull distribution then the survival and cumulative hazard function, are respectively,

$$S(t) = \begin{cases} 1-\left(1-e^{-\delta t^{\xi}}\right)^{\zeta} & ;t \in \mathbb{R}^+, \delta, \zeta, \xi \in \mathbb{R}^+ \\ 1 & ;otherwise \end{cases} \tag{11}$$

$$\Phi_0(t) = \begin{cases} -\log\left(1-\left(1-e^{-\delta t^{\xi}}\right)^{\zeta}\right) & ;t \in \mathbb{R}^+, \delta, \zeta, \xi \in \mathbb{R}^+ \\ 0 & ;otherwise \end{cases} \tag{12}$$

5.2. Generalised log-logistic distribution

Bacon (1993) used the log-logistic distribution for modelling saturation effects. The survival function of the log-logistic distribution is given by,

$$S(t) = (1+\delta t^{\xi})^{-1} \tag{13}$$

Due to having heavier tail in camparison to the gamma distribution, the log-logistic distribution can be more beneficial to be used for finance and insurance variables. The log-logistic distribution provides two parametric models for the survival analysis. Unlike the more commonly used Weibull distribution, it can have a non-monotonic hazard function: when $\xi > 1$ the hazard function is unimodal (when $\xi \leq 1$, the hazard decreases monotonically). The fact that the cumulative distribution function can be written in the closed form is particularly useful for the analysis of the survival data with censoring.

Lehmann family (Deshpande and Purohit, 2005) is a very useful family of life distributions generated from a given survival function and extensively used to model the effect of covariates. Let $S_0(t)$ be an arbitrary known survival function. If ζ is positive, then

$$S(t) = (S_0(t))^{\zeta} \tag{14}$$

is also a survival function. If, in particular, ζ is the positive integer n, then it represents the survival function of $min(X_1, ..., X_n)$ where X_i's are i.i.d. random variables with $S_0(t)$ as the common survival function. The hazards are proportional ζ times. Lehmann family is also known as the proportional hazards family. We use the same property and obtain, the survival function and the cumulative hazard rate as follows.

$$S(t) = \begin{cases} (1+\delta t^{\xi})^{-\zeta} & ;t \in \mathbb{R}^+, \delta, \zeta, \xi \in \mathbb{R}^+ \\ 1 & ;otherwise \end{cases} \tag{15}$$

$$\Phi_0(t) = \begin{cases} \zeta \log(1+\delta t^{\xi}) & ;t \in \mathbb{R}^+, \delta, \zeta, \xi \in \mathbb{R}^+ \\ 0 & ;otherwise \end{cases} \quad (16)$$

6. Proposed model

Due to group variation or frailty and individual variation described by the hazard function, a shared frailty model can be considered as a mixture model in survival analysis. Distribution of W is convergent, if dependence is positive. After substituting a cumulative hazard function for generalised Weibull and generalised log-logistic baseline distributions in equation (8)

$$S(t_{1j},t_{2j}) = \frac{1}{(1+\theta)}\left[\frac{\theta^{\mu+1}}{(\theta+\rho(t_{1j}+t_{2j}))^{\mu}} + \frac{\theta^{\theta(1+\theta-\mu)}}{(\theta+\rho(t_{1j}+t_{2j}))^{\theta(1+\theta-\mu)}}\right] \prod_{i=1}^{2}\left(1-\left(1-e^{-\delta_i t_{ij}^{\xi_i}}\right)^{\zeta_i}\right), \quad (17)$$

$$S(t_{1j},t_{2j}) = \frac{1}{(1+\theta)}\left[\frac{\theta^{\mu+1}}{(\theta+\rho(t_{1j}+t_{2j}))^{\mu}} + \frac{\theta^{\theta(1+\theta-\mu)}}{(\theta+\rho(t_{1j}+t_{2j}))^{\theta(1+\theta-\mu)}}\right] \prod_{i=1}^{2}(1+\delta_i t_{ij}^{\xi_i})^{-\zeta_i}, \quad (18)$$

here, equations (17), (18) can be called Model-I, Model-II respectively, which have been established for generalised Weibull and generalised log-logistic baseline distributions.

7. Likelihood Design and Bayesian Paradigm

For the study, n individuals have been considered. Observed failure times have been indicated by (t_{1j},t_{2j}). We are using the random censoring scheme. Let censoring time is indicated by c_{1j} and c_{2j} for j^{th} individual $(j = 1,2,3,...,n)$. Independence between censoring schemes and lifetimes of individuals has been presumed. The probability density function can be described for bivariate lifetime random variable of the j^{th} individual as

$$f_j(t_{1j},t_{2j}) = \begin{cases} f_1(t_{1j},t_{2j}), & ; t_{1j} < c_{1j}, t_{2j} < c_{2j}, \\ f_2(t_{1j},c_{2j}), & ; t_{1j} < c_{1j}, t_{2j} > c_{2j}, \\ f_3(c_{1j},t_{2j}), & ; t_{1j} > c_{1j}, t_{2j} < c_{2j}, \\ f_4(c_{1j},c_{2j}), & ; t_{1j} > c_{1j}, t_{2j} > c_{2j}. \end{cases}$$

The likelihood function will be

$$L(\underline{\Theta},\underline{\beta},\theta,\mu) = \prod_{j=1}^{n_1} f_1(t_{1j},t_{2j}) \prod_{j=1}^{n_2} f_2(t_{1j},c_{2j}) \prod_{j=1}^{n_3} f_3(c_{1j},t_{2j}) \prod_{j=1}^{n_4} f_4(c_{1j},c_{2j}), \quad (19)$$

where $\underline{\Theta}$, $\underline{\beta}$, θ and μ are the vector of baseline parameters and the vector of regression coefficients and frailty parameters respectively. Let n_1, n_2, n_3, and n_4 be the number of pairs for which the first and the second failure times (t_{1j}, t_{2j}) lie in the ranges $t_{1j} < c_{1j}, t_{2j} < c_{2j}$; $t_{1j} < c_{1j}, t_{2j} > c_{2j}$; $t_{1j} > c_{1j}, t_{2j} < c_{2j}$, and $t_{1j} > c_{1j}, t_{2j} > c_{2j}$ respectively and let

$$f_1(t_{1j}, t_{2j}) = \frac{\partial^2 S(t_{1j}, t_{2j})}{\partial t_{1j} \partial t_{2j}}; f_2(t_{1j}, c_{2j}) = -\frac{\partial S(t_{1j}, c_{2j})}{\partial t_{1j}},$$
$$f_3(c_{1j}, t_{2j}) = -\frac{\partial S(c_{1j}, t_{2j})}{\partial t_{2j}}; f_4(c_{1j}, c_{2j}) = S(c_{1j}, c_{2j}). \tag{20}$$

Substituting cumulative hazard rates $\Phi_{01}(t_{1j})$ and $\Phi_{02}(t_{2j})$ and survival function $S(t_{1j}, t_{2j})$ in equation (29) for Model-I and Model-II and by differentiating we get the likelihood function. The maximum likelihood method has crucial importance in computing efficient estimators. Inappropriately, due to a convergence problem, maximum likelihood failed to estimate the parameters, because Model-I and Model-II have thirteen-dimensional optimization problems. The Bayesian scenario has been discussed by several researchers for estimating parameters of the frailty models. For gamma and log-normal frailty models, the Bayesian paradigm has been contemplated by Santos and Achcar (2010). Weibull and piecewise exponential models have been discussed by Ibrahim et al. (2001) with gamma frailty. The joint posterior density function of parameters for given failure times is obtained as

$$\pi(\Theta, \theta, \mu, \underline{\beta_0}) \propto L(\Theta, \mu, \underline{\beta_0}) g_1(\zeta) g_2(\xi) g_3(\delta) g_4(\theta) g_5(\mu) \prod_{i=1}^{5} p_i(\beta_{0i\times 1})$$

where $g_i(.)$ indicates the prior density function with known hyperparameters of the corresponding argument for baseline parameters and frailty variance; $p_i(.)$ is prior density function for regression coefficient β_{0i} and the likelihood function is $L(.)$. An important assumption here is that all the parameters are independently distributed. In a similar way, the joint posterior density function can be written for without frailty models. To estimate the parameters of the models, hybrid Metropolis-Hastings algorithms have been used. The Geweke test (see Geweke, 1992) and Gelman-Rubin (see Gelman and Rubin, 1992) statistics have been used to monitor the convergence of a Markov chain to a stationary distribution.
Due to the high dimensions of conditional distributions, it is difficult to integrate out. Thus, it has been considered that full conditional distributions can be obtained as they are proportional to the joint distribution of the parameter of the model. The conditional distribution for single parameter δ with frailty is obtained as

$$\psi_1(\delta \mid \xi, \zeta, \theta, \mu, \underline{\beta_0}) \quad \propto \quad L(\delta, \xi, \zeta, \theta, \mu, \underline{\beta_0}) \cdot g_1(\delta) \tag{21}$$

the conditional distribution for single parameter δ without frailty is obtained as

$$\psi_1(\delta \mid \xi, \zeta, \underline{\beta_0}) \quad \propto \quad L(\delta, \xi, \zeta, \underline{\beta_0}) \cdot g_1(\delta).$$

Similarly, the conditional distributions for other parameters can be obtained.

8. Simulation Study

A simulation study has been executed to appraise the Bayesian estimation paradigm for Model-I and Model-II. Single covariate K_1 has been considered folliwng normal distribution. The frailty variable W is assumed to follow generalised Lindley distribution. Independence between lifetimes of individuals has been considered. Samples are generated using the subsequent mechanism,

1. From the binomial distribution with probability 0.6, 25 values for K_1 has been generated.

2. For known covariates, compute $\rho = e^{K_1\beta_1}$.

3. The distribution of lifetimes follow generalised Weibull and generalised log-logistic baseline distributions for given frailty W_j. 25 values of lifetimes have been generated. The conditional survival function for lifetime t_j $(j = 1,2,...,n)$ for given frailty $W_j = w_j$ and covariate K_1 is

$$S(t_j \mid w_j, K_1) \quad = \quad e^{-(\Phi_0(t_j)+w_j t\rho)}$$

   Equating $S(t_j \mid w_j, K_1)$ to random number, say $v_j (0 < v_j < 1)$ generated from $U(0,1)$ over $t_j \in \mathbb{R}^+$ we get:
   for Model-I and Model-II

$$v_j \quad = \quad \left(1 - \left(1 - e^{-\delta t_j^{\xi}}\right)^{\zeta}\right) * e^{-wt_j\rho},$$

$$v_j \quad = \quad (1 + \delta t_j^{\xi})^{-\zeta} * e^{-wt_j\rho} \quad \text{respectively.}$$

4. Censoring time c_j has been generated from $G(0.9, 0.01)$ for Model-I.

5. Observe the j^{th} survival time $t_j^* = min(t_j, c_j)$ and the censoring indicator χ_j for the j^{th} individual $(j = 1,2,...,25)$ where

$$\chi_j = \begin{cases} 1, & ; t_j < c_j \\ 0, & ; t_j > c_j \end{cases}$$

   thus we have data consisting of 25 pairs of survival times t_j^* and the censoring indicator χ_j.

Concurrently, with different priors and starting points, two chains have been operated. Both chains were recapitulated 100,000 times. Gelman-Rubin test values are very close to one. Due to small values of Geweke test statistic and corresponding p-values, the chains reach stationary distribution for both prior sets. The estimates of parameters were the same for both the priors, no impact of prior distributions has been found on posterior summaries. Here, the analysis for one chain has been exhibited because both the chains have shown the

same results. Tables 1 and 2 present the estimates and the credible intervals of the parameters for Models I and II based on the simulation study. The Gelman-Rubin convergence statistic values are nearly equal to one and also the Geweke test values are quite small, and the corresponding p-values are large enough to say that the chain attains stationary distribution.

9. Applicability on Kidney Infection Data

To elucidate the Bayesian estimation paradigm, kidney infection data of McGilchrist and Aisbett (1991) have been considered. This data consists of 38 patients and recurrence times (in days) of infection are given. Table 3 gives the p-values of goodness of fit test for Model I and Model II. Consequently, on the basis of p-values of K-S test it is clear that there is no statistical evidence to reject the hypothesis that data are from Model I and Model II in the marginal case and it can be assumed that they also fit for bivariate case. For frailty parameters, gamma distribution with very small shape and scale parameters (say, 0.0001) has been used. Additionally, it can be considered that regression coefficients are normally distributed with mean zero and high variance (say 1000). A similar type of prior was used in Ibrahim et al. (2001) and Santos and Achcar (2010). Thus for frailty parameters θ, μ and regression coefficients $\beta_{0i}, i = 1, ..., 5$, vague priors have been used. Because of no information about the baseline parameter, the prior distribution corresponding to baseline parameters is also considered flat. We considered two different vague prior distributions for baseline parameters, one is gamma distribution with shape and scale hyperparameters $\varepsilon_1, \varepsilon_2$ respectively and another is uniform distribution with interval (ν_1, ν_2). All the hyperparameters are known. Under the Bayesian paradigm, for both models, two parallel chains have been run. Also, two sets of prior distributions have been used with different starting points using the hybrid Metropolis-Hastings algorithm based on normal transition kernels. It can be said that estimates are independent of the different prior distributions because, for both sets of priors, estimates of parameters are approximately similar. We got an almost similar convergence rate of the Gibbs sampler for both sets of priors. Here, the analysis for one chain has been exhibited because both the chains have shown the same results. The Gelman-Rubin convergence statistic values are closely equal to one. The Geweke test statistic values are somewhat small, and the corresponding p-values are large enough to say that the chains reach stationary distribution. Tables 4-5 contained the values of posterior mean and the standard error with 95% credible intervals, the Gelman-Rubin statistics values, and the Geweke test with p-values for Model I and II. The AIC, BIC, and DIC values, given in Table 7, have been used to compare both models. Model-I holds the lowest possible values of AIC, BIC, and DIC. For Model-I and Model-II, the credible interval of all regression coefficients does not contain zero. It indicates that all covariates have a significant effect on both models. With a negative value, it is being indicated that age, sex, disease PKD are significant factors for kidney infection, having negative effects. But, with positive value diseases, GN and AN have a positive significant effect with a higher chance of infection. It is observed that female patients have a lower risk of kidney infection as compared to male counterparts.

10. Conclusions

A generalised Lindley additive frailty model under generalised Weibull and generalised log-logistic baseline distributions has been proposed. To fit the proposed model the hybrid M-H algorithm has been applied. Analysis has been done in R statistical software with self-written programs. The value of both frailty parameters for Model-I ($\theta = 2.29258, \mu = 1.38391$) and Model-II ($\theta = 2.12060, \mu = 1.28878$) is very high and corresponding variances are 1.434811 and 1.36565 by using equation (3.2). In Table 6, we calculate Kendall's τ measure of dependence by using equation (4.2). All these values are large enough to exhibit that there is a strong indication of heterogeneity among the patients in the population for the data set. To take the decision about all models, different tools have been utilized. With the lowest value of AIC, BIC, and DIC, from Table 7, and the value of Bayes factor for Model-I against Model-II (1.122368), it can be said that Model-I is better than the Model-II to analyze kidney infection data. For kidney infection data, all the covariates have been found statistically significant factors for both models (see Tables 4-5). Our proposed frailty models, Model-I and Model-II, are better as compared to the frailty models by Hanagal et al. (2017) and Hanagal and Pandey (2017a) with baseline generalised log-logistic distribution. In a similar way, with a minimum value of AIC, our proposed frailty models are better as compared to the frailty models by Pandey et al. (2018).

References

Bacon, R. W., (1993). A note on the use of the log-logistic functional form for modeling saturation effects. *Oxford Bulletin of Economics and Statistics*, 55, pp. 355–361.

Bin, H., (2010). Additive hazards model with time-varying regression coefficients. *Acta Math. Sci.*, 30B(4), pp. 1318–1326.

Clayton's, D. G., (1978). A model for association in bivariate life tables and its applications to epidemiological studies of familial tendency in chronic disease incidence. *Biometrica*, 65, pp. 141–151.

Cox, D. R., (1972). Regression Models and Life Tables (with Discussion). *Journal of Royal Statistical Society*, Series B, 34, pp. 187–220.

Deshpande, J. V., Purohit, S. G., (2005). Life Time Data: Statistical Models and Methods. *World Scientific*, New Jersey.

Elbatal, I., Merovci, F., & Elgarhy, M., (2013). A new generalized Lindley distribution. *Mathematical theory and Modeling*, 3(13), pp. 30–47.

Gelman, A., Rubin, D. B., (1992). A single series from the Gibbs sampler provides a false sense of security. In Bayesian Statistics 4 (J. M. Bernardo, J. 0. Berger, A. P. Dawid

and A. F. M. Smith, eds.). *Oxford University Press*. pp. 625–632.

Geweke, J., (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Bayesian Statistics 4 (eds. J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith), *Oxford University Press*, pp. 169–193.

Gupta, P., Pandey, A., and Tyagi, S., (2022). Comparison of Multiplicative Frailty Models under Generalized Log-Logistic-II Baseline Distribution for Counter Heterogeneous Left Censored Data, 1, pp. 97–114.

Hanagal, D. D., (2008a). Frailty regression models in mixture distributions. *Journal of Statistical Planning and Inference*, 138(8), pp. 2462–68.

Hanagal, D. D. (2019). Modeling Survival Data Using Frailty Models. 2nd Edition. *Springer*, Singapore.

Hanagal, D. D., & Pandey, A., (2014a). Inverse Gaussian shared frailty for modeling kidney infection data. *Advances in Reliability*, 1, pp. 1–14.

Hanagal, D. D., & Pandey, A., (2015a). Gamma frailty models for bivariate survival data. *Journal of Statistical Computation and Simulation*, 85(15), pp. 3172–3189.

Hanagal, D. D., Pandey, A., & Ganguly, A., (2017). Correlated gamma frailty models for bivariate survival data. *Communications in Statistics-Simulation and Computation*, 46(5), pp. 3627–3644.

Hanagal, D. D., & Pandey, A., (2017a). Shared inverse Gaussian frailty models based on additive hazards. *Communications in Statistics-Theory and Methods*, 46(22), pp. 11143–11162.

Hosmer, D. W., Royston, P., (2002). Using Aalen's linear hazards model to investigate time varying effects in the proportional hazards regression model. *Stat Journal*, 2(4), pp. 331–350.

Hougaard, P., (1985). Discussion of the paper by D.G. Clayton and J. Cuzick. *Journal of the Royal Statistical Society*, A, 148, pp. 113–14.

Hougaard, P., (1991). Modeling heterogeneity in survival data. *Journal of Applied Probability*, 28, pp. 695–701.

Hougaard, P., (2000). Analysis of Multivariate Survival Data. *Springer*, New York.

Ibrahim, J. G., Ming-Hui C. and Sinha, D., (2001). Bayesian Survival Analysis. *Springer*, Verlag.

Johnson, N. L., Kotz, S., (1975). A vector valued multivariate hazard rate. *Journal of Multivariate Analysis*, 5 (1), pp. 53–66. doi:10.1016/0047-259X(75)90055-X.

Lin, D. Y., Ying, Z., (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81(1), pp. 61–71.

Lindley, D. V., (1958). Fiducial distributions and Bayes's theorem. *Journal of the Royal Statistical Society*, B, 20, pp. 102–107.

McGilchrist, C. A., Aisbett, C. W.,(1991): Regression with frailty in survival analysis. *Biometrics*, 47, pp. 461–466.

Oakes, D.,(1982). Bivariate Survival Models Induced by Frailties. Journal of the American Statistical Association, 84(406), pp. 487–493.

Pandey, A., Lalpawimawha, L., & Bhushan, S., (2018). Additive shared inverse Gaussian frailty model. *Pak. J. Statist*, 34(4), pp. 311–330.

Pandey, A., Bhushan, S., Pawimawha, L., and Tyagi, S., (2020a).Analysis of Bivariate Survival Data using Shared Inverse Gaussian Frailty Models: A Bayesian Approach, Predictive Analytics Using Statistics and Big Data: Concepts and Modeling, *Bentham Books*, 14, pp. 75–88.

Pandey, A., Hanagal, D. D., Gupta, P., & Tyagi, S., (2020b). Analysis of Australian Twin Data Using Generalized Inverse Gaussian Shared Frailty Models Based on Reversed Hazard Rate. *International Journal of Statistics and Reliability Engineering*, 7(2), pp. 219–235.

Pandey, A., & Tyagi, S., (2021). Comparison of Multiplicative Frailty Models Under Weibull Baseline Distribution. *Lobachevskii Journal of Mathematics*, 42(13), pp. 3184–3195.

Pandey, A., Hanagal, D. D., Tyagi, S., and Gupta, P., (2021a). Generalized Lindley Shared Frailty Based on Reversed Hazard Rate. *International Journal of Reliability*, Quality and Safety Engineering, 2150040.

Pandey, A., Hanagal, D. D., and Tyagi, S., (2021b). Generalized Lindley Shared Frailty Models. *Statistics and Applications*, 19(2), pp. 41–62.

Pandey, A., Hanagal, D. D., Tyagi, S., & Gupta, P., (2022). Modeling Australian Twin Data Using Generalized Lindley Shared Frailty Models. *In Annual Conference of the Society of Statistics, Computer and Applications*, pp. 143–169. *Springer*, Singapore.

Santos, C. A., Achcar, J. A.,(2010). A Bayesian analysis for multivariate survival data in the presence of covariates. *Journal of Statistical Theory and Applications*, 9, pp. 233–253.

Shaked, M., Shantikumar, J. G. (1994). Stochastic Orders and Their Applications. *Academic Press*, New York.

Tyagi, S., Pandey, A., Hanagal, D. D., and Gupta, P.,(2021a). Bayesian inferences in generalized Lindley shared frailty model with left censored bivariate data. *Advance Research Trends in Statistics and Data Science*, pp. 137–157.

Tyagi, S., Pandey, A., Agiwal, V., and Chesneau, C.,(2021b). Weighted Lindley multiplicative regression frailty models under random censored data. *Computational and Applied Mathematics*, 40(8), pp. 1-24.

Tyagi, S., Pandey, A. & Chesneau, C., (2022a).Identifying the Effects of Observed and Unobserved Risk Factors Using Weighted Lindley Shared Regression Model. *J Stat Theory Pract 16*, 16, https://doi.org/10.1007/s42519-021-00241-9.

Tyagi, S., Pandey, A. & Chesneau, C.,(2022b). Weighted Lindley Shared Regression Model for Bivariate Left Censored Data. *Sankhya B.*, https://doi.org/10.1007/s13571-022-00278-1.

Vaupel, J. W., Manton, K. G. and Stallaed, E., (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, pp. 439–454.

Xie, X., Strickler,H.D., Xue, X., (2013). Additive hazard regression models: An application to the natural history of human papillomavirus. *Comput. Math. Methods Med.*, pp. 1–7.

Appendix

Tables and Figures

Table 1: Posterior Summary of Generalised Lindley Frailty with Baseline Generalised Weibull (Simulation Study: Model-I)

Parameter	*Estimate*	*s.e.*	*L.C.L.*	*U.C.L.*	*Geweke test*	$p-value$	*Gelman Rubin test*
burn-in period = 6900;		autocorrelation lag = 400					
$\zeta_1(22.5)$	23.07359	1.34778	20.32034	25.68016	0.00151	0.50060	1.01140
$\delta_1(0.013)$	0.01418	0.00291	0.00919	0.01884	-0.00694	0.49723	1.00058
$\xi_1(0.35)$	0.35559	0.03256	0.29407	0.42018	0.00562	0.50224	1.00174
$\zeta_2(22.5)$	22.77481	3.04714	17.77885	27.43331	-0.00067	0.49973	1.00054
$\delta_2(0.013)$	0.01395	0.00287	0.00916	0.01865	-0.00405	0.49839	0.99996
$\xi_2(0.33)$	0.33549	0.03187	0.27242	0.40088	-0.00716	0.49714	0.99996
$\theta(2.8)$	2.57442	0.50458	1.83115	3.55436	0.00639	0.50255	1.00289
$\mu(1.5)$	1.51839	0.18151	1.21484	1.93333	0.00094	0.50037	1.00154
$\beta_1(0.15)$	0.12616	0.06989	-0.00552	0.26851	-0.00119	0.49953	0.99996

Table 2: Posterior Summary of Generalised Lindley Frailty with Baseline Generalised Log-Logistic-II (Simulation Study: Model-II)

Parameter	*Estimate*	*s.e.*	*L.C.L.*	*U.C.L.*	*Geweke test*	$p-value$	*Gelman Rubin test*
burn-in period = 6900;		autocorrelation lag = 400					
$\zeta_1(4.5)$	4.42991	0.32462	3.78131	5.04944	0.00185	0.50074	1.00044
$\delta_1(0.02)$	0.02039	0.00291	0.01514	0.02481	-0.00352	0.49860	1.06288
$\xi_1(0.75)$	0.76296	0.06621	0.63513	0.87928	-0.00773	0.49692	1.01162
$\zeta_2(7.5)$	7.36200	0.55461	6.51050	8.40194	0.00984	0.50392	1.00001
$\delta_2(0.05)$	0.04829	0.00577	0.04045	0.05902	0.00597	0.50238	0.99996
$\xi_2(0.65)$	0.64983	0.06029	0.53887	0.77669	-0.00359	0.49857	0.99996
$\theta(4.8)$	4.65038	0.52427	3.85500	5.62394	0.01400	0.50559	1.00018
$\mu(2.5)$	2.50104	0.32257	1.90090	3.14192	0.01360	0.50543	0.99997
$\beta_1(0.15)$	0.13734	0.06798	-0.00299	0.27876	0.01239	0.50494	0.99996

Table 3: p-value of K-S statistics for goodness of fit test for Kidney Infection data set

Model	T_1 p-value	T_2 p-value
$Model-I$	0.7912	0.4490
$Model-II$	0.5722	0.6860

Table 4: Posterior Summary of Generalised Lindley Frailty with Baseline Generalised Weibull for Kidney Infection Data (Model-I)

Parameter	*Estimate*	*s.e.*	*L.C.L.*	*U.C.L.*	*Geweke test*	$p-value$	*Gelman Rubin test*
burn-in period = 6900;		autocorrelation lag = 400					
ζ_1	1.99061	0.06654	1.84351	2.10812	0.00268	0.50107	1.00003
δ_1	0.05530	0.00310	0.04954	0.06115	-0.00126	0.49950	1.00001
ξ_1	0.66315	0.02125	0.61705	0.70345	-0.00648	0.49741	1.00006
ζ_2	2.71016	0.06297	2.59528	2.83639	-0.00474	0.49811	0.99998
δ_2	0.06205	0.00311	0.05563	0.06818	0.00865	0.50345	1.00034
ξ_2	0.67052	0.02313	0.62961	0.71617	-0.00296	0.49882	0.99998
θ	2.29258	0.09757	2.11305	2.47659	0.00077	0.50031	0.99998
μ	1.38391	0.09709	1.20062	1.58699	-0.00608	0.49757	1.00025
β_1	-0.10576	0.01289	-0.13073	-0.08109	-0.00220	0.49912	0.99997
β_2	-8.88412	1.46382	-11.50565	-6.16638	0.00658	0.49912	1.00032
β_3	2.44371	0.33770	1.84343	3.11420	0.00775	0.50309	1.00075
β_4	1.61045	0.29506	1.08735	2.19532	-0.00539	0.49785	1.00093
β_5	-52.67579	27.25061	-101.04670	-4.02850	0.00950	0.50379	0.99996

Table 5: Posterior Summary of Generalised Lindley Frailty with Baseline Generalised Log-Logistic-II for Kidney Infection Data (Model-II)

Parameter	*Estimate*	*s.e.*	*L.C.L.*	*U.C.L.*	*Geweke test*	$p-value$	*Gelman Rubin test*
burn-in period = 6900;		autocorrelation lag = 400					
ζ_1	3.79268	0.10452	3.59085	4.01330	0.00174	0.50070	1.00009
δ_1	0.00160	0.00006	0.00148	0.00173	-0.00034	0.49987	1.00000
ξ_1	1.04034	0.04412	0.95236	1.11628	0.00441	0.50176	1.00096
ζ_2	4.30595	0.09619	4.11335	4.49439	-0.00039	0.49984	0.99997
δ_2	0.00043	0.00001	0.00041	0.00045	0.00431	0.50172	1.00010
ξ_2	1.25850	0.04593	1.16443	1.34386	0.00485	0.50194	0.99997
θ	2.12060	0.10701	1.92252	2.34629	0.00182	0.50072	0.99997
μ	1.28878	0.09992	1.10590	1.49725	-0.00126	0.49950	1.00025
β_1	-0.10630	0.01145	-0.12756	-0.08297	-0.00153	0.49939	0.99997
β_2	-67.94356	33.76583	-132.77210	-7.62369	0.00755	0.49939	1.00175
β_3	2.51987	0.26359	2.04163	2.97889	0.00187	0.50075	1.00056
β_4	1.51014	0.20139	1.14758	1.87047	-0.00018	0.49993	1.00104
β_5	-54.94150	31.09431	-111.91510	-3.56763	0.00406	0.50162	0.99997

Table 6: Kendall's τ Measure of Dependence

Model	Kendall's τ value
$Model-I$	0.297939
$Model-II$	0.303226

Table 7: AIC, BIC and DIC Comparison

Model	AIC	BIC	DIC
Model-I	685.3974	706.6861	664.4514
Model-II	686.2751	707.5637	665.7128