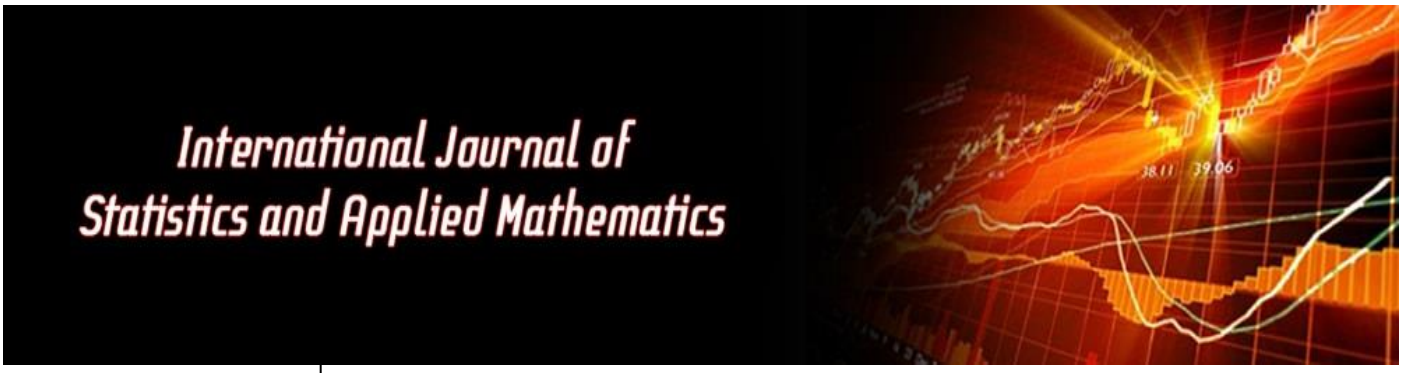
International Journal of Statistics and Applied Mathematics 2022; 7(5): 125-134



ISSN: 2456-1452
Maths 2022; 7(5): 125-134

www.mathsjournal.com
Received: 23-06-2022
Accepted: 27-07-2022

Nagaraj J
Research Scholar, Department of Statistics, Annamalai University, Annamalainagar, Chidambaram, Tamil Nadu, India

Parthasarathy S
Professor, Department of statistics, Annamalai university, Annamalainagar, Chidambaram, Tamil Nadu, India

Ponnuraja C
Scientist "E" ICMR-National Institute for Research in Tuberculosis (ICMR-NIRT), Chetpet, Chennai, Tamil Nadu, India

Corresponding Author:
Nagaraj J
Research Scholar, Department of Statistics, Annamalai University, Annamalainagar, Chidambaram, Tamil Nadu, India

Parametric frailty models under two-parameter Lindley distribution with applications to time-to-event analysis

Nagaraj J, Parthasarathy S and Ponnuraja C

DOI: https://doi.org/10.22271/maths.2022.v7.i5b.889

Abstract
A frailty model is a random effects model in survival or time-to-event analysis, where the random effect (the frailty) has a multiplicative effect on the hazard. Lindley distribution is one of the classical distributions, which is widely used in reliability and ordinary survival models (without frailty) but not in the frailty models. In recent years, Lindley distribution and its generalizations have played an important role in survival analysis due to its natural flexibility. In this study, we attempt to fit parametric frailty models with Two-parameter Lindley baseline distribution (TPLD) and apply them to the two real-life disease data sets of the (i) Recurrent asthma attacks in children's (Asthma attacks) and (ii) Culling of dairy heifer cow's (Culling) data. Comparison and assed the model's fitness were done using a minimum value of Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC). The study results revealed that TPLD with the Lognormal frailty model is a good choice for Culling data and Lindley with Gamma frailty model is the best for Asthma data. So, we suggest that the Two-parameter Lindley baseline distribution (TPLD) with frailty models are potential alternative models and will facilitate analyses of time-to-event data with covariates.

Keywords: Two-parameter Lindley distribution, frailty models, hazard function, time-to-event analysis

1. Introduction
Survival or time-to-event data analysis plays a major role in various fields such as biology, epidemiology, medicine and life sciences [1] and events may be death, recurrence, or any other outcome of interest [2] The measure of heterogeneity or unnoticed random effect shared by subjects is called "frailty", which is another essential component in survival analysis [3]. Clayton [4] first designed a model to account for such unobserved covariates in survival models and the term "frailty" was first introduced by Vaupel in 1979 [5] in the study of mortality. So, frailty (random effect) should be multiplied by the baseline hazard function and it will be given more accurate results than ordinary survival models (without frailty) [6]. Exponential, Weibull, Log-Logistic and Lognormal distributions are fitting with Gamma (GA), Inverse Gaussian (IG), Lognormal (LN) and Positive stable (PS) frailty distributions in most parametric survival models. In recent years, Lindley distribution and its generalizations have been significantly applied in lifetime data analysis due to its different hazard functions and shapes [8, 9] but not fitted with the frailty models. Therefore, in this study, we attempt to fit the parametric frailty models for Two-parameter Lindley distribution (TPLD) and use them in two real-life datasets. This paper is organized as follows. Section 2 deals with the properties of the TPLD distribution. Section 3 discusses frailty models, and then these methods were applied to two real-life data sets in Sections 4 and 5. Finally, conclusions are given in Section 6.

2. Two-parameter Lindley distribution (TPLD)
The Lindley distribution was introduced by Lindley in 1958 [10]. It is a mixture of gamma and exponential distributions with mixing proportions. It has the probability density function (p.d.f.)

$$f(t) = \frac{\lambda^2 (1+t)}{(\lambda+1)} e^{-\lambda t} \quad \lambda>0, t>0 \qquad (2.1)$$

Shanker and Mishra [11] suggested a Two-parameter Lindley distribution (TPLD) and proved a better fit than the one-parameter Lindley distribution for time-to-event analysis. Let us consider the continuous random variable "T" for an individual's lifetime in a particular population that follows a TPLD distribution. Then probability density function (p.d.f) is given by

$$f\ (t: \lambda, \alpha)\ = \frac{\lambda^2\ (1+\alpha t)}{\lambda+\alpha}\ e^{-\lambda t}\ \lambda, \alpha > 0 \text{ and } t > 0 \tag{2.2}$$

Then cumulative density function (c.d.f) is given by

$$\mathrm{F(t)} = 1 - \frac{\lambda+\alpha+\lambda\alpha t}{(\lambda+\alpha)} e^{-\lambda t}\ t>0, \lambda>0 \text{ and } \alpha > -\lambda \tag{2.3}$$

The probability of failure at time T is known as the survival rate ($S(t)$). For TPLD, $S(t\ ) = 1 - F\ (t\ )$, Therefore

$$\mathrm{S(t)} = \frac{\lambda+\alpha+\lambda\alpha t}{(\lambda+\alpha)} e^{-\lambda t}\ t>0, \lambda>0 \text{ and } \alpha > -\lambda \tag{2.4}$$

The hazard rate function is an instantaneous rate of failure given the survival until time "T", and it is defined as $h(t) = f(t)/s(t)$; the hazard rate of the TPLD given by

$$h(t) = \frac{\lambda^2\ (1+\alpha t)}{\lambda+\alpha+\lambda\alpha t} \tag{2.5}$$

where t>0, λ >0 and α > - λ and particular case: α = 1 the one-parameter Lindley distribution.
The hazard function can also be represented as the "cumulative hazard function". Therefore, the cumulative hazard function (H(t)) of the TPLD given by

$$H(t)\ = \int_0^t h(t)\ dt\ = -log\ (S(t));$$

where $h(t) = -(\frac{dlog\ S(t)}{dt})$

$$H(t) = -\ log(\frac{\lambda+\alpha+\lambda\alpha t}{(\lambda+\alpha)} e^{-\lambda t}) \tag{2.6}$$

3. Frailty models

Frailty models account for random effect (unobserved heterogeneity) in time-to-event analysis rather than ordinary models (without frailty) in survival analysis implicitly assuming that populations are homogenous [12], meaning all individuals have the same risk of an event. In frailty models, unobserved heterogeneity is assumed to represent different clusters, and clusters are considered to be independent and assume proportional hazards structure conditional on the random effect, "Z" [13]. Let us consider random effect "Z" to be a frailty variable (non-negative) indicates the individual-level risk of the study population. The conditional hazard then represents the frailty model as

$$h_{ij}\ (t \setminus Z_i)\ =\ h_0(t).Z_i.exp(x_{ij}^T\ \beta) \tag{3.1}$$

Where

i. j = subject (J: 1,2, ….n) and i = group (I:1, 2,…. G)
ii. $h_0(t)$ - is the baseline hazard function (here TPLD) for all individuals/subjects.
iii. The Z_i is an unobserved random effect common to all observations from cluster group "i" and assumed to be IID (independently and identically distributed) random variables with a common density function $f(z, \theta)$. Here, θ is the parameter of the frailty distribution.
iv. The factor $exp(x_{ij}^T\ \beta)$ gives that subject-specific contribution to the hazard. x_{ij} is the covariates vector for the subject "j" in the group "i," and β is the unknown regression coefficient vector.

In this paper, we attempt to fit four frailty distributions, namely, Gamma (Ga), Inverse Gaussian (IG), Lognormal (LN) and Positive Stable (PS), with TPLD baseline distribution. The properties of the above-mentioned frailty distributions are well documented in previous studies [13-15]. In Sections 3.1 - 3.3 we explained that Marginal log-likelihood, Laplace transform, estimation and prediction techniques thereafter, we illustrate the four frailty distributions which are used for comparison and analysis in Sections 3.4 - 3.7.

3.1. Marginal log-likelihood

A marginal likelihood approach is used for estimating the parameters in frailty models [16] parametric frailty models, the frailties seeming in the conditional likelihood can be integrated to maximize the marginal likelihood, leading to estimates of the model parameters [17,18] . In right-censored cluster survival data, the marginal log-likelihood estimation is obtained by the following assumptions (i) random variables are independent between the censoring time and the survival time, (ii) non-informative right-censoring and (iii) the covariate information and the marginal log-likelihood $U = \{U_{ij}; i \in I, j \in J_i\}$ [16].

$L_{marg\ (\psi,\beta,\xi;\ u\ |\ \tau)}=\sum_{i=1}^{G}\{[\sum_{j=1}^{ni}\ \delta_{ij}\ (\log(h_0(yij))+x_{ij}^T\beta)]+\log[(-1)^{d_i}L^{d_i}(\sum_{j=1}^{ni}H_0(yij)\exp(x_{ij}^T\beta)]$

$$-\log[L(\sum_{j=1}^{ni}H_0(yij)\exp(x_{ij}^T\beta)] \quad (3.2)$$

Where
$d_i = \sum_{j=1}^{ni}\ \delta_{ij}$ the no. of events in the $i-th$ cluster.

3.2. Laplace transform

The Laplace transform (L(s)) was first introduced by Hougaard [12] and is used to characterize the density functions of the frailty distributions. Further, unconditional hazards and survival functions can be easily stated in this approach. Hence, the maximum likelihood function can be represented with easy Laplace transforms, and it was essential in frailty models for parameter estimation [14]. $L^{(q)}(\cdot)$, the q[th] derivative of the Laplace transform [17] of the frailty distribution is given by

$$L(s) = E\ [exp(-Zs)]\ =\ \int_0^{\infty}\exp(-z_i s)\,f(z_i)dz_i \quad (3.3)$$

Where $L^{(q)}(\cdot)$ is the Higher-order derivatives of the Laplace transform up to $q\ =\ max\{d1,\ldots,dG\}$. Hence q[th] derivate is given by

$$L^{(q)}(s) = (-1)^{(q)}\,E(Z^{(q)}\exp(-zs)) \quad (3.4)$$

3.3. Estimation and prediction

Estimations of ψ, β, and ξ are found by maximizing the marginal log-likelihood; and this can simply be by calculating the higher-order derivatives $L^{(q)}(.)$ of the Laplace transform (L(s)) up to $q\ =\ max\{d1,\ldots,dG\}$. A combination of an expectation, and maximization is called the "EM algorithm", and it was used to predict the frailties. The frailty z_i is predicted by $z_i = E(Z\ |u_i,\tau_i\ ;\ \psi,\beta,\xi)$; where "u_i" and "τ_i" are the data and the truncation times of the i-th cluster. Therefore, the conditional expectation is given by

$$E(Z\ |u_i,\tau_i\ ;\ \psi,\beta,\xi) = -\frac{L^{(d_i+1)}\,[\sum_{j=1}^{ni}H_0(yij)\exp(x_{ij}^T\beta)]}{L^{d_i}\cdot[\sum_{j=1}^{ni}H_0(yij)\exp(x_{ij}^T\beta)]} \quad (3.5)$$

3.4 Gamma frailty (GA)

A continuous random variable X that takes any non-negative values and it follows a Gamma distribution if its pdf of the form

$$f(x) = \frac{\theta^{-\frac{1}{\theta}}x^{\frac{1}{\theta}-1}\exp(-x/\theta)}{\Gamma(1/\theta)}, \theta > 0,$$

and its denoted by $X\sim Gam*(\theta)$. where $\Gamma(.)$ is the gamma function. It corresponding to gamma distribution Gamma(μ, θ). mean (μ)=1, and variance = θ.
The associate Laplace transformation is given by

$$L(s) = (1+\theta s)^{-\frac{1}{\theta}}, s \geq 0,$$

and it is easy to show that, for q≥1,

$$L^q(s) = (-1)^q(1+\theta s)^{-q}\left[\prod_{l=0}^{q-1}(1+l\theta\right]\ L(s).$$

Therefore, in equation 7, we have

$$\log\left((-1)^q L^{(q)}(s)\right) = \left(q+\frac{1}{\theta}\right)\log(1+\theta s) + \sum_{l=0}^{q-1}\log(1+l\theta)$$

Which measures the close relationship between any two-event time from the same cluster in the multivariate case, can be computed as

$$T = \frac{\theta}{\theta+2} \in (0,1).$$

3.5 Inverse Gaussian frailty (IG)

The inverse Gaussian frailty distribution IG* (θ) has density

$$f(x) = \frac{1}{\sqrt{2\pi\theta}}x^{-\frac{3}{2}}.exp\left(-\frac{(x-1)^2}{2\theta x}\right), \theta > 0.$$

The mean and variance are 1 andθ, respectively. For the Laplace transform, one has

$$L(s) = exp\left(\frac{1}{\theta}\left(1 - \sqrt{1 + 2\theta s}\right)\right), s \geq 0,$$

and, for q≥1,

$$L^q(s) = (-1)^q(1 + 2\theta s)^{-\frac{q}{2}} \frac{K_{q-\left(\frac{1}{2}\right)}\left(\sqrt{2\theta^{-1}}(s + \frac{1}{2\theta}\right)}{K_{\left(\frac{1}{2}\right)}\left(\sqrt{2\theta^{-1}(s + \frac{1}{2\theta}}\right)}.L(s),$$

Where K is the modified Bessel function of the second kind [18].

$$K_\gamma(\omega) = \frac{1}{2}\int_0^\infty t^{\gamma-1}\, exp\left\{-\frac{\omega}{2}\left(t + \frac{1}{t}\right)\right\} dt\ \gamma \in \mathbb{R}, \omega > 0.$$

The general construction method to obtain the above equation and derivative of the Laplace transform for any distribution for which the moments of $W/z_i, T_I; \delta, \beta, \xi$, the conditional frailty given the data, are known.
Noting that

$K_{\frac{1}{2}}(\omega) = \sqrt{\frac{\pi}{2\omega}}\exp(-\omega)$, we have

$$log\big((-1)^q L^{(q)}(s)\big) = -\frac{q}{2}\log(2\theta s + 1) + \log\left(K_{q-\left(\frac{1}{2}\right)}(z)\right) - \left[\frac{1}{2}\left(l0g\left(\frac{\pi}{2z}\right)\right) - z\right] + \frac{1}{\theta}\big(1 - \sqrt{1 + 2\theta s}\big),$$

With $z = \sqrt{2\theta^{-1}}\left(s + \frac{1}{2\theta}\right)$, with multivariate data, an inverse Gaussian distribution frailty yields given by

$$T = \frac{1}{2} - \frac{1}{\theta} + 2\frac{\exp(2/\theta)}{\theta^2}\int_{2/\theta}^\infty \frac{\exp(-x)}{x} dx\ x \in (0,1/2)$$

3.6. Positive stable frailty (PS)
The family of the positive stable distribution with two parameters. A scale $\delta > 0$ and the so-called index $\alpha < 1$. The positive stable frailty distribution PS*(γ), with $\gamma = 1 - \alpha$, is produced by imposing $\delta = \alpha$.
The related probability density function is given by

$$f(u) = -\frac{1}{\pi u}\sum_{k=1}^\infty \frac{\Gamma(k(1-\gamma)+1}{k!}(-u^{-1})^k\, sin\big((1-\gamma)k\pi\big), \gamma \in (0,1).$$

Both the mean and variance are unknown. As a result, the variance of the frailty term does not correlate to the heterogeneity parameter. Because of this, we purposefully refer to it as "instead of" to prevent misunderstanding.
The accompanying Laplace transform has a much simpler shape than the probability density function.

$$L(s) = EXP(-s^{1-\gamma}), s \geq 0,$$

And for q≥1,

$$L^{(q)}(s) = (-1)^q(1-\gamma)s^{-\gamma})^q\left[\sum_{m=0}^{q-1}\Omega_{q,m}s^{-m(1-\gamma)}\right]L(s),$$

Where the $\Omega_{q,m}'s$ are polynomial of degree m, given recursively by

$$\Omega_{q,0} = 1,$$

$$\Omega_{q,m} = \Omega_{q-1,m} + \Omega_{q-1,m-1}\left\{\frac{q-1}{1-\gamma} - (q-m)\right\}, m = 1,2,\dots\dots, q-2,$$

$$\Omega_{q,q-1} = (1-\gamma)^{1-q}\frac{\Gamma(q-(1-\gamma))}{\Gamma(\gamma)}.$$

It follows that

$$log\ ((-1)^q L^{(q)}(s) = q\big(log(1-\gamma) - \gamma log(s)\big) + log\left[\sum_{m=0}^{q-1} \Omega_{q,m} s^{-m(1-\gamma)}\right] - s^{1-\gamma}$$

With clustered data, the Kendall's tau for positive stable distribution frailty is

$T = \gamma \in (0,1)$.

3.7 Lognormal frailty (LN)

Let as consider that the continuous random variable X with scale parameterθ, and lognormal frailty distribution LN* (θ) has density

$$f(x) = (2\pi\theta)^{-1/2} x^{-1} \exp\left\{\left[-\frac{(logx)^2}{2\theta}\right]\right\}, \theta > 0$$

If $X \sim LN(\theta)$, subsequently, the closed form of the Laplace transform does not exist. Consequently

$$L^q(s) = (-1)^q \int_0^\infty x^q exp(-xs)\, f(x) dx$$

$$L^q(s) = (-1)^q \frac{1}{\sqrt{2\pi\theta}} \int_0^\infty x^q exp(-xs) \frac{1}{x} exp\left(-\frac{1}{2\theta}(\log(x)^2\right) dx$$

we have (s≥0) needs to be roughly estimated. The modification of the variable u=log(x) has allowed us to:

$$L^q(s) = (-1)^q \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^\infty (\exp(u))^q exp(-\exp(u)s) \frac{1}{x} exp\left(-\frac{u^2}{2\theta}\right) du$$

$$L^q(s) = (-1)^q \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^\infty exp\left\{qu - \exp(u)\, s - \frac{u^2}{2\theta}\right\} du$$

Using the Laplace integral approximation, we can approximate this. Let

$$g(u; s, \theta) \coloneqq -qu + \exp(-\exp(u)\, s + \frac{u^2}{2\theta}$$

$$g^{'}(u; s, \theta) \coloneqq \frac{dg}{du}(u; s, \theta) = -q + \exp(u)\, s + \frac{u}{\theta}$$

$$g^{''}(u; s, \theta) \coloneqq \frac{d^2 g}{du^2}(u; s, \theta) = \exp(u)\, s + \frac{1}{\theta} > 0$$

In order to approximate g(.), the first three terms of its Tylor series expansion are used instead of û.

$$g(u; s, \theta) \approx g(\hat{u}; s, \theta) + (u - \hat{u}) g^{'}(u; s, \theta) + \frac{(u-\hat{u})^2}{2} g^{''}(u; s, \theta)$$

The value of û is chosen such that $g^{'}(u; s, \theta) = 0$, such that $L^q(s)$ can be approximated by

$$L^q(s) \approx (-1)^q \frac{1}{\sqrt{2\pi\theta}} exp\{-g(\hat{u}; s, \theta)\} * \int_{-\infty}^\infty exp\left\{-\frac{(u-\hat{u})^2}{2} g^{''}(\hat{u}; s, \theta)\right\} du$$

$$= (-1)^q \frac{1}{\sqrt{\theta}} exp\{-g(\hat{u}; s, \theta)\}\left[g^{''}(\hat{u}; s, \theta)\right]^{-1/2}$$

Recognizing the kernel of a normal density with a mean of û and a variance of $1/g^{''}(\hat{u}; s, \theta)$ leads to the last line. This is known as Laplace approximation.

4. Applications

Data set I: First, we applied the Asthma attacks data set: Recurrent asthma attacks in infants [14] to fit the four frailty models with the TPLD baseline. Asthma is occurring more and more frequently in infants. Therefore, objective of the study is a new application of an existing anti-allergic drug is administered to children who are at higher risk of developing asthma in order to prevent it. A prevention trial is set up with such children randomized to a placebo or drug, and the asthma events that developed over time are recorded in a diary. Typically, a patient has more than one asthma event. The different events are thus clustered within a patient and are ordered in time. This ordering can be considered in the model. The data set contains 232 (1776 observations) asthmatics infants between the age of 6 weeks and 24 months and the variables of (i) ID: Infant identification number (ii) Time (duration days; the time from the end of the previous event (asthma attack) to the start of the next event (iii) Status (Censored=0 or observed=1) event time) (iv) Drug (placebo=0 or drug=1) and (v) Fevent (First observation of the patient? 1=yes, 0=no).

Data set II: Second, we used a Culling data set: Culling of dairy heifer cows [14] to fit the four frailty models with the TPLD baseline. The objective of the study was time to culling in heifers as a function of the somatic cell count (SCC) measured between 5 and 15 days (measurement day) after calving. High somatic cell count (logarithm of somatic cell count (LogSCC) might be a surrogate marker for intramammary infections. Heifers with intramammary infections or expected to develop intramammary infections in the future are handling has expensive to keep. The dataset contains the data of 1702 observations that (i) Cow id: Cow's identifier (ii). Time (Time to culling (in days) (iii) Status (Censored (0) or observed (1) event time) (iv). Herd (Cluster:Herd identifyier) (v) LogSCC (Logarithm of the somatic cell count).

4.1. Data analysis

R studio version 1.2.50 was used to create the codes and function for TPLD distribution with frailty models and data analysis. Akaike's Information Criteria $(AIC = -2(loglikelihood) + 2P)$ and Bayesian Information Criteria $(BIC = -2(loglikelihood) + P(log/n))$ (where P is the number of parameters) were used to assess the fitness of the models and also identify the best model. The method of Kendall's tau was used to measure the relation between any two event times from the same cluster [19]). R packages of "Survival" [20], "parfm" [17], "frailtyEM" [21] and "frailtypack" [22] were referred to create the codes/function for TPLD.

5. Results

The Asthma attacks and Culling data sets were used to fit the parametric frailty models under the TPLD baseline. Table 1 shows the comparison results of the four frailty models for each data set. The model results revealed that TPLD with Gamma (Ga), Inverse Gaussian (IG), and Lognormal (LN) frailty models gave almost close results to covariates in both Asthma attacks and culling data. The TPLD baseline with Gamma (Ga) frailty distribution was found to be the best compared to other models for Asthma attacks data due to the lowest AIC (16625.32) and BIC (16646.47) values. Estimated hazard ratio [95% Wald CI] of significant (P <0.05) covariates are shown in Figure 1. The frailty values were predicted for each Asthma attack infants based on the TLPD with Gamma frailty model, as shown in Figure.2.

To check the model efficiency, AIC and BIC values were compared with four frailty models with four baseline distributions such as TPLD, Weibull, Lognormal and Log-Logistic for Asthma attacks data is shown in Figure 3-4. The results revealed that the Gamma frailty model is an excellent choice for this data because minimum AIC and BIC values were observed in all the baseline distributions. However, the lowest AIC and BIC values were recorded in the TLPD baseline with the Gamma frailty model (Figures 3-4).

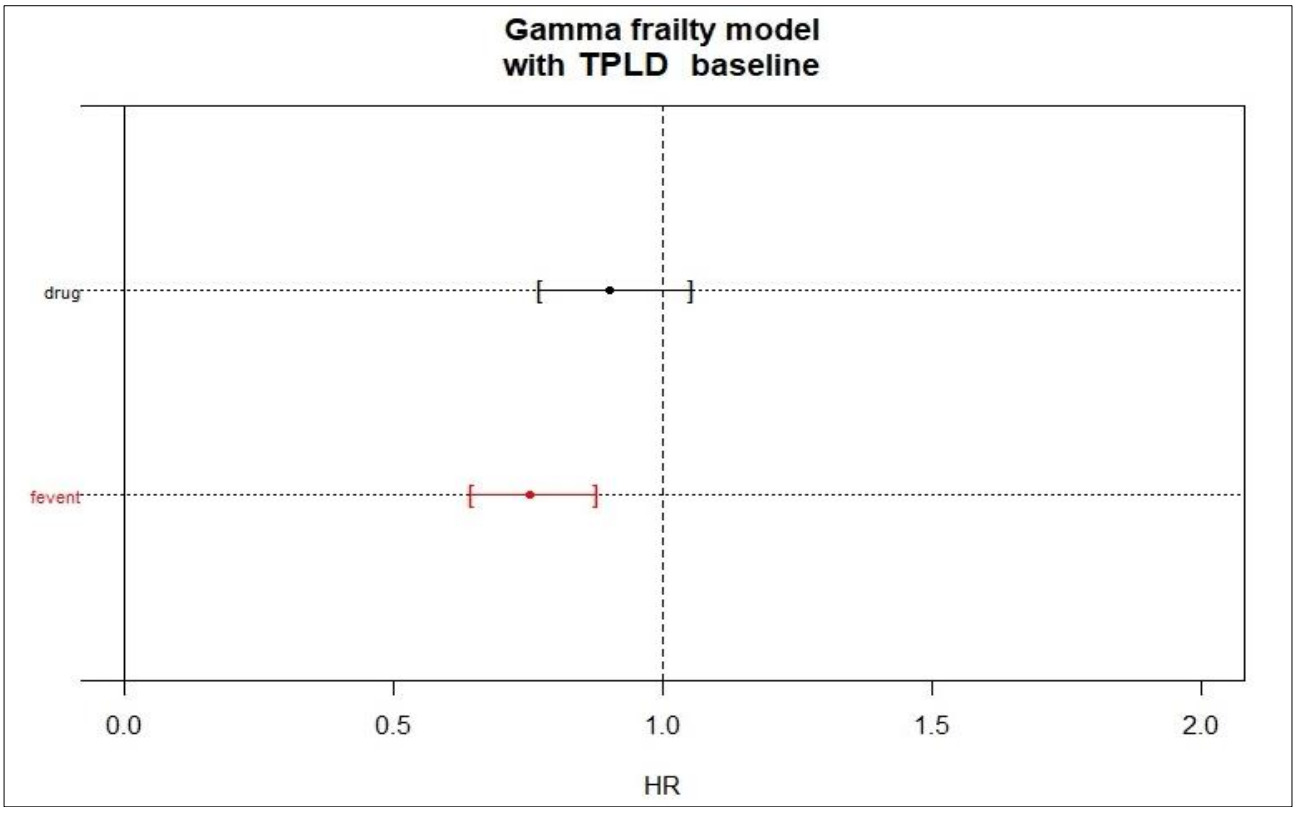


Fig 1: The Hazard Ratio for Gamma (GA) frailty Model with TPLD baseline for Asthma attacks data (Significant covariate with 95% CI has coloured in red)

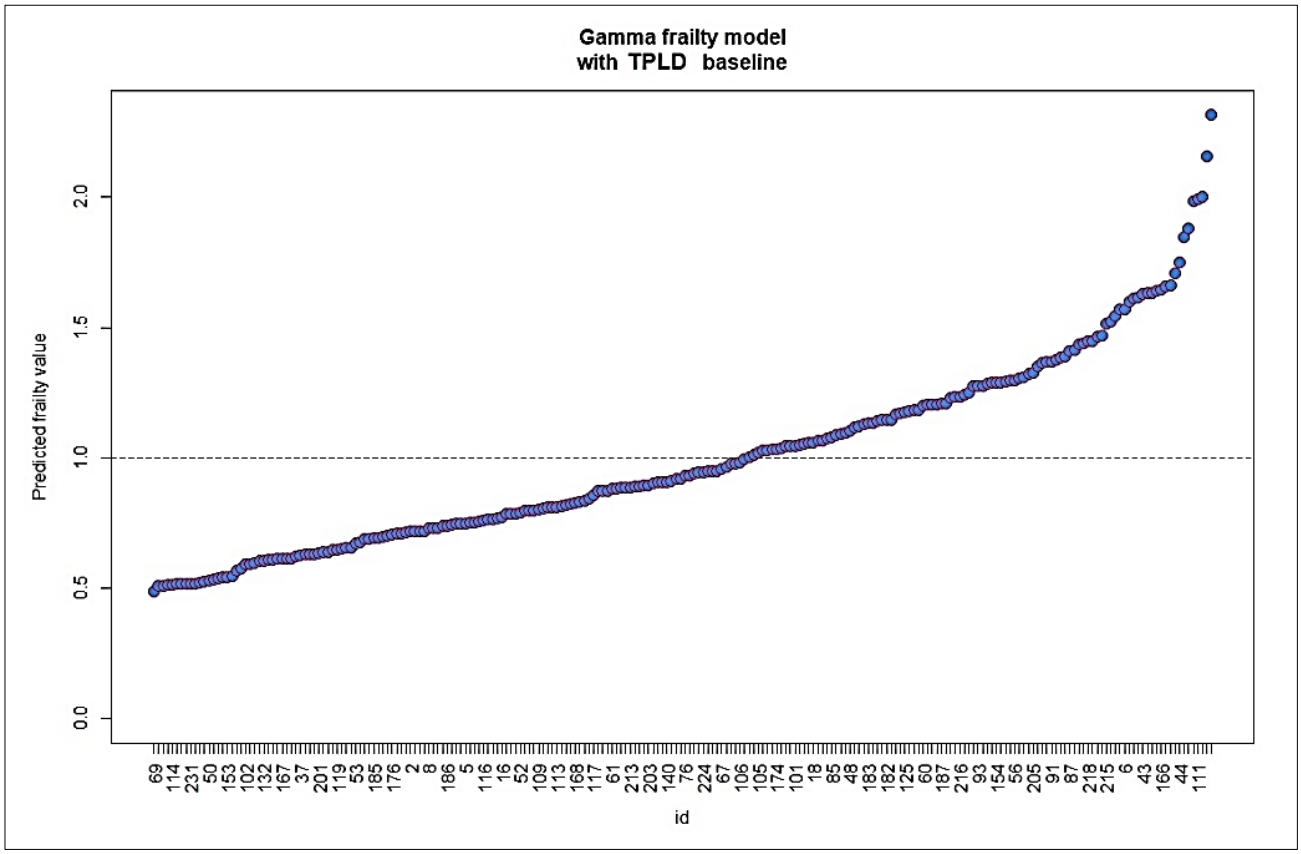


Fig 2: Predicted frailty values for each Asthma attack infants bsed on Gamma (GA) frailty model with TPLD baseline

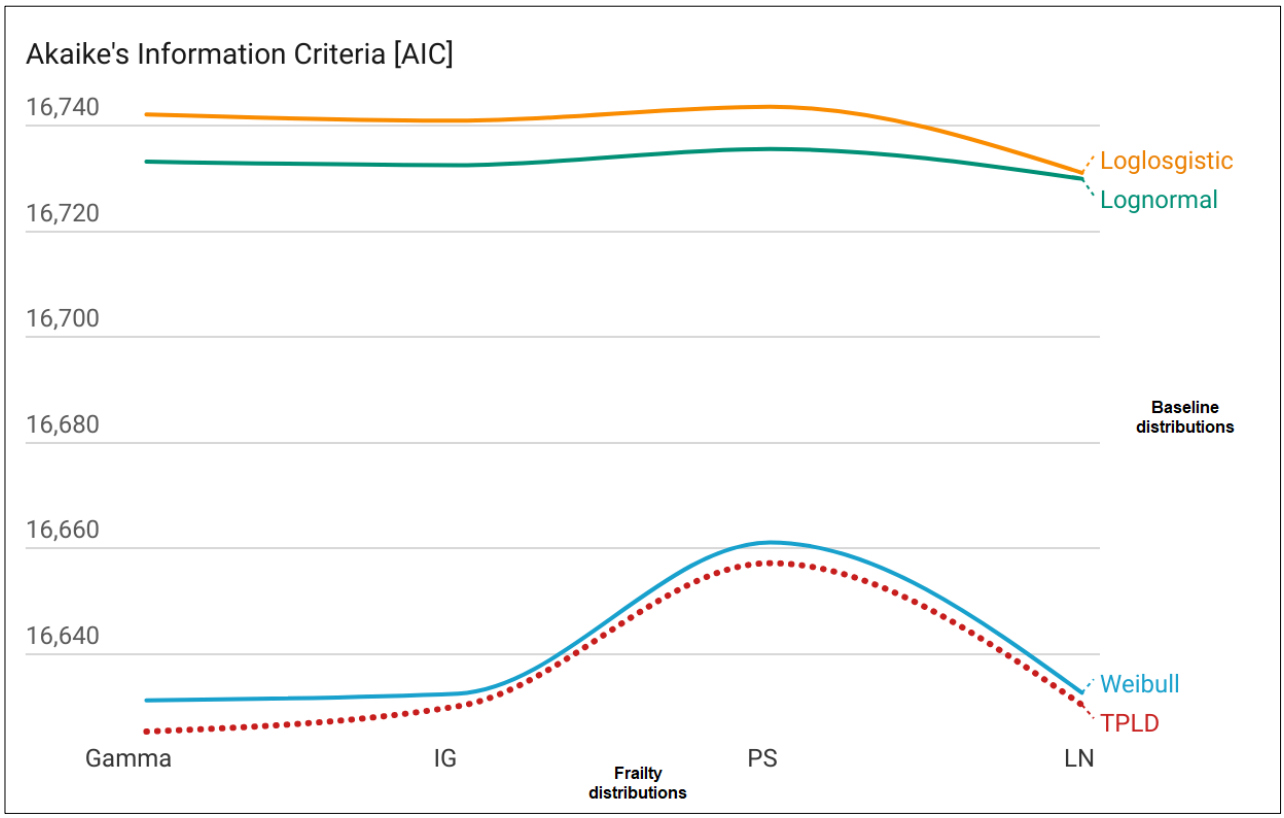


Fig 3: Comparison of AIC values for Asthma attacks data

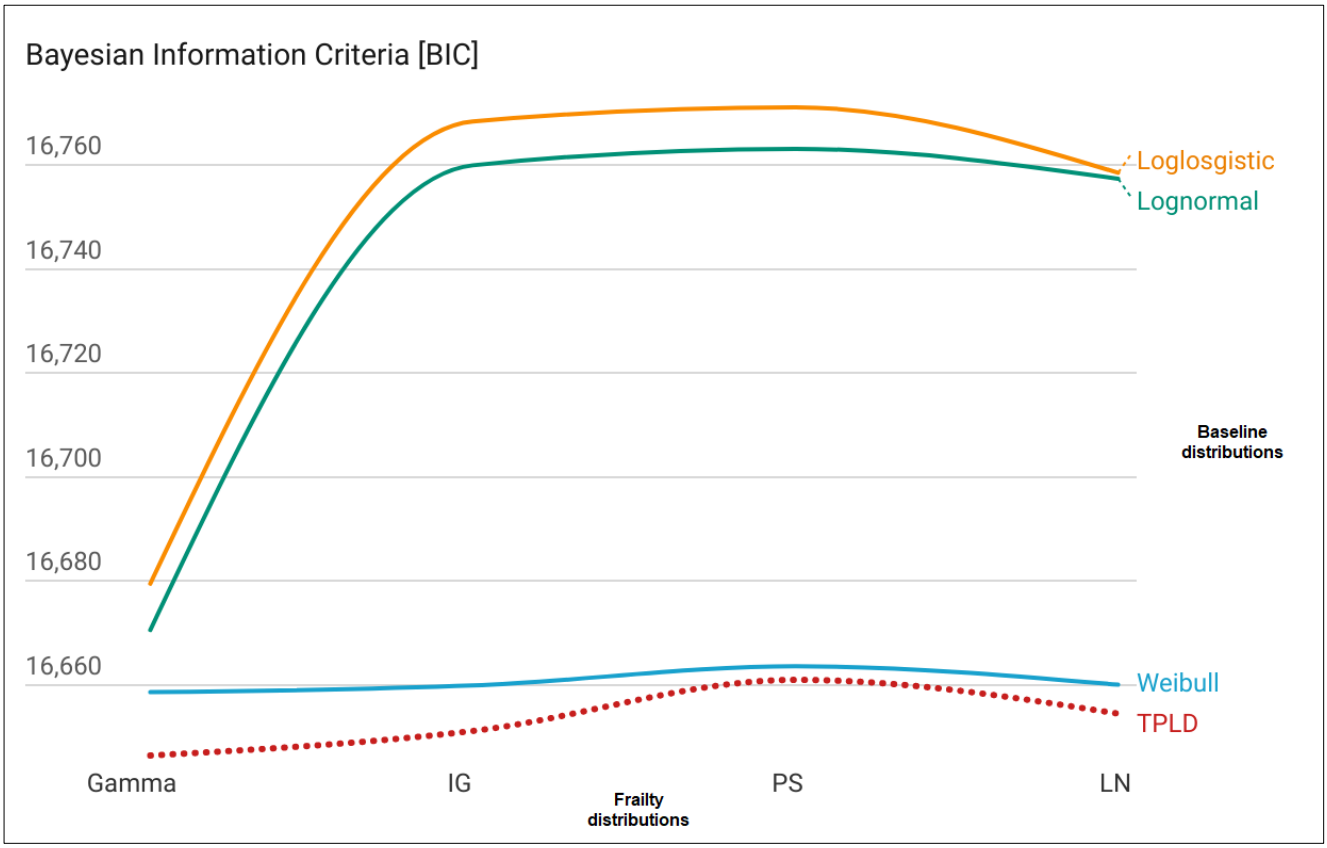


Fig 4: Comparison of BIC values for Asthma attacks data

Table 1: Frailty models comparison under Two Parameter Lindley baseline distribution (TPLD) for (I) Asthma attacks and (II) Culling data set

Data set	Parameters/ Covariates	Gamma (Ga)		Inverse Gaussian (IG)		Positive Stable (PS)		Lognormal (LN)	
		Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Asthma Attacks	*Frailty*	0.210	0.040	0.230	0.047	0.085	0.020	0.205	0.040
	λ	0.023	0.018	0.025	0.003	0.025	0.003	0.024	0.002
	α	0.765	0.003	0.866	0.018	0.850	0.019	0.865	0.018
	Drug	-0.103	0.080	-0.106	0.08	-0.089	0.077	-0.105	0.081
	Fevent	-0.276*	0.078	-0.289*	0.077	-0.347*	0.076	-0.302*	0.078
	AIC	16625.32		16630.12		16657.18		16630.37	
	BIC	16646.50		16651.20		16661.01		16654.52	
	Kendall's Tau	0.097		0.086		0.084		0.091	
Culling	*Frailty*	0.102	0.038	0.106	0.039	0.009	0.007	0.134	0.040
	λ	0.051	0.003	0.053	0.004	0.054	0.004	0.052	0.003
	α	0.547	0.090	0.548	0.091	0.542	0.090	0.545	0.089
	LogSCC	0.086*	0.019	0.087*	0.020	0.085*	0.020	0.090*	0.018
	AIC	14149.47		14144.18		14153.02		14140.24	
	BIC	14198.19		14197.89		14201.74		14193.96	
	Kendall's Tau	0.049		0.046		0.010		0.062	
*Significantly differed at 0.1% level (P<0.001)									

In the Culling data, TPLD with Lognormal (LN) frailty distribution was identified as the best model due to minimum AIC (14140.24) and BIC (14193.96) values. The estimated hazard ratio with confidence interval [95% Wald CI] for covariate (Figure 5) and frailty values were predicted each id based on this model is shown in Figure 6.

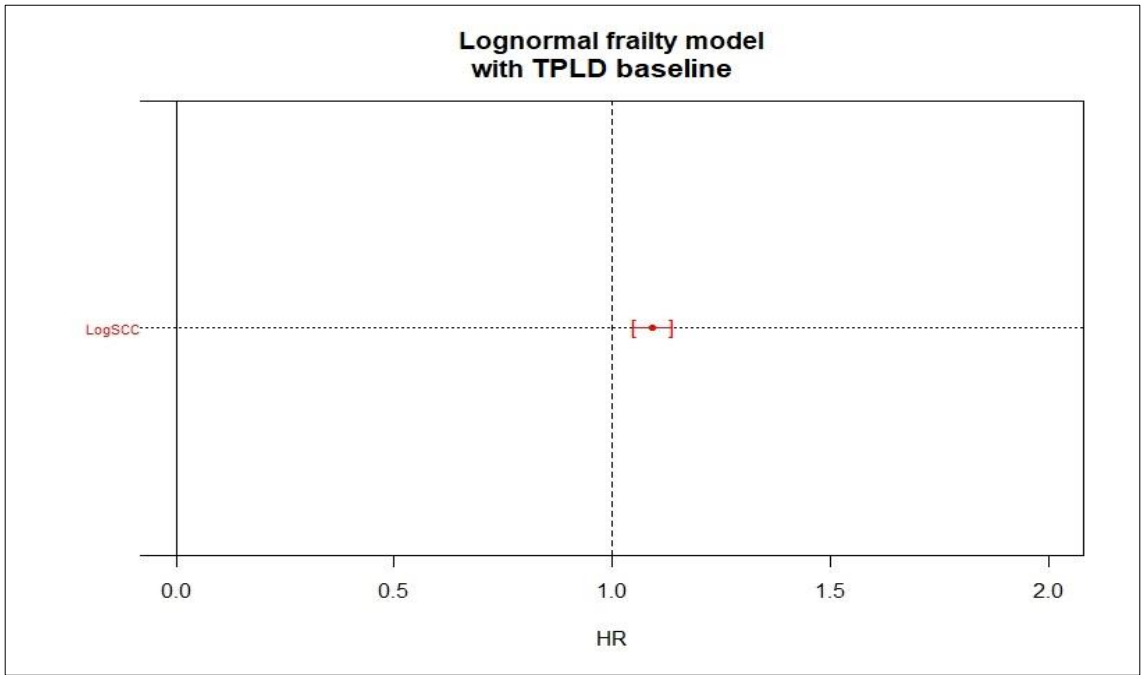


Fig 5: The Hazard Ratio for Lognormal (LN) frailty Model with TPLD baseline for Culing data (Significant covariate with 95% CI has coloured in red)

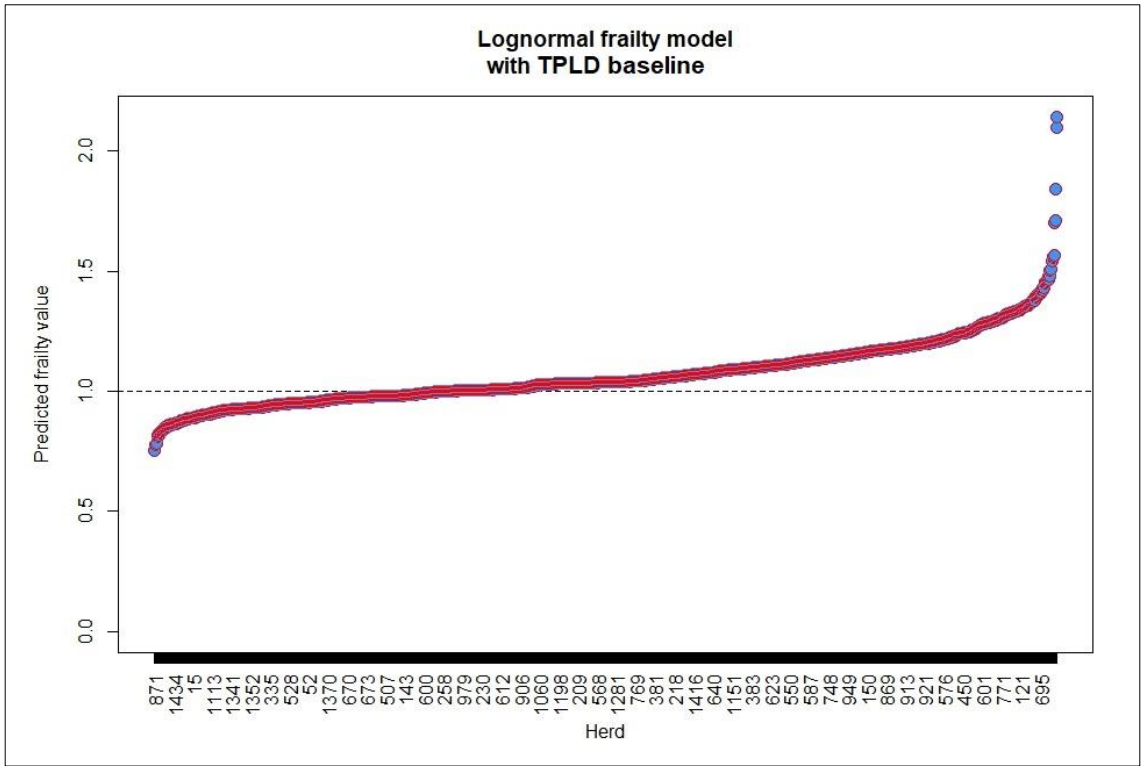


Fig 6: Predicted frailty values for each Herd (Cow id) based on Lognormal (LN) frailty model with TPLD baseline

6. Conclusion

In practice, identifying and fitting an appropriate baseline with frailty distribution is crucial and it has given better outcomes than other ordinary analyses in time-to-event data analysis. In this paper, a Two-parameter Lindley distribution (TPLD), has been proposed as a baseline distribution and fitted with parametric frailty models. Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) were used to identify the best model for Asthma attacks and Culling data. The TLPD with Gamma (GA) and Lognormal (LN) frailty distributions were identified as the best-fitted model for Asthma attacks and Culling data respectively. And also, the TPLD baseline with frailty models are showed better fits than other baseline distributions. So we suggest that TPLD baseline distribution with the frailty models is a potential alternative approach for time-to-event data analysis.

7. References

1. Abbas SA, Subramanian S, Ravi P, Ramamoorthy S, Munikrishnan V. An Introduction to Survival Analytics, Types, and Its Applications. In (Ed.), Biomechanics. Intech Open; c2019. https://doi.org/10.5772/intechopen.80953

2. Kumar M, Sonker PK, Saroj A, Jain A, Bhattacharjee A, Saroj RK. Parametric survival analysis using R: Illustration with lung cancer data. Cancer Rep (Hoboken). 2020;3(4):e1210. doi:10.1002/cnr2.1210
3. Theodor A, Balan, Hein Putter. A tutorial on frailty models. Statistical Methods in Medical Research. 2020;9(11):3424-3454. DOI:10.1177/0962280220921889
4. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika. 1978;(65):141-151.
5. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography. 1979;16:439-454.
6. David Hanagal D. Modeling Survival Data Using Frailty Models, Industrial and Applied Mathematics, Springer Nature Singapore Pte. Ltd, 2011, 295. https://doi.org/10.1007/978-981-15-1181-3.
7. Duchateau L, Janssen P. The Frailty Model, Series of Statistics for Biology and Health, Springer-Verlag New York, 2008. https://doi.org/10.1007/978-0-387-72835-3
8. Tomy LA. Retrospective study on Lindley distribution. Biom Biostat Int J. 2018;7(3):163-169. DOI: 10.15406/bbij.2018.07.00205
9. Shanker R, Fesshaye H, Selvaraj S. On modeling of lifetimes data using exponential and lindley distributions. Biom Biostat Int J. 2015;2(5):140-147. DOI: 10.15406/bbij.2015.02.00042
10. Lindley DV. Fiducial distributions and Bayes theorem, Journal of the Royal Statistical Society A. 1958;20(1):102-107.
11. Shanker R, Mishra A. A two parameter Lindley distribution. Statistics in transition new series. 2013;14(1):45-56.
12. Hougaard P. Life table methods for heterogeneous populations. Biometrika. 1984;71:75-83.
13. Wienke A. Frailty Models in Survival Analysis, Chapman & Hall/CRC, Boca Raton; c2010.
14. Duchateau and Janssen P. The Frailty Model, Series of Statistics for Biology and Health, Springer-Verlag New York. 2008. https://doi.org/10.1007/978-0-387-72835-3.
15. David D. Hanagal. Modelling Survival Data Using Frailty Models, Industrial and Applied Mathematics, Springer Nature Singapore Pte. Ltd, 2011, 295. https://doi.org/10.1007/978-981-15-1181-3.
16. Lam KF, Anthony Kuk YC. A Marginal Likelihood Approach to Estimation in Frailty Models, Journal of the American Statistical Association. 1997;92:439, 985-990. DOI: 10.1080/01621459.1997.10474053.
17. Marco Munda, Federico Rotolo, Catherine Legrand. “parfm”: Parametric Frailty Models in R, Journal of Statistical Software, Foundation for Open Access Statistics. 2012;51:1-20. http://hdl.handle.net/10.18637/jss.v051.i11
18. Balan TA, Putter H, Frailty EM. An R package for estimating semi parametric shared frailty models, Journal of Statistical Software. 2019;90(7):1-29. https://doi.org/10.18637/jss.v090.i07.
19. Daniel, Wayne W. "Kendall's tau". Applied Nonparametric statistics (2nd ed.). Boston: PWS-Kent, 1990, 365-377. ISBN 978-0-534-91976-4.
20. Therneau TM. “Survival”: A Package for Survival Analysis in S. R package version 2.44-1.1. 2019. URL https://CRAN.R-project.org/package=survival.
21. Balan TA, Putter H, Frailty EM. An R package for estimating semi-parametric shared frailty models, Journal of Statistical Software. 2019;90(7):1-29. https://doi.org/10.18637/jss.v090.i07.
22. Rondeau V, Mazroui Y, Gonzalez JR. frailtypack: An R package for the analysis of correlated survival data with frailty models using penalised likelihood estimation or parametrical estimation, Journal of Statistical Software. 2012;47(4):1-28. URL http: //www.jstatsoft.org/v47/i04/.