Advances and Applications in Statistics

http://www.pphmj.com
http://dx.doi.org/10.17654/0972361722031
Volume 75, 2022, Pages 119-134 **P-ISSN: 0972-3617**

LINDLEY DISTRIBUTION AS FRAILTY MODELS WITH APPLICATION TO LIFETIME DATA

J. Nagaraj[1,*], S. Parthasarathy[1] and C. Ponnuraja[2]

[1]Department of Statistics
Annamalai University
Annamalainagar, Chidambaram-608 002
Tamil Nadu, India
e-mail: nagarajicmr@gmail.com
statsarathy@yahoo.co.in

[2]ICMR-National Institute for Research in Tuberculosis (ICMR-NIRT)
Chennai - 600 031, Tamil Nadu
India
e-mail: cponnuraja@gmail.com

Received: November 8, 2021; Revised: January 20, 2022; Accepted: March 7, 2022
2020 Mathematics Subject Classification: Primary 60E05; Secondary 62N03.
Keywords and phrases: Lindley distribution, frailty models, Lindley with frailty models, survival analysis.
*Corresponding author

How to cite this article: J. Nagaraj, S. Parthasarathy and C. Ponnuraja, Lindley distribution as frailty models with application to lifetime data, Advances and Applications in Statistics 75 (2022), 119-134. http://dx.doi.org/10.17654/0972361722031

Published Online: March 29, 2022

Abstract

Unobserved heterogeneity is called frailty, measuring frailty and multiplying it with the baseline distribution is critical for clustered survival data analysis. Lindley distribution is the one among classical distribution, yet it has limited applications in lifetime data analysis. Therefore, the objective of the study is to fit the frailty models for Lindley distribution and to compare the results with other existing distributions such as Exponential, Weibull, Lognormal and Log-logistic to test the effectiveness. Two real-life data sets and simulated data were used to fit the baseline distributions with frailty models. The study results revealed that Lindley with Gamma frailty model is a good choice for kidney infection data and Lindley with Inverse Gaussian frailty model is the best fit for CGD (Chronic Granulomatous Disease) and the simulated data set. Further, Lindley with frailty models points out the lowest Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) values than other baseline distributions. So we suggest that Lindley baseline distribution with the frailty models is a potential alternative approach for clustered survival data analysis.

1. Introduction

Survival (or time to event, lifetime) analysis plays a vital role in the fields of medicine, biology, epidemiology, and life sciences [1]. Unobserved random effect shared by subjects is called *frailty*, and it has a significant effect on an individual's survival and hazard function. Therefore, estimating and including spontaneous impact (heterogeneity) to the models is crucial in the clustered survival data analysis [2]. The frailty term was introduced by Vaupel et al. [3] and was used in univariate analysis. This technique was applied to multivariate survival analysis by Clayton [4]. The frailty effect is multiplied with the baseline hazard function and assumed that it follows parametric distributions such as Gamma (Ga), Lognormal (LG), Positive Stable (PS), Inverse Gaussian (IG) and Power variance function (PVF) family [5]. Mostly Exponential, Weibull, Lognormal distributions are most commonly used as baseline distributions [6]. To arrive at a robust estimation,

we must choose the best baseline and frailty distribution depending on the data structure [5]. Lindley distribution is one of the classical distributions introduced by Lindley in [7], which gets importance for the different shapes of the hazard function [8]. It is often used in the field of reliability and is rarely applied for survival analysis [9]. The aim of the article is to fit the frailty models with Lindley baseline distribution and further, we have compared other popular baseline distributions for the same frailty models for identifying the best fit model. This paper is organized as follows. Section 2 deals with the basic properties of Lindley distribution, Section 3 discusses frailty models and Section 4 shows the applications of Lindley distribution with frailty models for real-life and simulation study data set. Finally, concluding remarks are given in Section 5.

2. Lindley Distribution

Let us consider that the non-negative continuous random variable T denotes the time to event of interest and follows Lindley distribution with scale parameter λ of a particular population. The probability density function (p.d.f.) and cumulative density function (c.d.f.) are given by equations (1) and (2), respectively:

(p.d.f.)

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

(c.d.f.)

$$F(t) = 1 - \frac{(\lambda + \lambda t + 1)}{(\lambda+1)} e^{-\lambda t}; \quad \lambda > 0, t > 0. \tag{2}$$

The survival rate $(S(t))$ is known as the probability of failure at time T $(S(t) = 1 - F(t))$; therefore, survival rate function is given by equation (3)

$$S(t) = \frac{(\lambda + \lambda t + 1)}{(\lambda+1)} e^{-\lambda t}. \tag{3}$$

The hazard rate function of T is denoted as $(h(t))$ and is given by equation (4)

$$h(t) = \frac{f(t)}{s(t)},$$

$$h(t) = \frac{\lambda^2(1+t)}{(\lambda + \lambda t + 1)}, \tag{4}$$

where $h(0) = \frac{\lambda^2}{(\lambda+1)}$ and $h(t)$ is an increasing function of "t" and "λ" and $\lambda^2/(\lambda+1) < h(t) < \lambda.$ The hazard function can be represented in the term of the cumulative hazard function is given by equation (5)

$$H(t) = \int_0^t h(t)dt = -\log[1 - F(t)] = -\log(S(t)), \tag{5}$$

where

$$h(t) = -\left(\frac{d \log S(t)}{dt}\right); \quad f(t) = h(t)e^{-H(t)};$$

$$S(t) = e^{-H(t)} = -\log\left(\frac{(\lambda + \lambda t + 1)}{(\lambda+1)} e^{-\lambda t}\right).$$

Solving the equation (5), we get

$$H(t) = \lambda t + \log(\lambda + 1) - \log(\lambda + \lambda t + 1).$$

Simplifying further,

$$H(t) = \lambda t + \log\left(\frac{(\lambda+1)}{(\lambda + \lambda t + 1)}\right).$$

3. Frailty Models

In frailty models, random effects are assumed to represent different clusters, and clusters are considered to be independent [10] and assume proportional hazards structure conditional on the random effect, "Z" [11]. Let

random effect "*Z*" be a non-negative frailty variable that indicates the individual-level risk of the study population. The conditional hazard then represents the frailty model as

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

Here, $j \in J = \{1, 2, ..., n\}$ is a subject and $i \in I = \{1, 2, ..., G\}$ is a group, where $h_0(t)$ is the common baseline hazard function for all subjects. The Z_i is the risk factor that is common for all subjects in the cluster group "*i*" and also assumed to be independently and identically distributed (IID) random variables with a common density function $f(z, \theta)$. Here, θ is the parameter of the frailty distribution. The factor $\exp(x_{ij}^T \beta)$ gives the subject-specific contribution to the hazard. x_{ij} is the covariates vector for the subject *j* in group *i*, and β is the regression coefficient vector.

Table 1. Hazard, cumulative hazard, and survival functions for parametric baseline distributions

Baseline distributions (Parameters)	**Hazard function** $(h_0(t))$	**Cumulative hazard function** $(H_0(t))$	**Survival functions** $(S_0(t))$
Lindley $(\lambda > 0)$	$\frac{\lambda^2(1+t)}{(\lambda + \lambda t + 1)}$	$\lambda t + \log\left(\frac{(\lambda+1)}{(\lambda+\lambda t+1)}\right)$	$\frac{(\lambda+\lambda t+1)}{(\lambda+1)} e^{-\lambda t}$
Exponential $(\lambda > 0)$	λ	λt	$\exp(-\lambda t)$
Weibull $(\rho, \lambda > 0)$	$\lambda \rho t^{\rho-1}$	λt^{ρ}	$\exp(-\lambda t^{\rho})$
Lognormal $(\mu \in \mathbb{R}, \sigma > 0)$	$= \frac{\phi\left(\frac{\log(t)-\mu}{\sigma}\right)}{\sigma t\left[1-\Phi\left(\frac{\log(t)-\mu}{\sigma}\right)\right]}$	$-\log\left[1-\Phi\left(\frac{\log(t)-\mu}{\sigma}\right)\right]$	$1-\Phi\left(\frac{\log(t)-\mu}{\sigma}\right)$
Loglogistic $(\alpha \in \mathbb{R}, \kappa > 0)$	$\frac{\exp(\alpha) k t^{k-1}}{[1+\exp(\alpha) t^k]}$	$\log[1+\exp(\alpha) t^k]$	$\frac{1}{[1+\exp(\alpha) t^k]}$

3.1. Baseline distributions and estimation of frailty

Four parametric frailty models, namely Gamma (Ga), Inverse Gaussian (IG), Lognormal (LN), and Positive stable (PS) were used to fit and compared with Lindley and other baseline distributions. Table 1 shows hazard, cumulative hazard, and survival functions for each baseline

distribution such as Lindley, Exponential, Weibull, Lognormal and Log-logistic. The frailty distributions mentioned above and their properties are well documented in previous studies [11-14]. Hence, the p.d.f. Laplace transformation (LS) and estimation of frailty for each distribution are summarized in Table 2.

3.2. Marginal log-likelihood

The marginal likelihood function is driven based on the assumption of conditional independence of lifespans given the frailty [11]. For parametric frailty models, the frailties appearing in the conditional likelihood can be integrated out to maximize the marginal likelihood, leading to estimates of the model parameters [14, 15]. In right-censored cluster survival data, the marginal log-likelihood of the observed data is under assumptions of (i) non-informative right-censoring (ii) independence between the censoring time and the survival time random variables (iii) given the covariate information (iv) the marginal log-likelihood of the observed data $U = \{U_{ij}; i \in I,\ j \in J_i\}$ [16].

For right-censored clustered survival data, the observation for subject $j \in J_i = 1, 2, ..., n_i$ from cluster $i \in I = \{1, 2, ..., G\}$ is the couple $U_{ij} = (y_{ij}, \delta_{ij})$, where $y_{ij} = \min(t_{ij}, c_{ij})$ is the minimum between the survival time t_{ij} and the censoring time c_{ij}, and where $\delta_{ij} = I(t_{ij} \le c_{ij})$ is the event indicator. Covariate information may also be collected; in this case, $U_{ij} = (y_{ij}, \delta_{ij}, x_{ij})$, where x_{ij} denotes the vector of covariates for the ijth observation. Further, if left-truncation is also present, truncation times T_{ij} are gathered in the vector:

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

where

$$d_i = \sum_{j=1}^{ni} \delta_{ij}$$

the number of events in the *i*th cluster.

3.3. Laplace transform

The Laplace transform to characterize the density functions of the frailty distribution, and unconditional survival and hazard functions can be easily expressed. Hence, the likelihood function can also be represented through the Laplace transform. Therefore, frailty distributions with easy Laplace transforms are essential; they allow for traditional maximum likelihood methods in parameter estimation [11, 17]. $L^{(q)}(\cdot)$, the *q*th derivative of the Laplace transform [13, 14] of the frailty distribution defined as

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

Higher-order derivatives $L^{(q)}(\cdot)$ of the Laplace transform up to $q = \max\{d1, ..., dG\}$. Hence, *q*th derivate is given by equation (6)

$$L^{(q)}(s) = (-1)^{(q)} E(Z^{(q)} \exp(-zs)). \tag{6}$$

3.4. Prediction

The EM algorithm is a combination of an expectation, and a maximization step, and this method was used to predict the frailties [18]. The frailty z_i is predicted by $z_i = E(Z \mid u_i, \tau_i; \psi, \beta, \xi)$, where u_i and τ_i are the data and the truncation times of the *i*th cluster. Therefore, conditional expectation becomes

$$E(Z \mid u_i, \tau_i; \psi, \beta, \xi) = -\frac{L^{(d_i+1)}\left[\sum_{j=1}^{ni} H_0(y_{ij}) \exp(x_{ij}^T \beta)\right]}{L^{d_i} \cdot \left[\sum_{j=1}^{ni} H_0(y_{ij}) \exp(x_{ij}^T \beta)\right]}.$$

Table 2. Probability density function (p.d.f), Laplace transform $(L(s))$ and estimation of frailty for parametric frailty distributions

Frailty distribution	**Probability density function (p.d.f)**	**Laplace transform for frailty** $L(s) = E[\exp(-Zs)]$	**Estimation of frailty** $\log(E[Z^q \exp(-Zs)]) = \log(-1)^q . L^{(q)}(s)$
Gamma frailty (G) $(Ga * \theta)$	$f(z) = \frac{\theta^{-1/\theta} Z^{\frac{1}{\theta}-1} \exp\left(-\frac{z}{\theta}\right)}{\Gamma(1/\theta)}$, $\theta > 0$	$(1+\theta s)^{-1/\theta}$ $s \geq 0$	$-(q + 1/\theta).\log(1+\theta s) + \sum_{l=0}^{q-1} \log(1 + l\theta)$
Positive Stable (PS) $(Ps * \nu)$	One parameter; $Gam(\mu, \theta)$: mean $(\mu) = 1$, variance $= \theta$ $f(z) = -\frac{1}{\pi Z} \sum_{k=1}^{\infty} \frac{\Gamma(k(1-\nu)+1)}{k!} \cdot(-z^{\nu-1})^k \sin((1-\nu)k\pi)$, $\nu \in (0, 1)$ Two parameters (δ, α), "ν" Not correspond to the variance Undefined mean and variance, so "ν" used instead of "θ"	$\exp(-s^{1-\nu})$, $s \geq 0$	$q(\log(1-\nu) - \nu.\log(s)) + \log\left[\sum_{m=0}^{q-1} \Omega_{q,m} s^{-m(1-\nu)}\right] - s^{1-\nu}$ $\Omega_{q,m} S$ are polynomials of degree m
Inverse Gaussian (IG) $(IG * \theta)$	$f(z) = \frac{1}{\sqrt{2\pi\theta}} Z^{-3/2} \cdot \exp\left(-\frac{(z-1)^2}{2\theta z}\right)$, $\theta > 0$ One parameter; $Gam(\mu, \theta)$: mean $(\mu) = 1$, variance $= \theta$	$\exp\left(\frac{1}{\theta}\left(1 - \sqrt{1+2\theta s}\right)\right)$, $s \geq 0$	$-\frac{q}{2}\log(2\theta s + 1) + \log(K_{q-\frac{1}{2}}(Z)) - \left[\frac{1}{2}\left(\log\left(\frac{\pi}{2z}\right)\right) - z\right] + \frac{1}{\theta}(1 - \sqrt{1+2\theta s})$ where $z = \sqrt{2\theta^{-1}\left(s + \frac{1}{2\theta}\right)}$
Lognormal (LN) $(LN * \theta)$	$f(z) = \frac{1}{z\sqrt{2\pi\theta}} \cdot \exp\left(-\frac{(\log z)^2}{2\theta}\right)$, $\theta > 0$ One parameter; $LN(\mu, \theta)$: mean $(\mu) = 0$, variance $= \theta$	For a lognormal frailty distribution no explicit evaluation of the Laplace transform is possible and also Kendall's τ no explicit formula exists (Duchateau and Janssen [12]). Hence we need Laplace approximation $L^{(q)}(s)$ (Marco et al. [14]) $(-1)^q \frac{1}{\sqrt{\theta}} \exp\{-g(w; s, \theta)\} \cdot [g^2(w; s, \theta)]^{-1/2}$	$\log\left[(-1)^q \frac{1}{\sqrt{\theta}} \exp\{-g(\breve{w}; s, \theta)\} \cdot [g^2(\breve{w}; s, \theta)]^{-1/2}\right]$ where mean $= \breve{w}$ and variance $= 1/g^2(\breve{w}; s, \theta)$

$L^{(q)}(\cdot)$: The qth derivative of the Laplace transform of the frailty distribution

4. Application to Real-life Data

Application I

For comparison of models, first, we used the kidney infection data set [19] to fit the four frailty models for five baseline distributions. The dataset contains the data of the first and second recurrence times (in days) of infection from the time of insertion of the catheter until it has to be removed. The measurements were recorded for 38 patients using portable dialysis equipment, totaling 76 observations (clusters). The dataset included five variables namely; Recurrence times, Indicator (0 = censored (catheter may have to be removed for reasons other than kidney infection), 1 = recurrence) with covariates of age, sex and diseases type.

Application II

Second, we used a "CGD" data set (i.e., Placebo-controlled randomized trial data of gamma interferon (γ-IFN) in chronic Granulomatous disease) to compare the models [20]. The study investigates the effectiveness of the γ-IFN in reducing the rate of serious infections in CGD patients. The data set contains survival times between recurrent CGD infections (gap times) of 135 patients (203 clusters) with the status censored (0) or not (1) and ten covariates of Treatment (0 = Placebo, 1 = γ-IFN), sex (0 = Male, 1 = Female), age (in years), height (in cm), weight (in kg), pattern of inheritance (0 = autosomal recessive, 1 = X-linked), Corticosteroids used at the entry time (0 = No, 1 = Yes), prophylactic antibiotics used at the time of entry (0 = No, 1 = Yes), hospital region (0 = U.S., 1 = Europe) and longitudinal years (accumulated time from the first infection in years).

Simulation study

A simulation study has been executed in a setting similar to a clustered survival data set structure including the covariate, survival time and censoring. A large dataset was simulated with a single covariate (X_1) from a

Binomial distribution $B[n,\ p = 0.5]$ with arbitrary parameter setting fixed throughout the entire study. The corresponding true regression coefficient is fixed as $\beta_1 = 1$. Survival time was randomly generated between 3 to 120 months considering several lifetime survival sets and we spawned a random censure following a uniform distribution on the interval from $U[0, 9]$. The frailty variables Z_i is assumed to follow any one of four frailty models (Gamma (Ga), Inverse Gaussian (IG), Lognormal (LN), and Positive stable (PS)). The simulated data contains the 1000 number of observations, i.e., 10 clusters and 100 individuals in each cluster $(n = 1000)$ and it was replicated 1000 times.

4.1. Data analysis

R packages of "Survival" [21], "parfm" [14], "frailtyEM" [15] and "frailtypack" [22] were used to create the codes/function for Lindley and other distributions. R studio version 1.2.50 was used for data analysis. The method of Kendall's tau was used to measure the association between any two event times from the same cluster [23]. Akaike's Information Criteria ((AIC = –2(loglikelihood) +2P); where P is the number of parameters) and Bayesian Information Criteria (BIC = –2(loglikelihood) + P(log/n)) were used to assess the model fitness.

4.2. Results

Comparison of four frailty models under five baseline distributions for kidney infection data is shown in Table 3. 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 Lindley baseline with the Gamma frailty model (Figures 1-2).

Table 3. Comparison of frailty models under Lindley and other baseline distribution for kidney infection data

Baseline/Frailty distribution	Parameters/ Covariates	Gamma (Ga)	Inverse Gaussian (IG)	Lognormal (LN)	Positive Stable (PS)
		Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Lindley	Frailty	0.548 (0.208)	1.058 (0.571)	0.752 (0.705)	0.291 (0.07)
	Λ	0.076 (0.029)	0.075 (0.032)	0.064 (0.031)	0.053 (0.025)
	Sex	-2.143 (0.408)^	-1.776 (0.396)^	-1.985 (0.697)*	-1.465 (0.408)^
	Age	0.003 (0.012)	-0.001 (0.012)	-0.001 (0.012)	-0.007 (-0.007)
	AIC	672.155	674.862	674.194	682.366
	BIC	681.478	684.185	683.517	691.689
	Kendall's Tau	0.215	0.229	0.220	0.293
Exponential	Frailty	0.301 (0.156)	0.375 (0.259)	0.342 (0.197)	0.112 (0.084)
	Λ	0.025 (0.014)	0.022 (0.013)	0.020 (0.011)	0.014 (0.008)
	Sex	-1.485 (0.396)^	-1.31 (0.373)^	-1.356 (0.382)^	-0.951 (0.348)*
	Age	0.005 (0.011)	0.004 (0.011)	0.005 (0.011)	0.004 (0.011)
	AIC	674.496	675.699	682.264	675.212
	BIC	683.819	685.022	691.587	684.535
	Kendall's Tau	0.131	0.125	0.139	0.112
Weibull	Frailty	0.510 (0.255)	0.677 (0.537)	0.589 (0.340)	0.139 0.134
	P	1.216 (0.152)	1.145 (0.141)	1.177 (0.141)	1.039 0.154
	Λ	0.013 (0.009)	0.013 (0.010)	0.010 (0.007)	0.011 0.009
	Sex	-1.912 (0.539)^	-1.481 (0.431)^	-1.626 (0.488)^	-0.973 (0.378)*
	Age	0.007 (0.012)	0.006 (0.012)	0.006 (0.011)	0.005 (0.011)
	AIC	674.376	676.627	682.315	675.726
	BIC	686.029	688.281	693.969	687.379
	Kendall's Tau	0.203	0.181	0.208	0.139
Log logistic	Frailty	0.106 (0.167)	0.099 (0.192)	0.191 (0.271)	0.001 (0.156)
	A	-5.845 (0.766)	-5.802 (0.752)	-5.970 (0.795)	-5.568 (1.004)
	K	1.489 (0.292)	1.476 (0.287)	1.469 (0.267)	1.428 (0.277)
	Sex	-1.006 (0.390)*	-0.967 (0.360) *	-1.049 (0.388)*	-0.843 (0.292)*
	Age	0.012 (0.012)	0.012 (0.011)	0.015 (0.013)	0.008 (0.011)
	AIC	685.184	685.304	685.699	684.818
	BIC	696.837	696.958	697.353	696.472
	Kendall's Tau	0.050	0.043	0.085	0.010
Lognormal	Frailty	0.999 (1.682)	1.000 (1.889)	1.000 (1.897)	0.500 (0.073)
	M	2.124 (0.021)	2.125 (0.026)	2.152 (0.133)	2.121 (0.018)
	Σ	0.561 (0.026)	0.557 (0.026)	0.589 (0.025)	0.554 (0.024)
	Sex	-1.742 (0.475)^	-1.742 (0.381)^	-1.742 (0.384)^	-1.742 (0.545)*
	Age	-0.018 (0.009)	-0.018 (0.012)	-0.023 (0.008)*	-0.022 (0.010)#
	AIC	678.849	679.196	680.467	678.882
	BIC	690.503	690.85	692.121	690.536
	Kendall's Tau	0.333	0.233	0.290	0.500

Significant at #5% level ($P < 0.05$); * 1% level ($P < 0.005$); ^0.1% level ($P < 0.001$)

Further among other frailty models, Lindley baseline distribution showed better results than other baseline distributions (Table 3 and Figures 1-2). The frailty models of Gamma, Inverse Gaussian and Lognormal with Lindley baseline have given almost close estimation values for this data. We noticed that the baseline distribution of log-logistic with frailty models is the least preferable option for kidney infection data because of high AIC and BIC values for all frailty models (Figures 1-2).

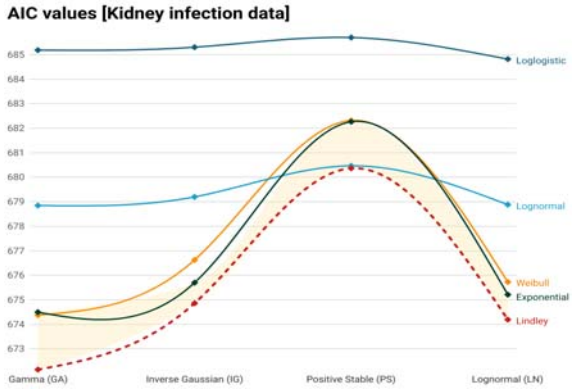

Figure 1. Comparison of AIC values for kidney infection data.

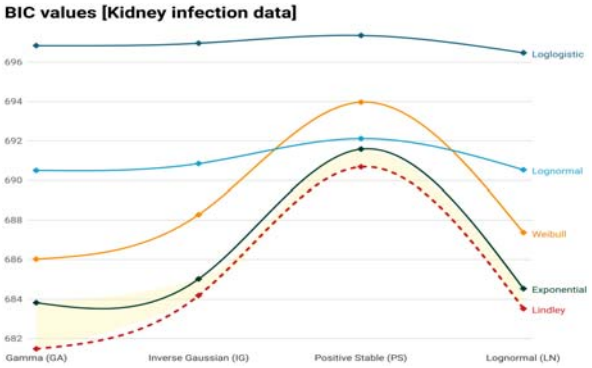

Figure 2. Comparison of BIC values for kidney infection data.

Lindley, Exponential and Weibull baseline distributions with the Inverse Gaussian (IG) have given almost close estimation with smaller AIC and BIC values for the CGD data set (Table 4). However, Lindley with the Inverse Gaussian frailty model is the best choice for this data because of the lowest

AIC and BIC values. Moreover, we noticed Lindley with the lognormal frailty model also gave close estimation values for this data. Further within each of the frailty models, Lindley baseline distribution shows better estimates than other baseline distributions (Table 4 and Figures 3-4). In this case, we noticed that the baseline distributions of Log logistic and Lognormal were not good choices due to high estimation values observed for all frailty models for CGD data (Figures 3-4). In a similar way, four frailty models were compared for simulated data. The Inverse Gaussian frailty model with Lindley baseline is best among all frailty models due to recording the lowest AIC and BIC values for simulated data, given by Table 5.

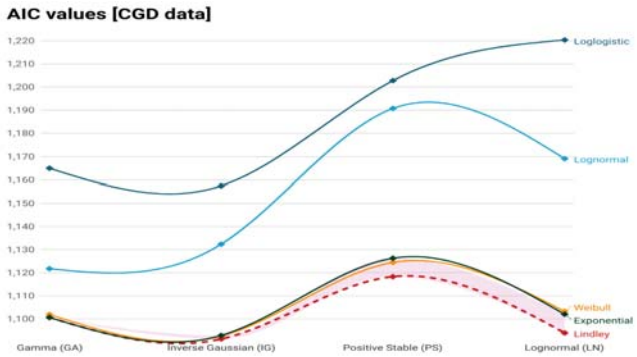

Figure 3. Comparison of AIC values for CGD data.

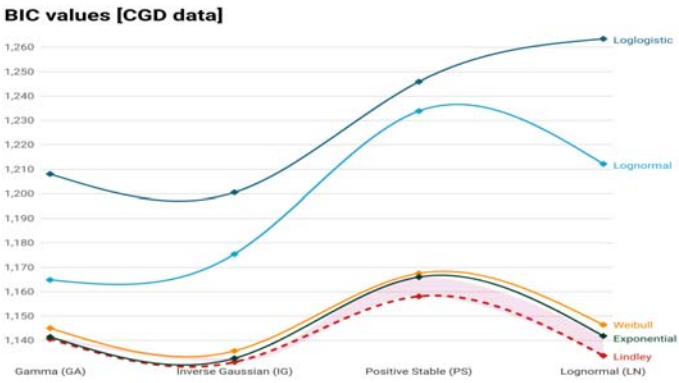

Figure 4. Comparison of BIC values for CGD data.

Table 4. Models wise AIC and BIC values comparison for CGD data

Baseline distribution	Akaike's and Bayesian Information Criteria	Frailty models			
		Gamma (GA)	Inverse Gaussian (IG)	Positive Stable (PS)	Lognormal (LN)
Lindley	AIC	1100.841	1091.376	1118.325	1093.991
	BIC	1140.600	1131.135	1158.084	1133.749
Exponential	AIC	1100.700	1092.913	1126.282	1102.076
	BIC	1141.448	1132.672	1166.04	1141.835
Weibull	AIC	1101.990	1092.639	1124.477	1103.340
	BIC	1145.061	1135.711	1167.548	1146.411
Loglogistic	AIC	1164.963	1157.465	1202.817	1220.386
	BIC	1208.035	1200.536	1245.889	1263.458
Lognormal	AIC	1164.843	1175.433	1233.841	1212.143
	BIC	1121.771	1132.362	1190.769	1169.071

Table 5. Comparison of frailty models under Lindley distribution for simulated data

Parameter/ Covariate	Gamma (Ga)		Inverse Gaussian (IG)		Positive Stable (PS)		Lognormal (LN)	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Frailty	0.149	0.084	0.166	0.1	0.118	0.056	0.152	0.09
λ	0.024	0.004	0.021	0.004	0.028	0.004	0.023	0.003
Treatment	–0.417*	0.141	–0.412*	0.136	–0.424*	0.143	–0.414*	0.138
AIC	1590		1580.11		1594.98		1587.24	
BIC	1611.32		1606.84		1617.71		1608.96	
Kendall's Tau	0.072		0.066		0.118		0.067	

*Significant at 0.1% level ($P < 0.001$)

5. Conclusion

In practice, exponential and the Weibull baseline distributions are widely used with frailty models in survival analysis. Identifying and applying both baseline and frailty distributions based on the data structure are essential to the model estimation. This paper attempted to fit the Lindley baseline distribution with four frailty models and identify the best model; simultaneously, we compared the results for most commonly used baseline distributions with the same frailty models for real-life data. We proved that

Lindley with frailty models has a good fit with the smallest AIC and BIC values than other baseline distributions for the real-life data applications. The study will help construct the new frailty models for Lindley and other baseline distributions and that may be used for the future applications of the Lindley distributions.

Acknowledgement

The authors thank the anonymous referees for their valuable suggestions and constructive criticisms which improved the presentation of the paper.

References

[1] T. G. Clark, M. J. Bradburn, S. B. Love and D. G. Altman, Survival analysis part I: basic concepts and first analyses, British Journal of Cancer 89(2) (2003), 232-38. https://doi.org/10.1038/sj.bjc.6601118.

[2] Theodor A. Balan and Hein Putter, A tutorial on frailty models, Statistical Methods in Medical Research 29(11) (2020), 3424-3454. https://doi.org/10.1177/0962280220921889.

[3] J. W. Vaupel, K. G. Manton and E. Stallard, The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography 16 (1979), 439-454.

[4] D. Clayton and J. Cuzick, Multivariate generalisations of the proportional hazards model, Journal of the Royal Statistical Society 148 (1985), 82-117.

[5] David D. Hanagal and Alok D. Dabade, Comparisons of frailty models for kidney infection data under Weibull baseline distribution, International Journal of Mathematical Modelling and Numerical Optimisation 5(4) (2014), 342-373.

[6] D. R. Cox and D. Oakes, Analysis of Survival Data, Chapman and Hall, 1984.

[7] D. V. Lindley, Fiducial distributions and Bayes theorem, Journal of the Royal Statistical Society A 20(1) (1958), 102-107.

[8] L. Tomy, A retrospective study on Lindley distribution, Biom. Biostat. Int. J. 7(3) (2018), 163-169. https://doi.org/10.15406/bbij.2018.07.00205.

[9] M. Okwuokenye and K. E. Peace, Size and power of tests of hypotheses on survival parameters from the Lindley distribution with covariates, Austin Biom and Biostat. 2(4) (2015), 1026.

[10] S. Banerjee, M. M. Wall and B. P. Carlin, Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota, Biostatistics 4(1) (2003), 123-142. https://doi.org/10.1093/biostatistics/4.1.123.

[11] A. Wienke, Frailty Models in Survival Analysis, Chapman & Hall/CRC, Boca Raton, 2010.

[12] L. Duchateau and P. Janssen, 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.

[13] David D. Hanagal, Modeling Survival Data Using Frailty Models, Industrial and Applied Mathematics, Springer Nature Singapore Pte. Ltd. 2011, p. 295. https://doi.org/10.1007/978-981-15-1181-3.

[14] Marco Munda, Federico Rotolo and Catherine Legrand, parfm: Parametric Frailty Models in R, Journal of Statistical Software, Foundation for Open Access Statistics 51(11) (2012), 1-20. http://hdl.handle.net/10.18637/jss.v051.i11.

[15] T. A. Balan, H. Putter and E. M. Frailty, An R package for estimating semi parametric shared frailty models, Journal of Statistical Software 90(7) (2019), 1-29. https://doi.org/10.18637/jss.v090.i07.

[16] Dilip C. Nath, Atanu Bhattacharjee and Ramesh K. Vishwakarma, Risk assessment in liver transplantation patients: A shared frailty parametric approach, Clinical Epidemiology and Global Health 4(1) (2016), 1-15.

[17] P. Hougaard, Frailty models for survival data, Lifetime Data Analysis 1 (1995), 255-274.

[18] G. Guo and G. Rodriguez, Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala, J. Am. Statist. Assoc. 87 (1992), 969-976.

[19] C. A. McGilchrist and C. W. Aisbett, Regression with frailty in survival analysis, Biometrics 47(2) (1991), 461-466.

[20] T. R. Fleming and D. P. Harrington, Counting Processes and Survival Analysis, New York, Wiley, 1991.

[21] T. M. Therneau, Survival: A Package for Survival Analysis in S. R package version 2.44-1.1, 2019. URL https://CRAN.R-project.org/package=survival.

[22] V. Rondeau, Y. Mazroui and J. R. Gonzalez, 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 47(4) (2012), 1-28. URL http: //www.jstatsoft.org/v47/i04/.

[23] P. Hougaard, Analysis of Multivariate Survival Data, Springer-Verlag, 2000.