Mathematics and Computer Science
2023; 8(4): 87-93
http://www.sciencepublishinggroup.com/j/mcs
doi: 10.11648/j.mcs.20230804.11
ISSN: 2575-6036 (Print); ISSN: 2575-6028 (Online)

Frailty Models Under Xgamma Distribution with Application to Survival Data

Ashok Kumar Palanisamy*, Muthukumar Madaswamy

Department of Statistics, PSG College of Arts and Science, Coimbatore, India

Email address:
ashokkumar.stat2023@gmail.com (Ashok Kumar Palanisamy), muthukumar@psgcas.ac.in (Muthukumar Madaswamy)

*Corresponding author

To cite this article:
Ashok Kumar Palanisamy, Muthukumar Madaswamy. Frailty Models Under Xgamma Distribution with Application to Survival Data. *Mathematics and Computer Science*. Vol. 8, No. 4, 2023, pp. 87-93. doi: 10.11648/j.mcs.20230804.11

Received: July 26, 2023; **Accepted**: August 14, 2023; **Published**: August 31, 2023

Abstract: Frailty models provide an alternative to proportional hazards models, which are designed to discover the properties of the unobserved heterogeneity in individual risks of disease and death. In spite of this distribution of the frailty is normally assumed to be continuous. In some circumstances, it is appropriate to recollect discrete frailty distributions. Generally, Gamma, Weibull, Exponential, Lognormal, and Log-logistic baseline distributions have fitted with frailty distribution. The Xgamma distribution among a unique finite aggregate of exponential and gamma distribution and allowance for the different shapes of the hazard function. The study aims to fit the above four distributions with the Xgamma baseline distribution and apply them to test popular actual-lifestyles statistics set. The study result revealed that Xgamma with Positive Stable (PS) frailty model is a good choice for the Veterans' Administration Lung Cancer study data set and Xgamma with Log-Normal (LN) frailty model is the best fit for the Culling dairy heifer cow's data set. Additionally, Xgamma identifies the baseline distribution with the lowest Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) values. The study result proved Xgamma distribution and its extended model for frailty distribution is the possible approach in a real-life time or survival analysis.

Keywords: Xgamma Distribution, Hazard Function, Survival Analysis, Parametric Frailty Models, Marginal LOG-Likelihood, Clustered Data Analysis

1. Introduction

Survival analysis is the study of time-to-event data, estimating the time elapsed from a given starting time to the prevalence of an event of interest. The study of survival data is prevalent in the medical field. However, the researcher can't usually examine the event due to censoring [1]. The event may be considered the time of death, recurrence or recovery from illness, and the onset (or) time of disease progression [2]. Furthermore, the learn population can be separated into clusters so that topics in the same cluster behave more cohesively than topics in different clusters [3].

Frailty models are becoming increasingly popular as a way to account for over-dispersion and/or clustering in survival data [4]. It is an extension of the proportional hazard model in which the hazard function is dependent on an unobserved random quantity, the so-called frailty, which operates multiplicatively on it [5-8]. The estimation of the frailty model can be parametric or semi-parametric including Gamma, Exponential, Lognormal, and Positive stable and Inverse Gaussian distribution [9]. In recent years, Lindley [10] and different baseline distributions have drawn the attention of researchers and practitioners in modeling time-to-event statistics units [11]. Similarly, Xgamma distribution is a mixture of exponential and gamma distribution with mixing proportion and was introduced by Subhradev Sen et. al. (2016) [12-13].

Therefore, in this article, we proposed a frailty model for Xgamma distribution and used them in two real-life data sets to compare the models. This is organized as follows. Section 2 deals with the properties of the Xgamma distribution. Section 3 discusses the frailty model and various frailty distributions with probability density function (PDF), Laplace transform (L (s)), and estimation of frailty distribution. Section 4, shows the applications of Xgamma distribution with frailty model for two real-life data sets and data analysis were applied in section 5. Finally, concluding remarks are given in section 6.

2. Xgamma Distribution

A continuous random variable X is called an Xgamma variant ie., $x \sim gamma\ (\theta)$, and its pdf is defined as

$$f(x) = \frac{\theta^2}{(1+\theta)}\left(1 + \frac{\theta}{2}x^2\right)e^{-\theta x}, x > 0, \theta > 0 \qquad (1)$$

The corresponding cumulative density function (cdf) is given by

$$F(x) = 1 - \frac{(1+\theta+\theta x+\frac{\theta^2}{2}x^2)}{(1+\theta)}e^{-\theta x}, x > 0, \theta > 0 \qquad (2)$$

The Survival Properties of Xgamma distribution are defined by

$$S_x(t)=1-F_x(t),$$

therefore

$$S(x) = \frac{\left(1+\theta+\theta x+\frac{\theta^2}{2}x^2\right)}{(1+\theta)}e^{-\theta x}, forgiven\ \theta \qquad (3)$$

The failure (hazard) rate function for continuous distribution with probability density function f(x), cumulative density function F(x), and survival function S(x) is defined as

$$h(x) = \lim_{\Delta x \to 0} \frac{P(X<x+\frac{\Delta x}{X}>x)}{\Delta x} = \frac{f(x)}{S(x)}$$

$$h(x) = \frac{\theta^2\left(1+\frac{\theta x^2}{2}\right)}{(1+\theta+\theta x+\frac{\theta^2}{2}x^2)}, x > 0 \qquad (4)$$

where $h(0) = \frac{\theta^2}{(1+\theta)} = f(0)$ and h (x) is an increasing function in x and θ with ${\theta^2}/{(1+\theta)} < h(x) < \theta$

The hazard function can be represented as the cumulative hazard function. Therefore, the cumulative hazard function of Xgamma distribution is

$$H(x) = \int_0^x h(x)dx = -\log\big(S(t)\big); where\ h(x) = -\left(\frac{dlogs(x)}{dx}\right)$$

$$H(x) = -logS(x) = -\log\left[\frac{(1+\theta+\theta t+\left(\frac{\theta^2x^2}{2}\right))}{(1+\theta)}\ e^{-\theta x}\right]$$

Solving this equation, we get

$$H(x) = \theta x + \log(1+\theta) - (1+\theta+\theta t+\left(\frac{\theta^2x^2}{2}\right))$$

To simplify further we get

$$H(x) = \theta x + \log\left(\frac{(1+\theta)}{\left(1+\theta+\theta x+\left(\frac{\theta^2x^2}{2}\right)\right)}\right) \qquad (5)$$

3. Frailty Models

In the proportional hazard model, the response variable is the 'Hazard'. The hazard is the chance of death given that patients have survived up to a given factor in time, or the risk for death at the second [2]. Most commonly, survival information is treated via the skill of the proportional hazard regression model [14]. However, the proper information based on such proportional hazard models requires a sample distribution that is independent and identical. However, in reality, subjects may expose to different risk levels. So, frailty, random effect, or unknown heterogeneity shout be multiple with baseline hazard. Hence frailties were essential considered in the analysis for more accurate results than normal survival analysis [7]. The frailty model is characterised in terms of conditional hazard as

$$h_{ij}(x/w_i) = h_0(x)w_i \exp\big(z_{ij}^T\beta\big) \qquad (6)$$

Here, $i \in I = \{1,2 \dots \dots M\}$ and $j \in J_i = \{1,2 \dots \dots n_i\}$, where $h_0(x)$- is the baseline hazard function, w_i the frailty time in group i, z_{ij} the vector of covariates for concern j in group i, and β the vector of regression coefficients. For the parametric approach, the Xgamma baseline hazard function $h_0(x)$(parametric function) and its parameter(θ) are estimated together with regression coefficient (β) and frailty (y_i) parameters.

3.1. Marginal Log-Likelihood and Laplace Transform

In the parameter established, estimation is based on the marginal likelihood, with frailties built-in through common conditional probability with respect to the frailty distribution. Under the assumption of non-informative right-censoring and arbitrary variables for the censoring and survival times, given the covariate information, the marginal log-likelihood of the observed records $w_{ij} = \big(y_{ij,}\delta_{ij}, x_{ij}\big)$ [15-17].

The statement of difficulty $j \in J_I = \{1,2,3, \dots n_i\}$ from cluster $i \in I = \{1,2,3, \dots M\}$ is couple $Z_{ij} = \big(y_{ij,}\delta_{ij}\big)$, where $y_{ij,} = \min(t_{ij}, c_{ij})$ is the minimum between the survival time t_{ij} and the censoring c_{ij}, and where $\delta_{ij} = I(t_{ij} \le c_{ij})$ is the event indicator. Covariate facts may also have been collected; in this instance, $Z_{ij} = \big(y_{ij,}\delta_{ij}, x_{ij}\big)$, where x_{ij} is the vector of covariance for the ij-th observation. Furthermore, if left-truncation is present, the truncation time T_{ij} is collected in the vector T.

$$L_{marg}(\delta, \beta, \xi;\ w/T) = \sum_{i=1}^{M}\Big\{\Big[\sum_{j=1}^{n_i}\delta_{ij}(\log\left(h_0\big(y_{ij}\big)\right) + z_{ij}^T\beta\Big] + \log\big[(-1)^{d_i}L^{d_i}\big(\sum_{j=1}^{ni}H_0\left(y_{ij}\right)\exp\big(z_{ij}^T\beta\big)\big)\big] - log\big[L\big(\sum_{j=1}^{ni}H_0\left(y_{ij}\right)\exp\big(z_{ij}^T\beta\big)\big)\big] \qquad (7)$$

Where $d_i = \sum_{j=1}^{ni}\delta_{ij}$ the number of events in the i-th cluster, and $L^{(q)}(.)$ the q-th derivative of the Laplace transformation [16] of the frailty distribution is described as

$$L(s) = E[ex\,p(-w_s)] = \int_0^\infty \exp(w_i\,s)f(w_i)dw_i, s \ge 0. \qquad (8)$$

3.2. Estimation and Prediction

To Estimates $\delta, \beta\ and\ \xi$ are obtained by using optimizing the log-likelihood; this can be done without difficulty if one is capable of computing higher order derivatives $L^{(q)}(.)$ of the Laplace transformation up to $q = max\{d1, d2 \dots . dM\}$. Hence q-th derivate is given by equation (8)

$$L^q(s) = (-1)^q E(w^q \exp(-ws)) \quad (9)$$

The frailty term w_i can be predicted by $\widehat{w_i} = E(W/z_i, T_I; \hat{\delta}, \hat{\beta}, \hat{\xi})$, where z_i and T_i are the data and truncation time of the i-th cluster, respectively [16, 17]. Therefore, conditional expectations become

$$E(W/z_i, T_I; \delta, \beta, \xi) = - \frac{L^{(d_i+1)}(\sum_{j=1}^{ni} H_0(h_0(x) y_i \exp(z_{ij}^T \beta))}{L^{(d_i)}(\sum_{j=1}^{ni} H_0(h_0(x) y_i \exp(z_{ij}^T \beta))}, \quad (10)$$

3.3. Gamma Frailty

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 it's denoted by $X \sim Gam*(\theta)$. where $\Gamma(.)$ denotes the gamma function. It relates to the 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

$$\log\left((-1)^q L^{(q)}(s)\right) = \left(q + \frac{1}{\theta}\right) \log(1 + \theta s) + \sum_{l=0}^{q-1} \log(1 + l\theta) \quad (11)$$

This analyses the close relationship between any two-event time from the same cluster [5] in the multivariate situation and can be computed as

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

3.4. Lognormal Frailty

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

$$f(x) = (2\pi\theta)^{-1/2} x^{-1} \exp\left\{\left[-\frac{(\log x)^2}{2\theta}\right]\right\}, \theta > 0 \quad (12)$$

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

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

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

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

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

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

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

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

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

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.

3.5. Inverse Gaussian Frailty

The density of the inverse Gaussian frailty distribution IG* (θ) is

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

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,

$$L^q(s) = (-1)^q (1 + 2\theta s)^{-\frac{q}{2}} \frac{K_{q-\left(\frac{1}{2}\right)}\left(\sqrt{2\theta^{-1}}(s + \frac{1}{2\theta})\right)}{K_{\left(\frac{1}{2}\right)}\left(\sqrt{2\theta^{-1}}(s + \frac{1}{2\theta})\right)} . L(s), \quad (13)$$

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 basic construction approach for obtaining the previously mentioned 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

$$log\left((-1)^q L^{(q)}(s)\right) = -\frac{q}{2}\log(2\theta s + 1) + \log\left(K_{q-\left(\frac{1}{2}\right)}(z)\right) - \left[\frac{1}{2}\left(l0g\left(\frac{\pi}{2z}\right)\right) - z\right] + \frac{1}{\theta}\left(1 - \sqrt{1 + 2\theta s}\right), \quad (14)$$

With $z = \sqrt{2\theta^{-1}}\left(s + \frac{1}{2\theta}\right)$, with multivariate data, an inverse Gaussian distribution frailty [5, 19] 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

The family of positive stable distributions with two parameters. A scale $\delta > 0$, as well as 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

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

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 Wang, Klein, and Moeschbereger (1995) found that, 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 polynomials of degree m, given recursively by

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

$$\Omega_{q,m} = \Omega_{q-1,m} + \Omega_{q-1,m-1}\left\{\frac{q-1}{1-\gamma} - (q-m)\right\}, m = 1,2,\ldots\ldots, q-2, \quad (15)$$

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

It follows that

$$log\left((-1)^q L^{(q)}(s) = q\left(log(1-\gamma) - \gamma log(s)\right) + log\left[\sum_{m=0}^{q-1}\Omega_{q,m}s^{-m(1-\gamma)}\right] - s^{1-\gamma} \quad (16)$$

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

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

4. Application to Real-Life Data

Application I: We consider the Veterans' Administration Lung Cancer data set [20-21] to fit the frailty model for the Xgamma distribution. The data set contains the Lung cancer data of the first and second recurrence time of 130 patients, 4 different clusters, and eight variables namely (i) Treatment (1=Standard; 2=Test) (ii) Cell type (1=Squamous, 2=Smallcell, 3=Adeno, 4=Large) (iii) Survival time (iv) Status (0=censored, 1= recurrence) (v) karnofsky Performance Score (100=good) (vi) Diagnostic time (Months) (vii) Age (in year) (viii) Prior therapy (0-No, 1-Yes).

Application II: We considered a culling data set [22-23]. The data set contains 13836 observations and 6 variables namely (i) Cow's Identifier (ii) Time to Culling (in the month) (iii) Status (0=Censored, 1= Observed) (iv) Herd: Herd Identifier (v) Time asses (somatic cell count day) (vi) Log SCC (Logarithm of the somatic cell count).

5. Data Analysis

R studio version 1.2.50 was used for analysis. Xgamma baseline distribution codes/function and frailty model were created based on R packages of "survival" [24], "parfm" [18], "frailtypack", "frailtyEM" [16], "Kendall's tua" [5] was used to measure the relationship between any two event times from the same cluster. The lowest value of Akaike's Information Criteria (AIC=log-likelihood) + 2 (P), where P is the number of parameters) and Bayesian Information Criteria (BIC=-2 (log-likelihood) +P(log/n) is used to identify the best model for real-life data.

6. Results

To the best of our knowledge, the perfm command in the R software can be used to generate a parametric frailty model with an Xgamma baseline distribution. Table 1 provides the Xgamma distribution for the baseline hazard distribution and four frailty distributions for each data. The model result provided that Xgamma distribution with Gamma (Ga), Lognormal distribution (LN), Inverse Gaussian (IG), and Positive Stable (PS) frailty model gave almost close result to covariates in Veterans Administration lung cancer study and culling of dairy heifer cows' data sets. The Xgamma distribution with Positive Stable frailty distribution was found to be best compared to other models for the Veterans' Administration Lung Cancer data set due to the lowest AIC (1451.167) and BIC (1471.607) values (Table 1). The estimated hazard ratio [95% Wald CI] of significant ($p<0.05$) covariates are shown in Figure 1. The frailty value was predicted for each Veterans Administration lung cancer data based on the Positive stable (PS) frailty model, as shown in Figure 2.

Table 1. *Result of Comparing four frailty models with Xgamma baseline distribution for Veterans' Administration Lung Cancer study data set.*

Parameters /Covariate	Frailty Models														
	Gamma			Log-Norma (LN)			Inverse Gaussian (IG)			Positive Stable (PS)			None		
	Estimate	**SE**	***P*-value**	**Estimate**	**SE**	***P*-value**	**Estimate**	**SE**	***P*-value**	**Estimate**	**SE**	***P*-value**	**Estimate**	**SE**	***P*-value**
Frailty Θ	0.207 0.363	0.18 0.131		1 0.403	0 0.145		0.321 0.373	0.414 0.141		0.232 0.378	0.125 0.143		- 0.402	- 0.136	
Treatment	0.212	0.199	0.286	0.099	0.186	0.595	0.205	0.2	0.305	0.183	0.202	0.363	0.098	0.181	0.586
KPS	-0.041	0.004	<0.001 ***	-0.038	0.004	<0.001 ***	-0.04	0.005	<0.001 ***	-0.041	0.004	<0.001 ***	-0.046	0.004	<0.001** **
Diagnosis time	-0.003	0.009	0.727	0.001	0.009	0.903	-0.003	0.009	0.752	-0.003	0.009	0.782	-0.002	0.009	0.856
Age	-1.019	0.007	0.006** *	-0.016	0.007	0.027*	-0.019	0.007	0.009** *	-0.019	0.007	0.011** *	-0.018	0.007	0.001** **
Prior	0.074	0.227	0.744	-0.085	0.221	0.701	0.076	0.228	0.738	0.069	0.227	0.76	-0.086	0.216	0.691
AIC	1462.18			1454.46			1458.38			1451.167			1451.17		
BIC	1482.62			1478.82			1474.91			1471.607			1471.61		

*Significantly differed at ***0.1% level (P < 0.001), **0.5% level (P < 0.005), *5% level (P < 0.05); KPS: Karnofsky Performance score; SE: Standard Error

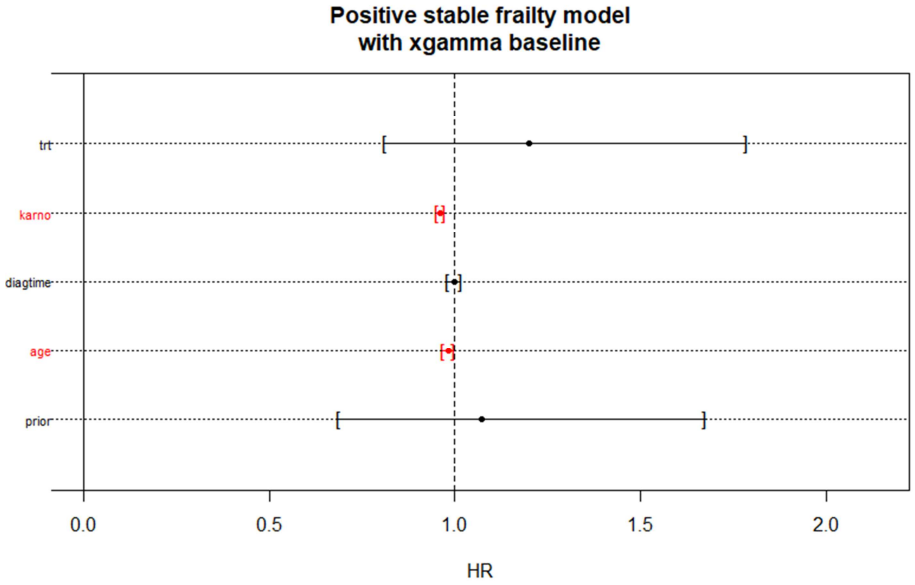

Figure 1. *Hazard ratio for covariates in Veterans' Administration Lung Cancer data by Lognormal (LN) frailty with xgamma baseline distribution (Significant covariate with 95% CI has coloured in red).*

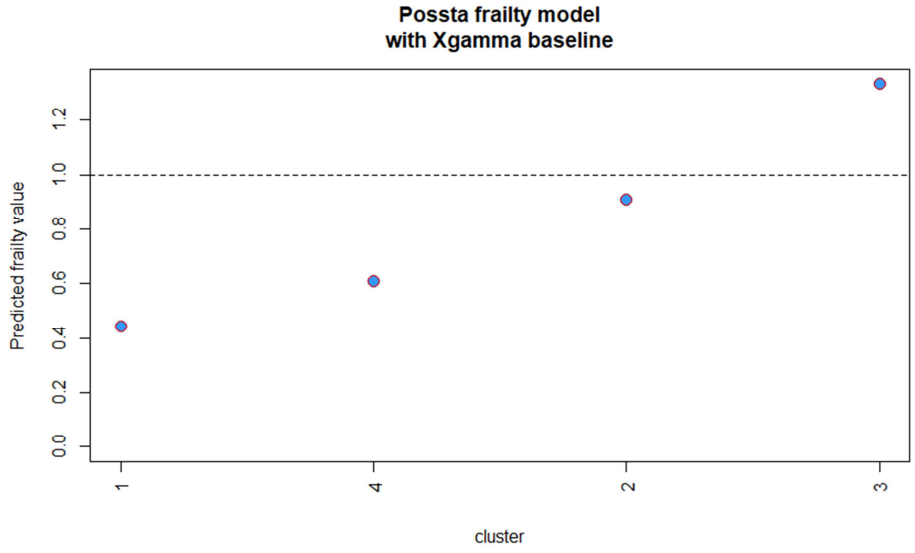

Figure 2. *Predicted frailty values for each veteran's Lung Cancer data based on Lognormal frailty model with xgamma baseline distribution.*

Table 2. *Result of Comparing four frailty models with Xgamma baseline distribution for Culling of dairy heifer cows' data set.*

Parameters/ Covariate	Frailty Models														
	Gamma			Log-Norma (LN)			Inverse Gaussian (IG)			Positive Stable (PS)			None		
	Estimate	**SE**	***P*-value**	**Estimate**	**SE**	***P*-value**	**Estimate**	**SE**	***P*-value**	**Estimate**	**SE**	***P*-value**	**Estimate**	**SE**	***P*-value**
Frailty	0.103	0.038		0.132	0.04		0.107	0.039		0.009	0.006		-	-	
Θ	0.101	0.005		0.098	0.005		0.101	0.005		0.1	0.005		0.1	0.005	
LSCC	0.08	0.02	<0.001***	0.079	0.201	<0.001***	0.08	0.02	<0.001***	0.079	0.02	<0.001***	0.08	0.019	<0.001***
AIC	14182.9			14179.1			14183.6			14188.9			14190.62		
BIC	14203.5			14199.8			14203.5			14209.6			14204.38		

*Significantly differed at ***0.1% level (P<0.001), **0.5% level (P<0.005), *5% level (P<0.05); LSCC: Logarithm of the somatic cell count; SE: Standard error

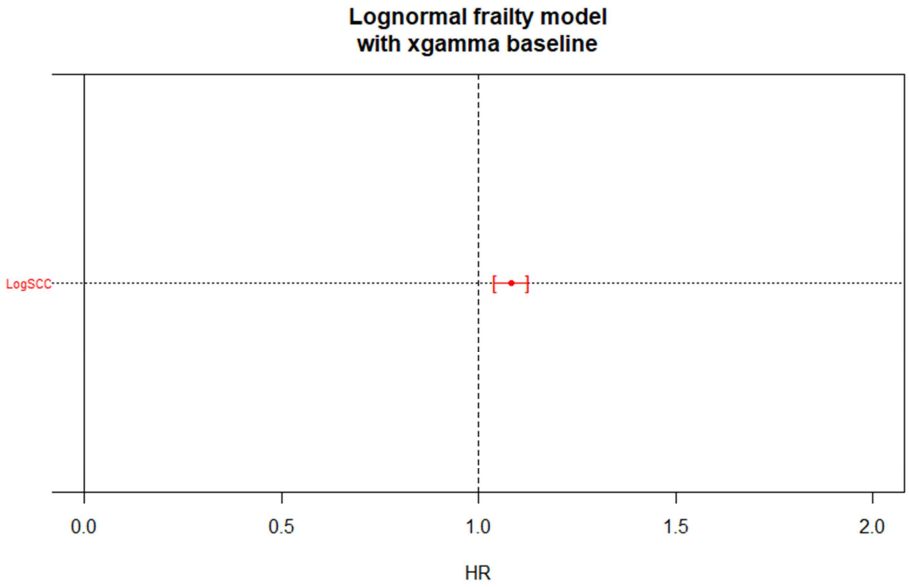

Figure 3. *Hazard ratio for covariates in Culling of dairy heifer cow's data by Lognormal (LN) frailty with xgamma baseline distribution (Significant covariate with 95% CI has coloured in red).*

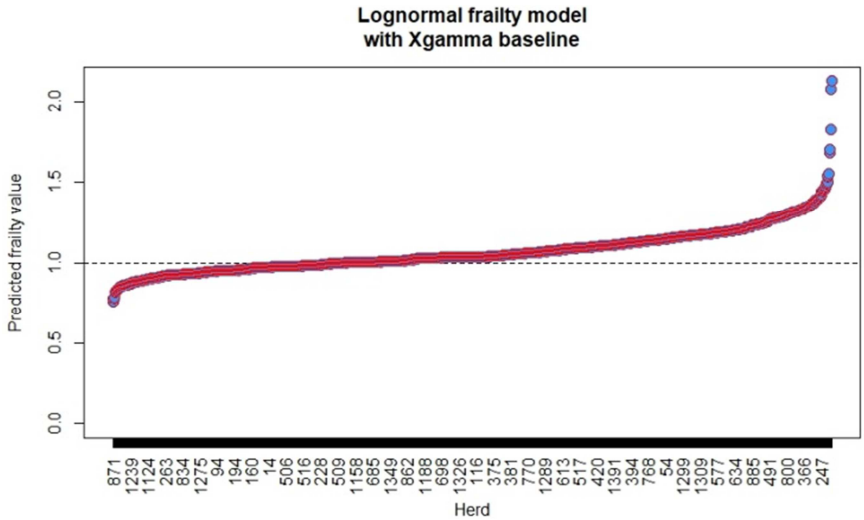

Figure 4. *Predicted frailty values for each culling of dairy heifer cow's data based on Lognormal frailty model with xgamma baseline distribution.*

In the culling of dairy heifer cow's study data, Xgamma distribution with Lognormal (LN) frailty distribution was identified as the best model due to minimum AIC (14179.1) and BIC (14199.8) values. The estimated hazard ratio with confidence interval [95% wald CI] of significant covariates are shown in Figure 3. The frailty values were predicted for culling of dairy heifer cows date based on the Lognormal frailty models, as shown in Figure 4.

7. Conclusions

Estimating frailty or random effect for survival models is essential and has produced better outcomes than ordinary models for lifetime data analysis. This research demonstrated the modeling and application of frailty models with the Xgamma distribution as the baseline hazard function. The

marginal likelihood estimation method was used to estimate and evaluate the significance of the parameters in the models under consideration to compare them.

To compare the four frailty models and non-frailty models with Xgamma baseline hazard, two real-life data sets for Veterans' Administration Lung Cancer, and culling of dairy heifer cow's data were used. The AIC and BIC were applied to determine which of the frailty models investigated provided the best fit for this data set. The study's findings revealed that (i) the Positive Stable (PS) model with Xgamma baseline hazard is the best model for the Veterans' Administration Lung Cancer data set, (ii) the best model for the culling of dairy heifer cow's study data is the Lognormal (LN) frailty model with Xgamma baseline hazard. The study results reveal that the Xgamma distribution with frailty outperformed the non-frailty model (Xgamma without frailty).

Frailty models with an Xgamma baseline distribution are the best alternative method for utilizing clustered survival data. In the survival analysis, the previous identification of an appropriate baseline and frailty distribution for estimating the parameter was determined to be the best technique.

Conflict of Interest

There is no conflict of interest.

References

[1] Klein JP, Moeschberger ML, Survival Analysis: Techniques for censored and Truncated Data. Springer-verlag, 2003.

[2] David G. Klenbaum and Mitchel Klein, Survival Analysis. A Self-Learning Text, Third Edition. Series of Statistics for Biology and health. Springer-Verlag New York. DOI: 10.1007-1-4419-6646-9, 2012.

[3] Duchateau L, Janssen P, Legrand C, Nguti R, Sylvester R, "The shared Frailty Model and the Power for Heterogeneity Test in Multicenter Trials". Computational Statistics & Data Analysis, 40 (3), 603-620, 2002.

[4] Vaupel JW, Manton KG, Stallard E, "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality". Demography, 16 (3), 439-454, 1979.

[5] Hougaard P, Analysis of Multivariate Survival Data. Lifetime Data analyses, 1 (3), 255-283, 2000.

[6] Dunhateau L, Janssen P, The Frailty Model. Series of Statistics for Biology and Health. Springer-Verlag. DOI: 10.1007/978-0-387-72835-3, 2008.

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

[8] Clayton, D., Cuzick, J, Multivariate generalizations of the proportional hazard model. Journal of the Royal Statistical Society (A) 148, 82-117, 1983a.

[9] Ibrahim J. G., Chen MH., Sinha D, Frailty Models. In Bayesian Survival Analysis. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-1-4757-3447-84, 2001.

[10] Lindley DV. Fiducially distribution and Bayes theorem. Journal of Royal Statistical Society A. 1958; 20 (1): 102-107.

[11] J. Nagaraj, S. Parthasarathy, C. Ponnuraja, "Lindley Distribution as Frailty Models with Application to Life Time Data". Advances and Applications in Statistics. 75, 119-134. http://dx.doi.org/10.17654/0972361722031, 2022

[12] Subhradev Sen, Sudhansu S. Maiti, N. Chandra "The Xgamma Distribution: Statistical Properties and Application". Journal of Modern Applied Statistical Methods. Vol. 15, No. 1, 774-788, 2016.

[13] Subhradev Sen, Sudhansu S. Maiti, N. Chandra, "Survival estimation in Xgamma distribution under the progressively type-II right censored scheme". Model Assisted Statistics and Applications 13 (2018) 107-121. DOI 10.3233/MAS-180423 2018.

[14] Cox DR, "Regression Model and Life-Tables". Journal of Royal Society B, 34 (2), 187-220, 1972.

[15] Van den Berg Gj, Drepper PM, "Inference for shared-Frailty Survival Models with Left-Truncated Data". Working Papers 12-5, University of Mannheim, Department of Economics. URL http://ideas.repec.org/p/mnh/wpaper/30729.html. 2012.

[16] Balan, TA, Putter H "Frailty EM: An R Package for Estimating Semi Parametric Shared Frailty Model". Journal of Statistical Software. 90 (7): 2019. DOI: 10.18637/jss.v090.i07. 2019.

[17] David D. Hanagal, "Modeling Survival Data Using Frailty Models. Industrial and Applied Mathematics". Springer Nature Singapore Pte Ltd (295 Pages). DOI: 10.1007/978-981-15-1181-3, 2011.

[18] Munda, Marco & Rotolo, Federico & Legrand, Catherine, "Parfm: Parametric Frailty Models in R", "Journal of Statistical Software, Foundation for Open Access Statistics, Vol. 51 (i11), 2012. http://hdl.handle.net.10.18637/jss.v051.i11.

[19] Hougaard P, "Frailty Models for Survival Data". Lifetime Data Analysis, 1 (3), 255-273, 1995.

[20] Duchateau L, Janssen P, The frailty model. Springer. New York: Springer-Verlag, 2008.

[21] De Vliegher. S, Barkema. H. W, Opsomer. G, et al., Association between somatic cell count in early lactation and culling of dairy heifers using Cox frailty models. J. Dairy Sci. 88, 560-568, 2005.

[22] D Kalbfleisch and RL Prentice, The Statistical Analysis of Failure Time Data. Wiley, New York, 1980.

[23] Byar, D. P., The Veterans Administration study of chemoprophylaxis of recurrent stage I bladder tumors: comparisons of a placebo, pyridoxine, and topical thiotepa. In Bladder Tumors and Other Topics in Urological Oncology (M. Pavone-Macaluso, P. H. Smith, and F. Edsmyn, eds.). New York: Plenum, pp. 363-370, 1980.

[24] Therneau TM, Survival: A Package for Survival Analysis in S. R. Package version 2.44-1.1, https://cran.r-project.org/package=survival. 2019.