ISSN 2063-5346

ESTIMATION OF WEIGHTED-XGAMMA FRAILTY MODEL WITH APPLICATION OF SURVIVAL DATA

P. ASHOK KUMAR[1*], M. MUTHUKUMAR[2]

Article History: Received: 02.07.2023 Revised: 15.07.2023 Accepted: 23.07.2023

Abstract

The study aims to provide a new frailty model for modelling unobserved heterogeneity in survival data. We proposed a Weighted-Xgamma (Wxg) distribution as frailty to investigate the statistical characteristics of the distribution and Laplace transform function that may be used to calculate hazard and marginal survival functions. To fit the models, Weighted-Xgamma distribution as frailty and parametric distributions such as Exponential, Weibull, Log-Logistic, and Lognormal as baseline distribution were used. The Expectation-Maximization (EM) algorithm is suggested to estimate the parameter of the models. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were used to assess the model fitness. To fit the proposed model, a well-known veterans Administration lung cancer study data set was applied. The study results revealed that the Weighted-Xgamma (Wxg) frailty distribution shows a better fit than the other frailty models. So we suggested the Weighted-Xgamma (Wxg) frailty model is an alternative approach for survival analysis.

Keywords: Frailty Models. Xgamma distribution. Weighted-Xgamma frailty. Cluster survival data. EM-algorithm.

[1*]*Research Scholar, Department of Statistics, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India. Email ID: ashokkumar.stat2023@gmail.com*

[2]*Assistant Professor, Department of Statistics, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India. Email ID:muthustat2014@gmail.com*

DOI:10.48047/ecb/2023.12.9.204

1. INTRODUCTION

In biological, epidemiological, and clinical studies, there are two types of models for analysing censored survival data: survival models and frailty models. The Cox Proportional Hazard (PH) model in survival analysis is a logical extension of the Frailty model. When an unobserved source of heterogeneity exists in the data, the standard statistical strategy known as the Cox-proportional hazard model [1] is no longer appropriate. The frailty model can be used in this situation to model unobserved heterogeneity among challenges or groups.

Several authors mentioned frailty models, the gamma [2], and Clayton's (1978) [3] random impact model of bivariate survival, which was a widespread breakthrough, helped popularize the concept of shared relative risk. The parametric distributions for Uniform, Weibull, Lognormal, Positive stable, Inverse Gaussian, Compound Poisson, and Exponential family distributions are also included [4,5,6].

Additionally, Oakes used this model to research the connection between non-negative random variables [7]. we recommend books for "survival analysis: strategies for censored and truncated data" [8]. Because frailty is random, it is usually modelled using a distribution known as the frailty distribution [9-13]. In the closing decade, the modelling of the inverse Gaussian frailty model and assessment of special frailty model were for analyzing more than a few real-life information [14]. Xgamma distribution is the mixture of exponential and gamma distribution with mixing share and get significance for the different shapes of the hazard characteristic [15].

The article's intention is as follows. In section 2, discuss the brief overview of frailty models and propose the Weighted-Xgamma frailty model. A popular parametric—baseline distribution of Exponential, Weibull, Log-logistic, and Lognormal has been discussed in Section 3. In section 4, we use the real-data to demonstrate the applicability of our concept. Finally, Section 5 provides concluding observations of the study.

2. FRAILTY MODELS

Let's consider the Cox PH model and an unobserved heterogeneity source that is no longer calculated in this model using covariates. In a univariate frailty model, it is assumed that random results represent various information and those clusters are impartial [11] and have proportional risks that depend on the random effect, "W" [10]. Let random impact "Z" be a non-negative frailty variable that represents the population's individual-level risk. The frailty model is therefore represented by the conditional hazard as

$$h_{ij}(x/w_i) = h_0(x_i)w_i \exp\left(z_{ij}^T\beta\right) \quad (2.1)$$

Here, j= subject (j=1,2…..n) and i=group (I=1,2….G), the place $h_0(x)$- is baseline hazard function (Here Weighted-Xgamma) , w_i the frailty time in team i,z_{ij} the vector of covariates for subject j in group i, and β the vector of regression coefficients. Additionally, the baseline hazard function is multiplied by the frailty w_i factor. As a result, if $w_i > 1 \; or \; w_i < 1,$ respectively, frailty w_i increases or decreases the likelihood that the relevant event will occur. The Cox PH model [1] is gained as a special instance when $w_i = 1 \; for \; all \; i,$

The i^{th} the subject is obtained from (2.1) as follows:

$$S(x/w_i z_i) = \exp\left[-w_i h_o(x_i)\exp\left(z_{ij}^T\beta\right)\right], \quad (2.2)$$

Where $h_o(x_i) = \int_0^x h_o(s)ds$ is the time-dependent baseline hazard function. As a result, given W=w_i, the conditional survival function (2.2) represents the likelihood of a i^{th}subject surviving unit timex_i.

We must incorporate the conditional survival function (2.2) on frailty in order to obtain the marginal survival function, which means we may stop relying on unseen quantities. Keep in mind that this is equivalent to computing the frailty distribution's Laplace transformation. Using the integrating $S(x/w_i z_i)$ given in (2) on W=w_i, we may obtain if f(z) is the frailty distribution.

$$S(x/w_i z_i) = \int_0^\infty exp\left[-w_i h_0(x_i)\exp\left(z_{ij}^T\beta\right)\right] f(w_i)dw_i = L_f\left(h_o(x_i)\exp\left(z_{ij}^T\beta\right)\right) \quad (2.3)$$

Where $L_f(.)$ denotes the frailty distribution's Laplace transform. As a result, the marginal hazard function can be calculated using eq. (2.3).

$$h(x_i/z_i) = -\frac{h_0(x_i)\exp\left(z_{ij}^T\beta\right) L_f'(h_0(x_i)\exp\left(z_{ij}^T\beta\right)}{L_f\left(h_o(x_i)\exp\left(z_{ij}^T\beta\right)\right)} \quad (2.4)$$

where $L_f'(x) = \frac{\partial}{\partial x}L_f(t)$. As a result, both the marginal survival function and the hazard function (described above) assess the survival and risk of a randomly selected subject from the study population [10].

As previously discussed, estimating each marginal survival and hazard function necessitates the use of a frailty distribution with a Laplace transformation on the closed form. Such marginal functions in this paper, in particular for them $F\sim WXG(\theta)$model, take the form of.

2.1. Xgamma distribution

If 'Y' is a random variable, then the probability density function (pdf) of a weighted distribution is defined as [16].

$$f(y) = \frac{w(y)f_0(y)}{E[w(y)]}. \quad (2.5)$$

Where f$f_0(y)$ is a pdf and w(y) is a non-negative weight function.

We take here $w(y) = y^r for\ r = 1,2, \ldots.$, and $f_0(y)$ is the pdf of Xgamma distribution [16] i.e.,

Here, we assume that $f_0(y)$ is the pdf of the Xgamma distribution and that $w(y) = y^r for\ r = 1,2, \ldots.$, [16]

$$f_o(y) = \frac{\theta^2}{(1+\theta)}\left(1+\frac{\theta}{2}y^2\right)e^{-\theta y}; y>0, \theta>0 \quad (2.6)$$

If we consider $W(Y) = Y^r for\ r = 1,2 \ldots \ldots.$, then E[w(Y)] is nothing more than the r^{th} order raw moment of the Xgamma distribution, so $E[Y^r] = \frac{r!(\theta+r+a_r)}{\theta^r(1+\theta)}$;

where $a_r = a_{r-1} + r,\ \ r = 1,2 \ldots \ldots$wit.h $a_0 = 0$ and $a_1 = 2$, which simplifies to

$$E[Y^r] = \frac{r!}{\theta^{r(1+\theta)}}\left[\theta + \frac{(1+r)(2+r)}{2}\right]; for\ r = 1,2 \ldots. \quad (2.7)$$

The r^{th} order moment Weighted-Xgamma distribution has the following distribution after (2.7). If the pdf is of the form, then a non-negative continuous random variable, Y, is said to follow a Weighted-Xgamma distribution with parameters r and θ if it is pdf is of the form

$$f(y) = \frac{2\theta^{r+2}}{r!\,[2\theta+(1+r)(2+r)]}\left(y^r + \frac{\theta}{2}y^{r+2}\right)e^{-\theta y}; y>0, >0, r = 1,2,3 \ldots. \quad (2.8)$$

We denoted by $Y\sim WXG(r,\theta)$.

The weighted-Xgamma distribution's k^{th} for k=1,2,.....order raw moment in equation (2.8) is given by

$$E[Y^K] = \frac{(r+k)!}{r!\,\theta^k}\left[\frac{2\theta + (1+r+k)}{2\theta + (1+r)(2+r)}\right] \quad (2.9)$$

The Weighted-Xgamma distribution's cumulative distribution function (cdf) is

$$F(y) = P(Y \le y) = \frac{2\theta}{r!\,[2\theta + (1+r)(2+r)}\left[\gamma(r+1,\theta y) + \frac{1}{2\theta}\gamma(r+3,\theta y)\right] \quad (2.10)$$

The lower incomplete gamma function is defined as $\gamma(a,y) = \int_0^y u^{a-1}e^{-u}du$ subsequently comes the survival function (SF).

$$S(y) = P(Y > y) = \frac{2\theta}{r!\,[2\theta + (1+r)(2+r)}\left[\Gamma((r+1,\theta y) + \frac{1}{2\theta}\Gamma(r+3,\theta y)\right], \quad (2.11)$$

Where $\Gamma(a,y) = \int_y^\infty u^{a-1}e^{-u}$ is the upper incomplete gamma function. The failure rate (FR) or hazard rate (HR) function is obtained as

$$h(y) = \frac{f(y)}{S(y)} = \frac{\theta^{r+1}\left(y^r + \frac{\theta}{2}y^{r+2}\right)e^{-\theta y}}{\left[\Gamma(r+1,\theta) + \frac{1}{2\theta}\Gamma(r+3,\theta y)\right]}; y,\theta > 0 \quad (2.12)$$

2.2. Weighted-Xgamma frailty Model

Consider that the frailty variable 'X' in the conditional for frailty model provided in (2.1) follows the suggested WXG distribution (2.8), with E[X]=1. To determine the parameters of the next model, it is important to make this assumption [17]. As a result of employing the choice parameterization of the WXG distribution in terms of imply [18], the frailty pdf grows to be,

$$= \frac{e^{-w\left(\sqrt{r(r+1)(r+2)}\right)}\left(\sqrt{r(r+1)(r+2)}\right)^{(r+2)}\left[(2w^r) + +\sqrt{r(r+1)(r+2)}\right]}{r!\left[2\left(\sqrt{r(r+1)(r+2)}\right) + (r+1)(r+2)\right]} \quad (2.13)$$

Where r>0 is the (known) Shape Parameter. It is critical to notice in the context of the frailty model that the frailty pdf is unimodal and skewed to the right. [6,12,19].

In general, the quantity of unobserved heterogeneity in a population finding is entirely determined by the variance of the frailty distribution. Assume that the pdf (2.13) is a frailty distribution, the variance is $\sigma^2 = 4\left(\left(\sqrt{r(r+1)(r+2)}\right) + (r+1)(r+2)\right)^{-1}$. As a result, as 'r' increases, the variance reduces, and it tends to be finite as r approaches zero. As a result, lower r values suggest more unobserved heterogeneity among people.

According to its variance, the frailty pdf's (2.13) Laplace transform is provided by

$$\mathcal{L}_f(s) = \left(1 + \frac{S(\sigma^2(\sigma^2+4)(\sigma^2+8)}{2(\sigma^2+2)(\sigma^2+4)}\right)^{-\frac{12}{\sigma^2(\sigma^2+3)(\sigma^2+4)}-1}\left(1 + \frac{s\sigma^2}{4}\right) \quad (2.14)$$

If we evaluate eq. (2.14) at $s = h_0(x_i)\varepsilon_i$, where $\varepsilon_i = \exp\left(z_{ij}^T\beta\right)$ for the sake of simplicity, Assuming WXG frailty, we find that marginal survival function (2.3) is defined by

$$S(x/w_i z_i) = \left(1 + \frac{h_0(x_i)\varepsilon_i(\sigma^2(\sigma^2+4)(\sigma^2+8)}{2(\sigma^2+2)(\sigma^2+4)}\right)^{-\frac{12}{\sigma^2(\sigma^2+3)(\sigma^2+4)}-1}\left(1 + \frac{h_0(x_i)\varepsilon_i\sigma^2}{4}\right) \quad (2.15)$$

Then, the corresponding marginal hazard function (2.4) becomes

$$h(x_i/z_i) = h_0(x_i)\varepsilon_i\left(\frac{12+\sigma^2(\sigma^2+3)(\sigma^2+4)}{2(\sigma^2+2)(\sigma^2+4)+h_0(x_i)\varepsilon_i(\sigma^2(\sigma^2+4)(\sigma^2+8)}\right) - \frac{\sigma^2}{4+h_0(x_i)\varepsilon_i\sigma^2} \quad (2.16)$$

3. PARAMETRIC BASELINE DISTRIBUTIONS

The baseline hazard is described as a parametric feature in the parametric approach, and the vector of its parameter, let's say, is evaluated along with the regression coefficient and the frailty parameters. Exponential, Weibull, Log-logistic, and Lognormal baseline distributions were examined in this article. For each distribution, Table 1 offers the hazard, cumulative hazard feature, and survival functions.

Table 1: The Parametric baseline distributions with hazard function, Cumulative Hazard function, and Survival functions.

Baseline distribution (Parameters)	**Hazard function** $(\boldsymbol{h_0(x)})$	**Cumulative Hazard function** $(\boldsymbol{H_0(x)})$	**Survival function** $(\boldsymbol{S_0(x)})$
Exponential $(\boldsymbol{\theta} > 0)$	θ	θx	$\exp(-\theta x)$
Weibull $(\boldsymbol{\rho}, \boldsymbol{\theta} > 0)$	$\theta\rho x^{-1}$	θx^{ρ}	$\exp(-\theta x^{\rho})$
Loglogistic $(\boldsymbol{\alpha} \in \mathbb{R}, \boldsymbol{k} > 0)$	$\frac{\exp(\alpha)kx^{k-1}}{[1+\exp(\alpha)\,x^k]}$	$\mathrm{Log}[1+\exp(\alpha)\mathrm{x}^{\mathrm{k}}]$	$\frac{1}{[1+\exp(\alpha)\mathrm{x}^{\mathrm{k}}]}$
Lognormal $(\boldsymbol{\mu} \in \mathbb{R}, \boldsymbol{\sigma} > 0)$	$\frac{\emptyset\left(\frac{\log(x)-\mu}{\sigma}\right)}{\sigma t\left[1-\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]}$	$-log\left[1 - \phi\left(\frac{\log(tx)-\mu}{\sigma}\right)\right]$	$1-\phi\left(\frac{\log(x)}{\sigma}\right)$

The most popular method for estimating the parameters in Frailty designs is the marginal Log-likelihood approach [20]. The frailties were taken into account by averaging the conditional log-likelihood with the frailty distribution in consideration. For right-censored cluster survival data, the following assumption is used to estimate the marginal log-likelihood. The random variable is unbiased between the censoring time and the survival time and occasions are non-informative

right-censoring information and the marginal log-likelihood of the commentary facts $X = \left(x_{ij}; i \in I. j \in J_i\right)$ [21].

$$l_{mar}(\psi, \beta, \xi; x/\tau) = \sum_{i=1}^{G} \left\{ \left[\delta_{ij}\left(\log(h_0)\log(y_{ij}\right) + z_{ij}^T\beta\right)\right] + log\left[(-1)^{d_i}\mathcal{L}^{(d_i)}\left(\sum_{j=1}^{n_i} H_0\left(y_{ij}\right)\exp(z_{ij}^T\beta)\right)\right] - log\left[\mathcal{L}\left(H_0(T_{ij})\exp\left(z_{ij}^T\beta\right)\right)\right] \right\}, \quad (3.1)$$

Where $d_i = \sum_{j=1}^{ni} \delta_{ij}$ the variety of activities in i -th cluster, and Laplace radically change L(s) used to be first brought through Hougaard [19] and is used to represent the density features of the frailty distribution. Further, unconditional hazards and survival features can be without problems mentioned in the approach. The $\mathcal{L}^{(q)}(.)$

The q-th derivative of the frailty distribution's Laplace transform is defined as

$$\mathcal{L}(S) = E(\exp(-U_i s)\, f(U_i) dU_i, \;\; s \geq 0. \quad (3.2)$$

Where $\mathcal{L}^{(q)}(.)$ is the Higher-order derivatives of the Laplace transform up to *q= max {d1,d2……dG}*. Hence q-th derivate is given by

$$\mathcal{L}^{(q)}(S) = (-1)^{(q)} E\left(U^{(q)}\exp(-Us)\right) \quad (3.3)$$

3.1. Estimation and prediction

The Estimate value of ψ,β, and ξ are bought by way of optimizing the marginal log-likelihood and it can without difficulty simplify by using calculating the greater order derivatives $\mathcal{L}^{(q)}(.)$ Of the Laplace seriously change up to q=max{d1,d2,…..dG}. The aggregate of parameter estimates and prediction is Symbolic differentiation may be carried out in R, with the aid of the usage of the "EM" algorithm and it used to be used to predict frailties. The frailty term u_i can be predicted by $\hat{U}_i = E(U/x_i, \tau_i; \psi, \beta, \xi)$, with $x_i \; and \; \tau_i$ the data and the truncation time of i -th cluster [21-22]. Therefore, the conditional expectation is given with the aid of

$$E(U\Big/x_i, \tau_i; \psi, \beta, \xi) = \frac{\mathcal{L}^{(d_i+1)}\left(\sum_{j=1}^{n_i} H_o\left(y_{ij}\right)\exp(z_{ij}^T\beta)\right)}{\mathcal{L}^{(d_i)}\left(\sum_{j=1}^{n_i} H_o\left(y_{ij}\right)\exp(z_{ij}^T\beta)\right)} \quad (3.4)$$

4. APPLICATIONS

In this section, we demonstrate the significance of the suggested frailty model by analysing real-life information primarily on Veteran's Lung Cancer collecting data [23]. In contrast to Gamma, Log-Normal, and Inverse Gaussian (IG) frailty modes, which employ Proportional Hazard models like Exponential, Weibull, Log-Logistic, and Log-Normal baseline distributions, the results obtained using the WXG frailty model. We provide the factor estimates and corresponding standard errors for each equipped model.

To determine the best models from all fitted models to the data, the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Bayes components are provided. A useful tool for evaluating the size of the difference between four BIC values is the Bayes issue (BF). The interpretation of twice the natural logarithm of the Bayes problem is taken into consideration while choosing the best fit [24–25]. The models under examination all have the same baseline hazard functions. Finally, we evaluate the goodness-of-fit of the four chosen models using the Cox-Snell residuals.

The Veterans Administration lung Cancer dataset is from the retrospective survey. The data set contains the first and second recurrence times of 130 observations, four 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 (Month) (vii) Age (in Year) (vi) Prior therapy (0=No, 1=Yes).

4.1 Data analysis

R studio Version 1.2.50 was once used to create the code and characteristics of Weighted-Xgamma distribution with frailty models and information analysis. The R code and characteristics are primarily based on programs of "survival" [26], "perfm" [20], and" frailtypack" [21]. The Akaike's Information Criteria (AIC Marco =-log-likelihood) +2(P), the place P is the quantity of parameters) and Bayesian Information Criteria (BIC=-2(Log-Likelihood) + P(log/n) have been used to perceive the pleasant model for existence time data. The technique of Kendall's tau used to be used measure the relation between any two occasions from the equal cluster.

4.2 Results and discussion

Table 2 provides summaries of the frailty models that take into account all of the discovered covariates and include baseline hazard function that are Exponential, Weibull, Log-Logistic, and Lognormal. Among independent variables regarded in the models, there is proof that Treatment, Cell type, survival time, Status, Karnofsky overall performance Score, Diagnostic time (Month), Age (in Year), and Prior remedy are full-size elements in the survival time of patients, regardless of the model, in view that treatment, Karnofsky overall performance rating and Age are significant factors in survival time of patients, regardless of the model, considering 95% confident intervals (CIs) of regression coefficients Bj (j=1,2,…), Calculated by$[\widehat{\beta} \pm 1.96XSE(\hat{\beta}_j)]$, do not zero. In contrast, for the majority of models, treatment, diagnostic time, and prior therapy were not significant. We keep it in the model anyway because it is a clinically important covariate. The estimated of unobserved heterogeneity in the equipped frailty models is statistically distinguishable from zero (p-values of LR check are less than 0.0001), indicating there is a certain degree of unobserved heterogeneity in the data. The estimated frailty variances, on the other hand, are greater for frailty models with the Exponential baseline hazard function, resulting in greater significant heterogeneity amongst the patients. Figure 1 shows the Kaplan-Meier estimate of the survival function for the Veterans Administration Lung Cancer data set. The survival rate appears to trend reasonably close to zero when the time is long, according to the estimated curve.

Table 2. Comparisons of four frailty and non-frailty models under four baseline distributions for Veterans Administration lung Cancer study

The exponential baseline hazard function					
Parameters	WXG MLE (SE)	Gamma MLE (SE)	Log-Normal MLE (SE)	Inverse Gaussian MLE (SE)	No-frailty MLE (SE)
Frailty	0.146 (0.137)	0.128 (0.113)	0.135 (0.127)	0.144 (0.137)	0.145(0.137)
Θ	0.065 (0.045)	0.062 (0.045)	0.059 (0.042)	0.063 (0.045)	0.059(0.040)
Treatment	0.215 (0.193)	0.221 (0.194)	0.214 (0.194)	0.214 (0.194)	(0.139)0.181)
Karnofsky Performance score	-0.032 (0.005) ^	-0.031 (0.005) ^	-0.031 (0.005) ^	-0.031 (0.005) ^	-0.031 (0.005) ^
Diagnosis time	-0.002 (0.009)	-0.001 (0.009)	-0.001 (0.009)	-0.001 (0.009)	-0.003(0.009)
Age	-0.005 (0.009)	-0.005 (0.009)	-0.005 (0.009)	-0.005 (0.009)	-0.001(0.009)
Prior	0.008 (0.220)	0.006 (0.227)	0.006 (0.227)	0.007 (0.227)	-0.129(0.218)
AIC	1457.671	1463.535	1459.946	1460.42	1464.309
BIC	1476.742	1483.975	1480.386	1480.86	1484.829
Kendall's tau	0.06	0.06	0.06	0.06	0.07
Weibull baseline hazard function					
Frailty	0.144 (0.002)	0.148 (0.130)	0.782 (0.000)	0.170 (0.161)	
Θ	1.125 (0.051)	1.055 (0.07)	1.080 (0.057)	1.055 (0.071)	0.982(0.064)
B	0.050 (0.041)	0.051 (0.039)	0.038 (0.031)	0.051 (0.040)	0.063(0.046)
Treatment	0.221 (0.200)	0.240 (0.197)	0.237 (0.202)	0.232 (0.198)	0.137(0.181)
Karnofsky Performance score	-0.032 (0.005) ^	-0.33 (0.005) ^	-0.032 (0.005) ^	-0.032 (0.005) ^	-0.034(0.005) ^
Diagnosis time	0.002 (0.000)	0.001 (0.009)	0.001 (0.000)	0.001 (0.009)	-0.003(0.009)
Age	-0.005 (0.009)	-0.006 (0.009)	-0.005 (0.009)	-0.005 (0.009)	-0.001(0.009)
Prior	0.008 (0.228)	0.006 (0.228)	0.040 (0.001)	0.007 (0.228)	-0.125(0.218)
AIC	1461.572	1470.645	1472.339	1465.747	1476.031
BIC	1484.236	1494.005	1495.699	1489.105	1496.471
Kendall's tau	0.052	0.069	0.07	0.069	
Lognormal baseline hazard function					
Frailty	0.493 (0.398)	1	1 (0.435)	1	1
Θ	0.092 (0.017)	0.040 (0.566)	0.036 (0.253)	0.047 (0.026)	0.035(0.165)

B	0.986 (0.017)	0.981 (0.424)	0.970 (0.160)	0.959 (0.001)	0.982(0.156)
Treatment	-0.001 (0.149)	-0.003 (0.177)	-0.003 (0.185)	-0.003 (0.171)	-0.003(0.167)
Karnofsky Performance score	-0.020 (0.004) ^	-0.019 (0.006) ^	-0.020 (0.004) ^	-0.019 (0.009) ^	-0.023(0.004) #
Diagnosis time	-0.001 (0.009)	-0.001 (0.009)	-0.001 (0.009)	-0.001 (0.009)	-0.002(0.009)
Age	-0.016 (0.005)	-0.012 (0.007)	-0.014 (0.007)	-0.012 (0.004)	-0.020(0.006)
Prior	-0.002 (0.212)	-0.001 (0.217)	-0.001 (0.221)	-0.001 (0.216)	-0.001(0.208)
AIC	1479.732	1481.567	1481.732	1482.131	1482.448
BIC	1502.088	1504.924	1505.091	1505.491	1505.888
Kendall's tau	0.227	0.313	0.314	0.323	0.324
Log Logistic baseline hazard function					
Frailty	0.126 (0.118)	0.119 (0.136)	1.041 (0.797)	0.114 (0.121)	
Θ	-6.343 (0.451)	-6341 (0.468)	-6.657 (0.916)	-63341 (0.468)	-6.017(0.434)
B	1.281 (0.110)	1.276 (0.120)	1.253 (0.111)	1.277 (0.120)	1.266(0.128)
Treatment	0.420(0.916) #	0.418 (0.194) #	0.401 (0.227) #	0.416 (0.195) #	0.338(0.181)
Karnofsky Performance score	-0.022 (0.005) ^	-0.021 (0.005) ^	-0.024 (0.006) ^	-0.021 (0.005) ^	-0.23(0.004) ^
Diagnosis time	0.004 (0.009)	0.004 (0.009)	0.002 (0.009)	0.004 (0.009)	0.00(0.008)
Age	0.021(0.007) *	0.020 (0.007) *	0.013 (0.010) *	0.021(0.007) *	0.023(0.007) #
Prior	0.079 (0.2321)	0.077 (0.237)	0.086 (0.247)	0.078 (0.231)	-0.033(0.221)
AIC	1476.175	1483.241	1477.197	1477.897	1483.151
BIC	1500.535	1506.601	1500.557	1501.257	1510.591
Kendall's tau	0.048	0.056	0.048	0.049	0

^ significant differed at 0.1% level(P<0.001), #0.5% level (0.005), *5% level (P<0.05)

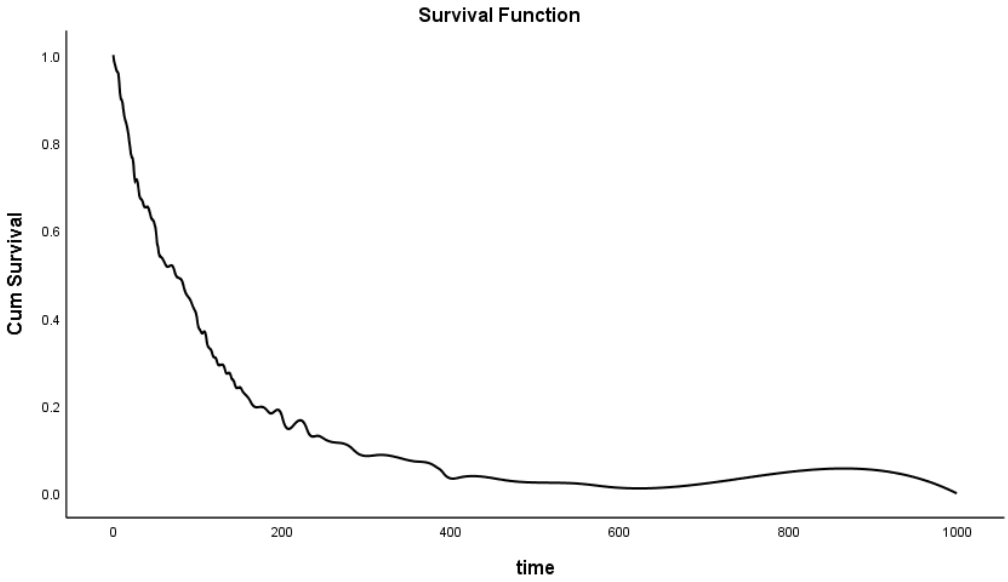

Fig.1 Estimated Survival probability (Kaplan-Meier) for the Veterans' Administration Lung Cancer study

Under the AIC and BIC values, the outfitted frailty models produced a better match than the no-frailty models, regardless of the baseline hazard function. This result was once expected, given that the sample contains some unobserved heterogeneity that the no-frailty models cannot capture. Additionally, the AIC and BIC values of the three WXG frailty functions are lower than those of the conventional models under the same baseline hazard function. As a result, they provide the quality that is appropriate for the data.

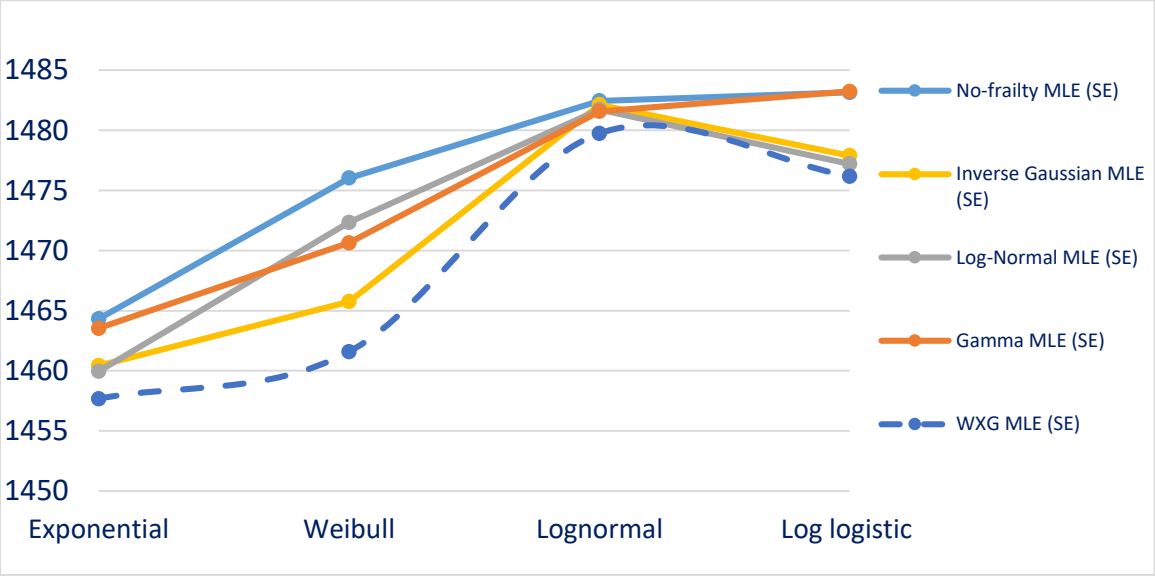

Fig.2 Comparison of AIC values for Veteran's Administration Lung Cancer data

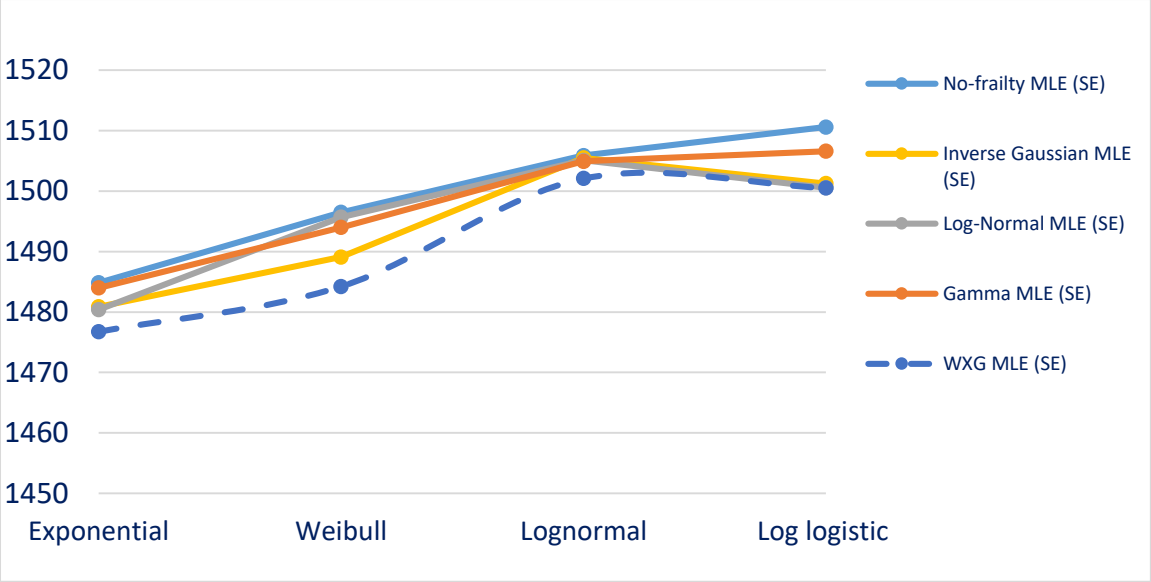

Fig.3 Comparison of BIC values for Veteran's Administration Lung Cancer data

In terms of AIC and BIC, the WXG frailty model with Exponential baseline hazard function exceeds the WXG frailty model with Weibull, Log-Logistic, and Lognormal hazard functions to determine the goodness-of-fit of the WXG frailty model with Exponential baseline hazard function.

Figures 2 and 3 show that the Exponential baseline hazard feature in the WXG frailty model has an excellent goodness-of-fit for the Veterans Administration lung cancer data since it is close to the Identity line. The WXG frailty model no longer exhibits this behaviour when using any other baseline hazard characteristics. So, as our working model, we choose the WXG frailty model with exponential baseline hazard characteristic.

5. CONCLUSIONS

The traditional Cox Proportional hazard model is no longer appropriate for modelling survival data if there is unobserved heterogeneity in the study population. As a result, its application may result in inaccurate estimates of the regression effects. Alternatively, to capture unobserved heterogeneity and improve accuracy, a frailty time should be considered in the Cox PH model. In this study, we present a new frailty model for modelling unobserved variability in survival data. In this case, the frailty distribution is the WXG with unit implies. Each marginal survival and hazard feature was chosen as the baseline hazard characteristic to force the four WXG frailty models when calculating the Laplace transform of this frailty distribution. We stated that the WXG frailty model can accommodate a treatment function with Exponential, Weibull, Log-Logistic, and Lognormal baseline hazard characteristics. We compared the results of our three WXG frailty modes to Gamma, Log-Normal, and Inverse Gaussian frailty models, as well as the Standard Cox model, to a real-life Veterans Administration lung cancer dataset. According to AIC, BIC, and BF, the three WXG frailty models provided the best fits to the Veterans Administration lung Cancer dataset. We concluded that the WXG frailty model with Exponential baseline hazard introduced the best shape to the analysed date set after comparing the three selected frailty models.

REFERENCES

1. Cox DR (1972) Regression models and life-tables. J Roy Stat Soc: Ser B (Methodol) 34(2):187–202
2. Vaupel JW, Manton KG, Stallard E (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16(3):439–454
3. Clayton DG (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65(1):141–151
4. Vaupel JW, Yashin AI (1983) The deviant dynamics of death in heterogeneous populations
5. Hougaard P (1986) Survival models for heterogeneous populations derived from stable distributions. Biometrika 73(2):387–396
6. Aalen OO (1988) Heterogeneity in survival analysis. Stat Med 7(11):1121–1137
7. Oakes D (1989) Bivariate survival models induced by frailties. J Am Stat Assoc 84(406):487–493
8. Klein JP, Moeschberger ML (2006) Survival analysis: techniques for censored and truncated data. Springer
9. Science & Business Media, BerlinDuchateau L, Janssen P (2007) The frailty model. Springer Science & Business Media, Berlin
10. Wienke A (2010) Frailty models in survival analysis. CRC Press, Boca Raton
11. Ibrahim J, Chen M, Sinha D (2001) Bayesian survival analysis springer series in statistics. Springer, New York, pp 978–981
12. Hougaard P (2012) Analysis of multivariate survival data. Springer Science & Business Media, Berlin
13. Sinha D, Dey DK (1997) Semiparametric bayesian analysis of survival data. J Am Stat Assoc 92(439):1195–1212
14. Hanagal DD, Dabade AD (2015) Comparison of shared frailty models for kidney infection data under exponential power baseline distribution. Commun Stat Theory Methods 44(23):5091–5108
15. Subhradev Sen, Sudhansu S.Maiti, N.Chandra (2016). "The Xgamma Distribution: Statistical Properties and Application".Journal of Modern Applied Statistical Methods. Vol.15, No.1, 774-788.
16. Patil, G.P., Rao, C.R and Zelen, M. (1988). Weighted Distribution, in Encyclopedia of Statistical Science, Vol.9, S. Kotz & N.L. Johnson, eds, Wiley, New York,p.565-571.
17. Elbers C, Ridder G (1982) True and spurious duration dependence: The identifiability of the proportional hazard model. Rev Econ Stud 49(3):403–409
18. Subhadev Sen, N.Chandra and Sudhansu S.Maiti (2017) The Weighted-Xgamma Distribution: Properties and Application, Journal of Reliability and statistics Studies, Vol.10, Issue 1(2017): 43-58.
19. Hougaard P (1995) Frailty models for survival data. Lifetime Data Anal 1(3):255–273
20. 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.
21. 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.
22. 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

23. De Vliegher.S, Barkema. H.W, Opsomer. G, et al. (2005). Association between somatic cell count in early lactation and culling of dairy heifers using Cox frailty models. J. Dairy Sci. 88, 560-568
24. Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90(430):773–795
25. Vilca F, Santana L, Leiva V, Balakrishnan N (2011) Estimation of extreme percentiles in Birnbaum Saunders distributions. Compute Stat Data Anal 55(4):1665–1678
26. Therneau TM (2019). Survival: A Package for Survival Analysis in S.R. Package version 2.44-1.1, https://cran.r-project.org/package=survival.