International Journal of Statistics and Applied Mathematics 2023; 8(1): 130-135

ISSN: 2456-1452
Maths 2023; 8(1): 130-135

https://www.mathsjournal.com
Received: 08-12-2022
Accepted: 10-01-2023

P Ashok Kumar
Ph.D. Research Scholar, Department of Statistics, PSG College of Arts & amp; Science, Coimbatore, Tamil Nadu, India

M Muthukumar
Assistant Professor, Department of Statistics, PSG College of Arts & amp; Science, Coimbatore, Tamil Nadu, India

Corresponding Author:
P Ashok Kumar
Ph.D. Research Scholar, Department of Statistics, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India

Parametric frailty models using two-parameter xgamma distribution with an application of survival analysis

P Ashok Kumar and M Muthukumar

DOI: https://doi.org/10.22271/maths.2023.v8.i1b.937

Abstract
When the actual observation measurement is expensive and difficult, an effective approach for estimating the population parameters is applied, such as ranked set sampling. In a survival or time-to-event analysis, a frailty model is a random effects model where the random effect (the frailty) has a multiplicative effect on the hazard. One of the traditional distributions, the Xgamma distribution, is frequently utilized in reliability and regular survival models (frailty-free), but not in frailty models. Due to its inherent flexibility, the Xgamma distribution and its generalizations have become significantly important in survival analysis in recent years. In this study, we attempt to fit two-parameter X-gamma baseline distribution (TPXGD) with parametric frailty models and apply them to the two real-life data sets. The study results revealed that TPXGD with Gamma frailty model is a good choice for Kidney infection data and recurrent asthma attack in children. To make studies of time-to-event data with covariates easier, we propose the Two-parameter Xgamma baseline distribution (TPXGD) with frailty models as suitable alternate models.

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

1. Introduction
In many disciplines, including biology, epidemiology, medicine, and the life sciences, survival or time-to-event data analysis is crucial [1]. Events can be death, recurrence, or any other result of interest [2]. Another crucial element of survival analysis is called "frailty," which is a measure of the heterogeneity or unrecognized random effect that all participants share [3]. To account for such unobserved factors in survival models, Clayton [4] was the first to build a model, and Vaupel [5] coined the term "frailty" in the study of mortality in 1979. As a result, adding frailty (a random effect) to the baseline hazard function will produce results that are more accurate than those from regular survival models (which do not include frailty) [6]. The majority of parametric survival models fit exponential, Weibull, log-logistic, and lognormal distributions with gamma (GA), inverse gaussian (IG), lognormal (LN), and positive stable (PS) frailty distributions. Due to its many hazard functions and forms, the Xgamma distribution and its generalizations have been extensively used in lifetime data analysis recently [7, 8], but they have not been fitted with the frailty models. Due to this, we attempt to fit the parametric frailty models for the Two-parameter X-gamma distribution (TPXGD) in this paper and apply them to two real-world datasets. The structure of this essay is as follows. The characteristics of the TPXGD distribution are covered in Section 2. Frailty models are covered in Section 3, and in Sections 4 and 5, these techniques are applied to two real-life data sets. Finally, Section 6 provides conclusions.

2. Two-parameter Xgamma distribution (TPXGD)
A random variable X follows the xgamma distribution if its probability density function (pdf) is given by

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

and its cumulative distribution function (CDF) is

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

The two-parameter xgamma distribution is a novel distribution that is presented by [9] as an extension of the xgamma distribution. This is done to create a more flexible distribution when modeling real data sets due to the widespread use of the xgamma distribution in various survival analyses. When a random variable, X, exhibits the TPXG distribution, the probability density function and cumulative distribution function are, respectively, provided by

$$f(x,\beta,\theta) = \frac{\theta^2(\beta\theta x^2+2)e^{-\theta x}}{2(\theta+\beta)}; \; x > 0, \theta > 0 \quad (3)$$

$$\mathrm{F}(\mathrm{x},\beta,\theta) = 1 - \frac{(\frac{\beta x^2}{2}+\beta x\theta+\theta+\beta)}{\theta+\beta} \mathrm{e}^{-\theta x} \; \mathrm{x} > 0, \theta > 0, \beta > 0 \; (4)$$

We obtain the xgamma distribution with parameter as a special case of the TPXG distribution for =1 in (3).
The distribution's r-th order moment is determined by

$$E[x^r] = \frac{r!}{2\theta^r(\beta+\theta)}[2\theta + \beta(1+r)^2(2+r)]; r = 1,2,3 \ldots \ldots \ldots \ldots$$

The characteristic function (CF) and hazard function $H(x,\beta,\theta)$ of the model are, respectively, given by

$$\emptyset_x(t) = E[e^{itx}] = \frac{\theta^2}{\beta+\theta}[(\theta - it)^{-1} + \beta\theta(\theta - it); t\epsilon\mathbb{R}, i = \sqrt{-1}]$$

$$H(x,\beta,\theta) = \frac{\theta^2(\beta\theta x^2+2)}{2\theta+\beta(\theta x(\theta x+2)+2} \quad (5)$$

3. Frailty Model

A multiplicative hazard model called a frailty model has three parts: a random effect, a baseline hazard function, and a term that simulates the impact of observable variables. A frailty model's fundamental principle is to include an unmeasured "random" influence in the hazard function to take into account subject variability at the level of the individual subject. Under the proportional hazard model, the hazard function at time t for the i-th subject is $(t_{i,}x_i, c_i)$, where i=1, 2, 3,., n denotes the observed.

$$h(t, x_i, \beta) = h_0(t)\exp(x_i'\beta) \quad (6)$$

A frailty model integrates the value of an additional unmeasured covariate called the frailty, denoted z_i, in the hazard function, producing a hazard function.

$$h(t, x_i, \beta) = h_0(t)\, z_i \exp(x_i'\beta) \quad (7)$$

This concept is expanded to the model with time-varying covariates with the typical notational change. The hazard function that has been altered by the addition of frailty is represented by the subscript equation in equation (7). With the TPXGD baseline distribution, we attempt to fit four frailty distributions in this paper: Gamma (Ga), Inverse Gaussian (IG), Lognormal (LN), and Positive Stable (PS) distributions.

Table 1: Probability density function (p.d.f), Laplace transform (L(s)) and estimation of frailty parametric frailty distributions

Frailty distribution	Probability density function (p.d.f)	The Laplace transformation from frailty L(s) = E[exp(-zs)}	Estimation of frailty $\log\left((-1)^q L^{(q)}(s)\right)$
Gamma frailty (G) (Ga*λ)	$f(x) = \frac{\lambda^{-\frac{1}{\lambda}} x^{\frac{1}{\lambda}-1}\exp(-x/\lambda)}{\Gamma(1/\lambda)}, \theta > 0,$	$(1+\lambda s)^{-\frac{1}{\lambda}}, s \geq 0,$	$-\left(q+\frac{1}{\lambda}\right)\log(1+\lambda s) + \sum_{l=0}^{q-1}\log(1+l\lambda)$
Inverse Gaussian (IG)(IG* λ)	$f(x) = \frac{1}{\sqrt{2\pi\lambda}} x^{-\frac{3}{2}}.exp\left(-\frac{(x-1)^2}{2\lambda x}\right), \lambda > 0.$	$exp\left(\frac{1}{\lambda}\left(1-\sqrt{1+2\lambda s}\right)\right), s \geq 0$	$-\frac{q}{2}\log(2\lambda s+1) + \log\left(K_{q-\left(\frac{1}{2}\right)}(z)\right) - \left[\frac{1}{2}\left(l0g\left(\frac{\pi}{2z}\right)\right) - z\right] + \frac{1}{\lambda}\left(1-\sqrt{1+2\lambda s}\right), where\ x = \sqrt{2\lambda^{-1}(s+\frac{1}{2\lambda})}$
Lognormal (LN) (LN* λ)	$(2\pi\lambda)^{-1/2}x^{-1}\exp\left\{\left[-\frac{(logx)^2}{2\lambda}\right]\right\}, \lambda > 0$	For a lognormal frailty distribution, there is no explicit evaluation of the Laplace transformation, and also Kendall's τ no explicit formula exists (Duchateau and Janssen [12]). Hence, we need Laplace approximation $L^q(s)$ (Macro *et al.* [14]) $(-1)^q\frac{1}{\sqrt{\lambda}}exp\{-g(\hat{u};s,\lambda)\}\left[g^{''}(\hat{u};s,\lambda)\right]^{-1/2}$	$log\left[(-1)^q\frac{1}{\sqrt{\lambda}}exp\{-g(\hat{u};s,\lambda)\}\left[g^{''}(\hat{u};s,\lambda)\right]^{-1/2}\right]$ Where mean = û and variance = $1/\hat{u}(\hat{u};s;\lambda)$
Positive Stable (PS) (PS*u)	$-\frac{1}{\pi u}\sum_{k=1}^{\infty}\frac{\Gamma(k(1-\gamma)+1}{k!}(-u^{-1})^k sin((1-\gamma)k\pi), \gamma \in (0,1)$	$EXP(-s^{1-\gamma}), s \geq 0,$	$EXP(-s^{1-\gamma}), s \geq 0,$

3.1 Marginal Hazard Model and log-likelihood function

The frailty model interprets the regression coefficients differently than the cox model does. Assume that there are n independent clusters and that each cluster has K Subjects to fix the notation. A failure of the J kind is possible for each topic. The marginal mixed baseline hazard model is used to calculate the failure time in the scenario of the jth type of failure on the subject k in cluster I.

$$h_{ijk}\left(t/x_{ijk}(t)\right) = h_{oj}(t)exp\left\{\beta^T X_{ijk}(t)\right\} \quad (8)$$

Where $\beta = (\beta_1, \beta_2, \ldots . \beta_d)^T$ is a vector representing the unknown regression coefficient, $X_{ijk}(t)$ maybe an external time-dependent vector and the baseline hazard functions $h_{oj}(t)$ and $h_o(t)$ are not known. Inferences have relied on the log-likelihood technique because the marginal model approach does not establish a correlation structure for the failure time of a cluster [10] [327]. In right-censored cluster survival data, the marginal log-likelihood and covariate information are assumed to be independent between the censoring time and the survival time, the right censoring is assumed to be non-informative, and the random variables are assumed to be independent between the censoring time and the survival time. The covariate information and the marginal log-likelihood $w = \{w_{ij}; i \in I, j \in J_i\}$ [11].

$$l_{mar}(\psi, \beta, \xi; w/\tau) = \sum_{i=1}^{G} \left\{ \left[\delta_{ij} \left(\log(h_0) \log(y_{ij}) + z_{ij}^T \beta \right) \right] + log \left[(-1)^{d_i} \mathcal{L}^{(d_i)} \left(\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(z_{ij}^T \beta) \right) \right] - log \left[\mathcal{L}\left(H_0(T_{ij}) \exp\left(z_{ij}^T \beta \right) \right) \right] \right\}, \quad (9)$$

Where $d_i = \sum_{j=1}^{ni} \delta_{ij}$ the variety of activities in i -th cluster

3.2 Laplace transform

The density function of the frailty distribution is described by the Laplace transform, and it is simple to construct unconditional survival and hazard functions. Because of this, the Laplace transform can also be used to express the probability function. As a result, frailty distributions with simple Laplace transforms are crucial because they enable the use of the standard maximum likelihood method for parameter estimation [12, 13].

$L^{(q)}(.)$, the q-th derivative of the Laplacce transform [14] of the frailty distribution defined as

$$\mathcal{L}(S) = E(\exp(-z_i s) f(z_i) dz_i, s \geq 0. \quad (10)$$

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

$$\mathcal{L}^{(q)}(S) = (-1)^{(q)} E\left(z^{(q)} \exp(-zs) \right) \quad (11)$$

3.3 Estimation and prediction

The EM algorithm, which combines an expectation and a maximization step, was utilized to forecast the frailties [15]. The frailty z_i is predicted by the formula $z_i = E(Z/W_i, \tau_i; \psi, \beta, \xi)$, where w_i and τ_i are the data and truncation times of the i-th cluster, respectively. As a result, conditional expectation transforms into

$$E(w/x_i, \tau_i; \psi, \beta, \xi) = \frac{\mathcal{L}^{(d_i+1)} \left(\sum_{j=1}^{n_i} H_o(y_{ij}) \exp(z_{ij}^T \beta) \right)}{\mathcal{L}^{(d_i)} \left(\sum_{j=1}^{n_i} H_o(y_{ij}) \exp(z_{ij}^T \beta) \right)} \quad (12)$$

Application to Real-life Data

Application -I

First, we fitted the four frailty models with the TPXGD distribution using the kidney infection data set [16] for model comparison. The dataset includes information on the first and second times an infection recurred (within a day) from the moment the catheter was inserted until it needed to be removed. 38 patients who used portable dialysis equipment had their measures taken, for a total of 76 observations (cluster). The dataset had five variables, including recurrence time, indicator (0=Censored, meaning the catheter could need to be removed for reasons other than kidney infection), and recurrence (1=recurrence), along with covariates for age, sex, and illness type.

Application -II

Second, we used an Asthma attacks data set: Recurrent attack in infants [17] to fit the four frailty models with the TPXGD baseline. Asthma is occurring more and more frequently in infants. Therefore, the objective of the study is a new application of an existing ant-allergic drug administered to children who are at higher risk of developing asthma 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. A patient usually experiences several asthma attacks. Thus, inside a patient, the various events are grouped and historically ordered. In the model, this ordering is a possibility. The data set includes 232 (1776 observation) asthmatic infants between the ages of 6 weeks and 24 months, along with the variables (i)ID: infant identification number (ii)Time (duration days: the time between the end of one event (an asthma attack) and the next event (iii)Status (Censored=0 or observed=1 event time (iv) Drug (placebo=0, drug=1), and (v) fervent (First observation of the patient (1=yes and 0=No).

4.1 Data analysis

The code and function for the TPXGD distribution using frailty models and data analysis were developed using R studio version 1.2.50. The models' fitness was evaluated using Akaile's Information Criteria (AIC = -2 (loglikelihood) +2P) and Bayesian Information Criteria (BIC = -2 (loglikelihood) +P (log/n)) (where P is the number of Parameters). The relationship between any two event times from the same cluster was measured using Kendall's tau methodology [19]. The "survial," [20] "parfm," [21] "frailtyEM,"[22] and "frailtypack" [23] R packages were used to create the code and function for TPXGD.

5. Results

The parametric frailty models under the TPXGD baseline were fitted using the kidney infection and asthma data sets. The comparison findings for each data set for the four frailty models are presented in Table 2. According to the model results, TPXGD with Gamma (Ga), Inverse Gaussian (IG), Positive Stable (PS), and Lognormal (LN) frailty models provided outcomes that were remarkably similar to variables in both kidney infection and asthma attacks. Due to its lower AIC (671.871) and BIC (680.962) values compared to other models for kidney infection data, the TPXGD baseline with Gamma (Ga) frailty distribution was found to be the most efficient.

Table 2: Frailty models comparison under Two-Parameter Xgamma baseline distribution (TPXAD)

Data set	Parameters/Covariates	Gamma (Ga)		Inverse Gaussian (IG)		Positive Stable (PS)		Lognormal (LN)	
		Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Kidney Infection data	Frailty	0.521	0.198	1.023	0.496	0.347	0.21	0.583	0.321
	λ	1.216	0.152	1.176	0.145	1.132	0.154	1.179	0.139
	θ	0.012	0.008	0.013	0.008	0.012	0.008	0.011	0.009
	Sex	-2.043*	0.434	-1.483*	0.422	-1.027*	0.384	-1.982*	0.678
	Age	0.006	0.018	0.007	0.012	0.005	0.011	0.006	0.01
	AIC	671.87		676.62		674.89		682.32	
	BIC	680.96		688.27		686.37		692.02	
	Kendall's Tau	0.215		0.227		0.291		0.219	
Asthma Attacks	Frailty	0.21	0.039	0.221	0.049	0.086	0.023	0.086	0.041
	λ	0.024	0.017	0.025	0.003	0.024	0.004	0.023	0.003
	θ	0.734	0.002	0.866	0.018	0.854	0.022	0.685	0.014
	Drug	-0.102	0.08	-0.104	0.082	-0.098	0.083	-0.025	0.081
	Fevent	-0.273*	0.074	-0.281*	0.073	-0.342*	0.072	-0.291*	0.064
	AIC	16625		16631		16653		16627	
	BIC	16646		16654		16662		16652	
	Kendall's Tau	0.098		0.087		0.082		0.092	
*Significantly differed at 0.1% level (p<0.001)									

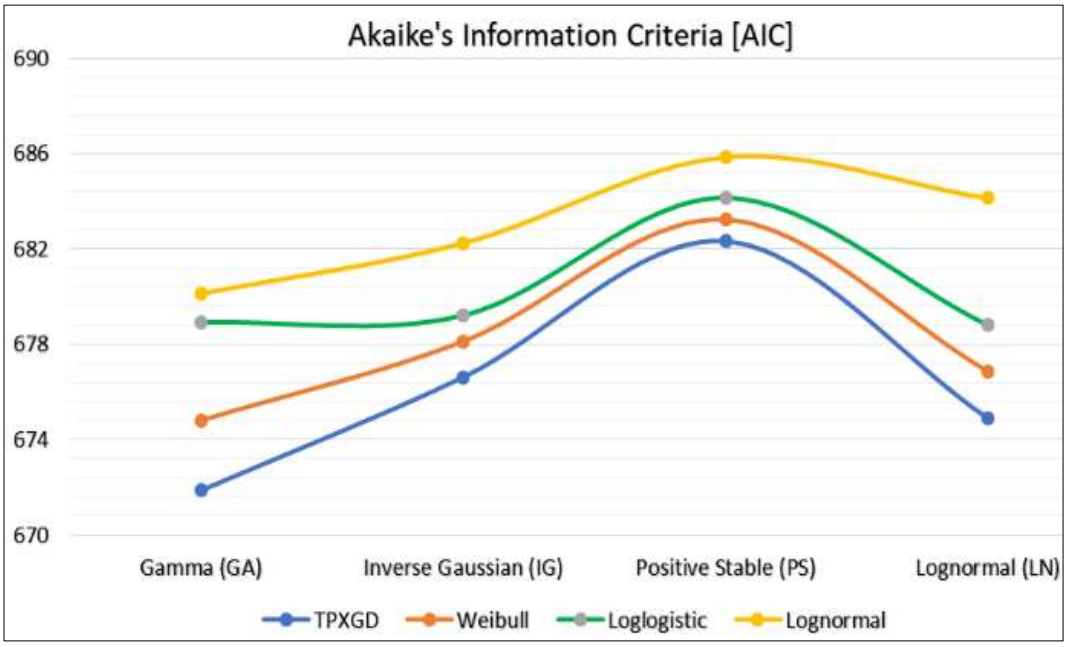

Fig 1: Comparison of AIC values for Kidney Infection data

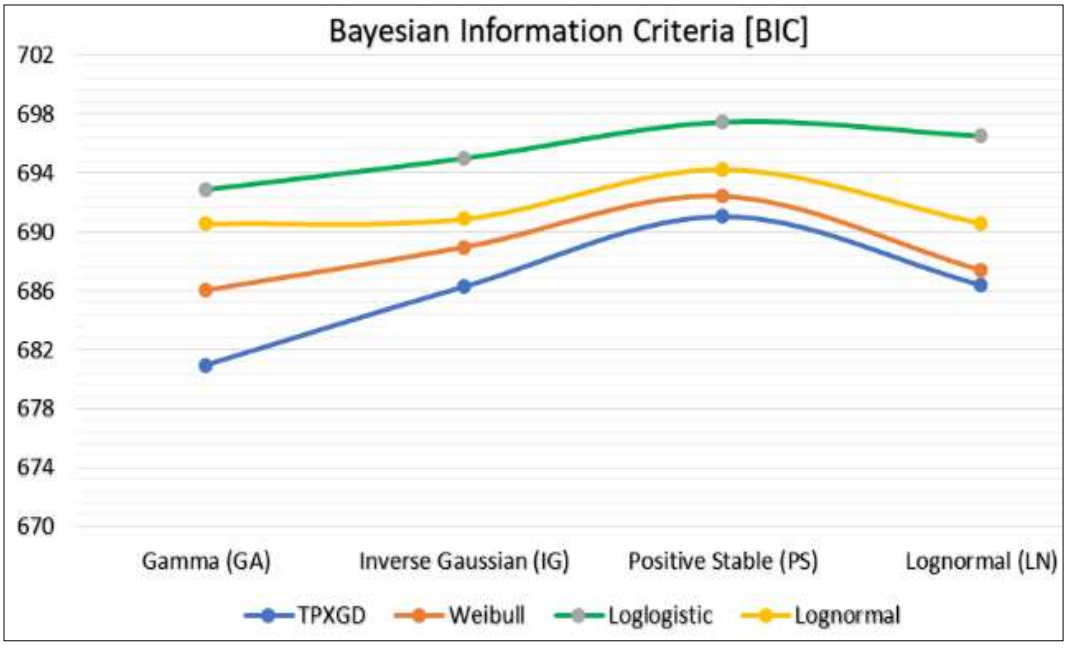

Fig 2: Comparison of BIC values for Kidney Infection data

Figure 1-2 shows the comparison of AIC and BIC values for the kidney infection data with four frailty models with four baseline distributions, including TPXGD, Weibull, Lognormal, and Log-Logistic.

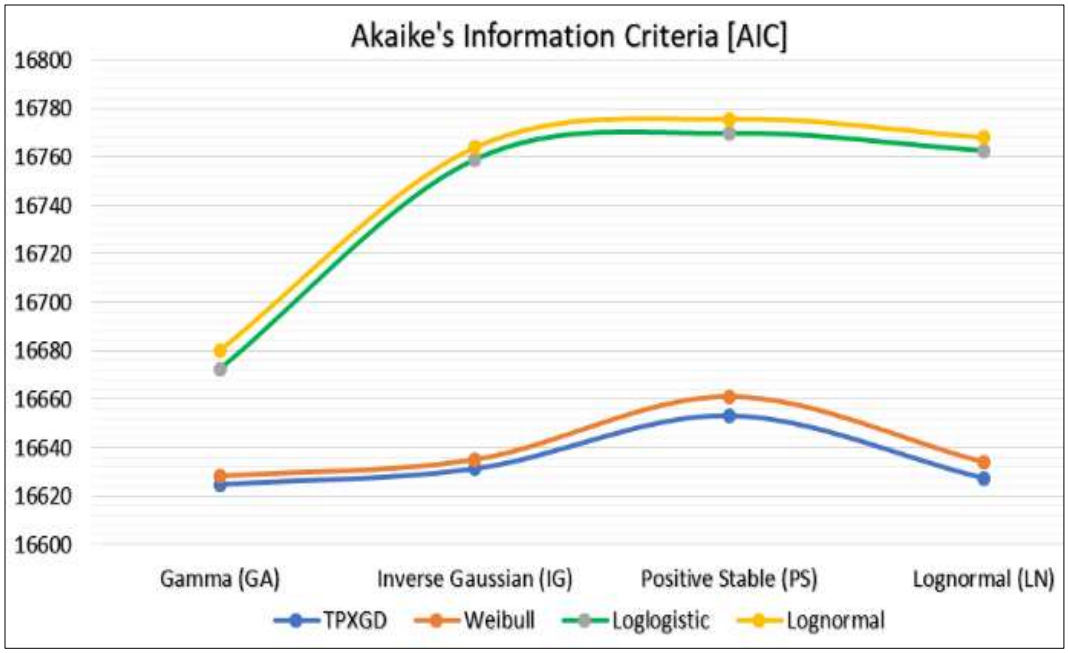

Fig 3: Comparison of AIC values for Asthma Attacks data

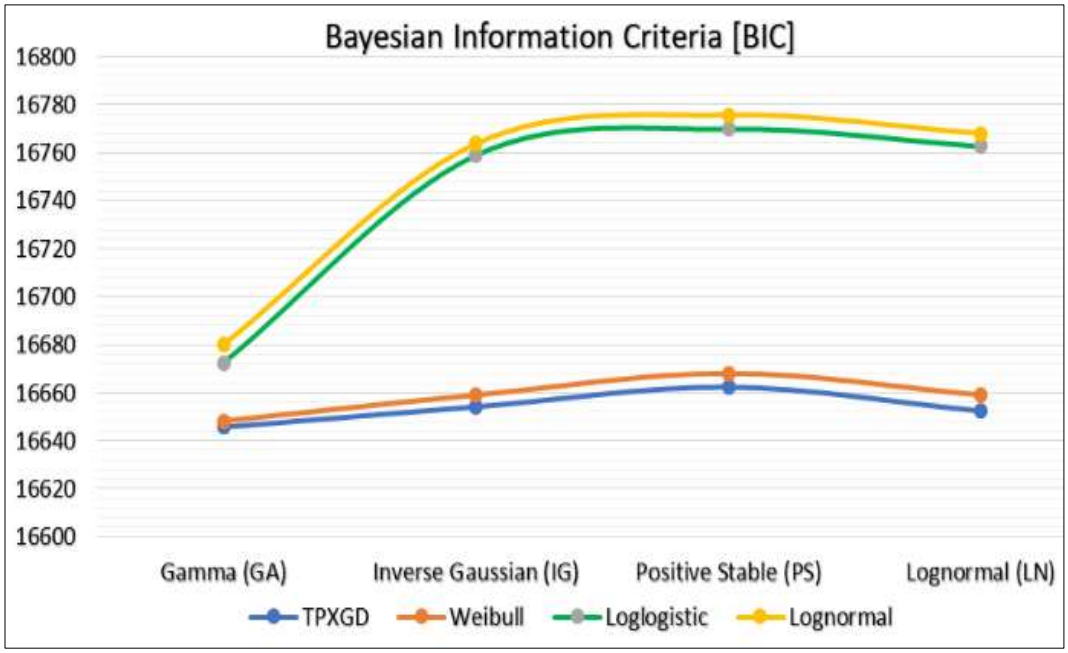

Fig 4: Comparison of BIC values for Asthma Attacks data

Similar to this, the lowest AIC (16624.8) and BIC (16645.7) values among other models for the data on asthma attacks resulted in the conclusion that the TPXGD baseline with Gamma (Ga) frailty distribution was the most efficient. Because each of the baseline distributions had minimal AIC and BIC values, the results showed that the Gamma frailty model is a perfect fit for these data. However, the TPXGD baseline with the Gamma frailty model had the lowest AIC and BIC values (Figures 3-4).

6. Conclusion

In actual practice, selecting and fitting a baseline with a frailty distribution has proven to be essential and has produced better results than other standard methods in time-to-event data analysis. In this study, a baseline distribution called the Two-Parameter X-Gamma Distribution (TPXGD) has been presented. It has been fitted with parametric frailty models. To choose the ideal model for the data sets on kidney infection and asthma attacks, Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) were utilized. For data on kidney infections and asthma attacks, respectively, the TPXD with Gamma (GA) frailty distributions were found to be the best-fit model. Additionally, compared to other baseline distributions, the TPXGD baseline with frailty models demonstrated better fits. Therefore, we propose that TPXGD baseline distribution with the frailty models is a possible alternate method 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(3):439-454.
6. Hanagal D. Modeling Survival Data Using Frailty Models, Industrial and Applied Mathematics, Springer Nature Singapore Limited Company; c2011. p. 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; c2008. https://doi.org/10.1007/978-0-387-72835-3
8. Sen S, Maiti SS, Chandra N. The xgamma Distribution: Statistical Properties and Application. Journal of Modern Applied Statistical Methods. 2016;15(1):774-788.
9. Sen S, Chandra N, Maiti SS. On Properties and Applications of a Two-Parameter Xgamma Distribution, Journal of Statistics Theory and Application. 2018 Dec;17(4):674-685. DOI: 10.2991/jsta.2018.17.4.9.
10. Wei LJ, DY, Weisseld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distribution. Journal of the American Statistical Association. 1989;84(408):1065-1073.
11. 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.
12. Hougaard P. Life table methods for heterogeneous populations. Biometrika. 1984;71:75-83.
13. Duchateau, Janssen P. The Frailty Model, Series of Statistics for Biology and Health, Springer-Verlag New York; c2008. https://doi.org/10.1007/978-0-387-72835-3.
14. Munda M, Rotolo F, Legrand C. PARFM: Parametric Frailty Models in R, Journal of Statistical Software, Foundation for Open Access Statistics. 2012;51(11):01-20. http://hdl.handle.net/10.18637/jss.v051.i11
15. Guo G, Rodriguez G. Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala, Journal of the American Statistical Association. 1992;87(420):969-976.
16. McGilichrist CA, Aisbett CW. Regression with frailty in survival analysis, Biometrics. 1991;47(2):461-466.
17. Adeleke KA, Gurprit Grover. Parametric Frailty Models for Clustered Survival Data and Posters. 2009;17:132.
18. Adeleke K, Grover G. Parametric frailty models for clustered survival data: Application to recurrent asthma attack in infants. Journal of Statistics Applications & Probability Letters. 6(3):89-99. http://dx.doi.or/10.18576/jsap/060301.
19. Daniel, Wayne W. Kendall's tau. Applied nonparametric statistics (2nd Ed.). Boston: PWS-Kent. ISBN: 978-0-534-91976-4; c1990. p. 365-377.
20. Therneau TM. Survival: A Package for Survival Analysis in SR Package Version; c2019, 2.44-1.1. URL https://CRAN.Rproject.org/package=survival.
21. Munda M, Rotolo F, Legrand C. PARFM: Parametric Frailty Models in R, Journal of Statistical Software, Foundation for Open Access Statistics. 2012;51:1-20. https://hdl.handle.net/10.18637/jss.v051.i11
22. 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.
23. Rondeau V, Mazroui Y, Gonzalez JR. frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation, Journal of Statistical Software. 2012;47(4):1-28. URL http: //www.jstatsoft.org/v47/i04/.