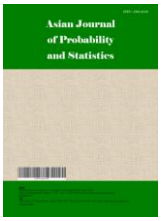

Asian Journal of Probability and Statistics

6(4): 54-65, 2020; Article no.AJPAS.55283
ISSN: 2582-0230

Application of Survival Analysis of TB Patients Using Parametric Model: A Case Study of General Hospital Bayara

Samson Daniel[1*], K. E. Lasisi[1] and Jerry Banister[1]

[1]Abubakar Tafawa Balewa University, Bauchi, Nigeria.

Authors' contributions

This work was carried out in collaboration among all the authors. Author SD designed the study and carried out the research and the analysis. Authors KEL and JB interpreted the result. All the authors approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2020/v6i430169
Editor(s):
(1) Dr. Oguntunde, Pelumi Emmanuel, Covenant University, Nigeria.
Reviewers:
(1) Shakila Bashir, Forman Christian College (A Chartered University), Pakistan.
(2) Kamil Alakuş, Ondokuz Mayıs University, Turkey.
Complete Peer review History: http://www.sdiarticle4.com/review-history/55283

Original Research Article

Received: 01 February 2020
Accepted: 13 March 2020
Published: 21 March 2020

Abstract

Aim: We evaluate the performance of parametric models, mixture of generalized gamma frailty model with Gompertz distribution and compare it with Cox proportional hazard model that is commonly used in the analysis of TB patients and also by [1].
Place and the Duration of the Study: The study was carried out in Bauchi State, Nigeria from January, 2017 to January, 2020.
Methodology: In this study secondary data was used and gotten from the patients' treatment card and TB registers from January 2015 to December 2017. The covariates used were, drug, age, marital status, smoking habit, educational level, weight, category, and risk factor. We used AIC and BIC selection tool to select the model with the lowest value and then compare it with Cox hazard model. Data analysis was done in Stata version 14.
Results: The result of the analysis shows that mixture of frailty model with Gompertz baseline distribution has the lowest AIC and BIC value when compared to Cox Proportional model therefore shows a better goodness of fit for our dataset.
Conclusion: We therefore conclude that mixture of frailty model with Gompertz baseline distribution model can serve as an alternative to Cox Proportional Model.

Keywords: Tuberculosis (TB); gamma frailty model; Gompertz distribution; Cox proportional hazard model.

**Corresponding author: E-mail: samson198690@gmail.com;*

1 Introduction

Tuberculosis (TB) remains one of the main causes of morbidity and mortality in many countries and is considered a global health problem [2]. Although it can be treated if diagnose on time, every year not less than nine million people develop active disease and two million die from it [3] WHO has coordinated multiple and various efforts towards global TB control. One of the goals of the Global Plan to Stop TB is to reduce TB deaths to the barest minimum [4]. In recent studies evidence has shown that TB prevalence and mortality have been under-estimated in many high-burden countries with rising estimates from Nigeria. Nigeria being the most populous country in Africa has an estimated population of over 180 million people. Among the identified 22 high burden countries, Nigeria is ranked 4th in the high burden countries and 1st in Africa with an estimate of 460,000 new cases and 5,000 leprosy per year [5].

Survival analysis is a branch of statistics for analyzing the expected duration of time until one or more events happen. Despite the advantages of parametric model due attention has not been given to this model in the study of non HIV TB patients as a control group. Hence this study tries to compare the performance of several parametric models and compare it with Cox proportional model that was often used by researches in modeling time to event data in survival analysis.

A number of studies have been conducted in which various survival regression methods are compared with parametric models to model time to event data [6,7,8,9,3,10].

2 Materials and Methods

Secondary data was obtained from the treatment card and TB register of the patients in the hospital. The following covariates were used in the analysis drug, age, marital status, smoking habit, educational level, weight, category and risk factor. Kaplan Meier curve was used to estimate the survival pattern of the patients; log rank test was used to identify factors affecting TB survival. Parametric model (exponential, weibull, log logistic, log normal, Gompertz distribution), and compared with Cox proportional model. The model with the lowest AIC and BIC was selected as the best goodness of fit for the dataset.

2.1 Method of analysis

This study included 255 non HIV TB patients in General Hospital Bayara, Opportunistic and infection unit Bauchi. Parametric model, mixture of generalized gamma frailty model based on Gompertz baseline distribution, will be analyze and compared with Cox proportional model in the analysis of non HIV Tb Patients as a control group. Data of the patients were collected and recorded over time from the Tb register and treatment cards of the patients. The parametric models considered in this study are as follows:-

- Propose mixture of generalized gamma frailty model based on Gompertz baseline distribution [11]

The conditional likelihood for the frailty model is given as

$$L = \prod_{i=1}^{G} \prod_{j=1}^{ni} \left[h\left(t_{ij} \backslash z_{i,} X_{ij}\right)\right]^{\delta_{ij}} S\left(t_{ij} \backslash z_{i,} X_{ij}\right)$$

Integrating L over the entire range of the frailty variable Z. the unconditional likelihood function L is obtained as:

$$L = \iint \ldots \int \prod_{i=1}^{G} \prod_{j=1}^{ni} \left[h\left(t_{ij} \backslash z_{i,} X_{ij}\right)\right]^{\delta_{ij}} S\left(t_{ij} \backslash z_{i,} X_{ij}\right) f(z_i) dz_i$$

Where f(.) denotes the pdf of the generalized gamma frailty distribution (GGD (b,d,k)) given by

$$f(z: b, d, k) = \frac{d}{\Gamma(\mathrm{k})b}\left(\frac{z}{b}\right)^{dk-1} e^{\left(\frac{z}{b}\right)^{d}}, \quad z, b, d, k > 0$$

Note that d and k are the shape parameters and b is the scale parameter of the GGD, to make the parameters of the model identifiable we set the mean of the frailty parameter equal to 1 hence

$$E(z) = \frac{b\Gamma(k + \frac{1}{d})}{\Gamma(k)} \Longrightarrow b = \frac{\Gamma(k)}{\Gamma(k + \frac{1}{d})}$$

This gives

$$V(z) = \frac{\Gamma(\mathrm{k})\Gamma(\mathrm{k} + \frac{2}{\mathrm{d}})}{\Gamma(\mathrm{k} + \frac{1}{\mathrm{d}})^2} - 1$$

- Gompertz Distribution

$$f(t) = \omega e^{\alpha t} e^{-\frac{\omega}{\alpha}[e^{\alpha t} - 1],} \quad \omega > 0, t > 0$$

Parametric models used in this study

- Exponential model

$$f(t) = \lambda \exp(-\lambda t), \qquad t > 0 \quad \lambda > 0$$

- Weibull model

$$f(t) \frac{\alpha}{t}\left[\frac{t}{\lambda}\right]^{\alpha} exp\left(-\left[\frac{t}{\lambda}\right]^{\alpha}\right) 0 < t < \infty, \lambda, \alpha > 0$$

- Log-normal model

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} exp\left(-\frac{1}{2\sigma^2}[\ln t - \ln \mathcal{T}\mathrm{m}]\right), \qquad 0 < t < \infty$$

- Log-logistic model

$$f(t) = \frac{\alpha \lambda [\alpha \lambda]^{\lambda - 1}}{[1 + [\alpha t]^{\lambda}]^{2}}, \qquad \alpha > 0, \lambda > 0 \text{ and } t > 0$$

- Cox proportional hazard model

$$\lambda(t/X_i) = \lambda_o(t) + \beta_i X_{ij} + \cdots + \beta_p X_{ip} = \lambda_o(t) + X_i.\beta$$

The data was analyzed in Stata version 14 using all the models and then compared with (Cox proportional model) used by [1].

3 Data Analysis

This chapter provides the results of the analysis done on expert's opinion on the best alternative model in the study of non HIV TB patients, and then compared with Cox model.

3.1 Presentation of tables and results

A total of 255 non HIV TB patients were followed for a period of 18 months.

Table 3.0. Demographic and clinical data of non HIV TB patients

Covariates and health factors	Total	Death	Percentage of covariates
Sex: Male	157	10	61.57
Female	98	5	38.43
Marital status: Single	93	5	36.47
Married	159	10	62.36
Widow	2	0	0.78
Divorce	1	0	0.39
Smoking record: Yes	17	12	6.67
No	238	3	93.33
Education level: Non formal	181	10	70.98
Primary	8	0	3.14
Secondary	28	2	10.98
Tertiary	38	3	14.90
Category of TB: I	79	6	30.98
II	127	8	49.80
III	49	1	19.22
Weight: <20	46	1	18.04
22 – 30	51	1	20.00
>31	158	13	61.96
Age: 15 – 25	91	2	35.69
26 – 35	84	2	32.94
36 – 59	68	5	26.67
>60	12	6	4.70

Table 3.1. Summary of survival data

```
. stset months, failure (status==0) scale(1)
        failure event :     status == 0
obs. time interval    :     (0, months)
 Exit on or before    :     failure
```

255	total observations
0	exclusions
255	observations remaining, representing
15	failures in single-record/single-failure data
4173	total analysis time at risk and under observation
	at risk from t = 0
	earliest observed entry t = 0
	last observed exit t = 39

The table above indicates 255 total observations. 15 failures and 4,173 total analysis times at risk. There were no exclusions. The event of interest is recovery from TB (0 = failure, 1 = recovery).

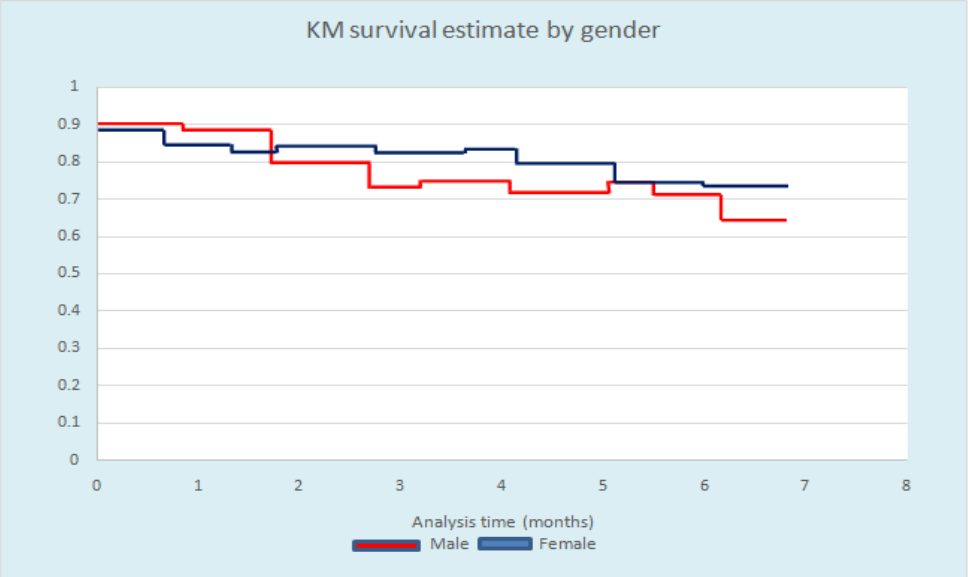

Figure 3.0. Kaplan Meier graph by gender

The Kaplan Meier graph above indicates that gender is not significant in predicting survival since there is no much difference between the subjects.

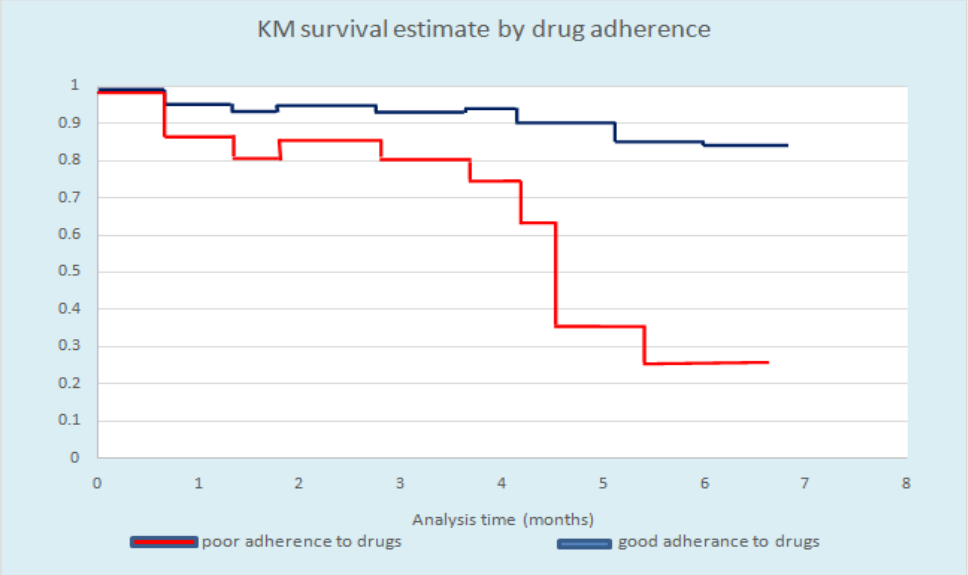

Figure 3.1. Kaplan Meier survival estimate by drug adherence

Figure 3.1 shows that survival is dependent on drug adherence, those with poor adherence end up having complications and dying than those who adhere. Drug adherence is significant in the survival of TB patients.

Table 3.2. Log rank test by smoking habit

. sts test _SmokingHabit, logrank

failure _d: status ==0
analysis time _t months

log-rank test for equality of survivor functions

_Smoking habit	Events observed	Events expected
1	3	6.84
2	12	8.16
Total	15	15.00

Chi2 (1) = 3.99
Pr>chi2 = 0.0459

The log rank test above shows that the significant level of the test is less than (α =0.05) indicating that smoking habit is affecting the survival of TB patients.

Table 3.3. Log rank test by risk factor

. sts test _riskfactor, logrank
failure _d : status ==0
analysis time _t : months

log-rank test for equality of survivor functions

_risk factor	Events observed	Events expected
0	12	13.33
1	0	0.31
2	2	0.44
3	0	0.33
4	1	0.10
5	0	0.48
Total	15	15.00

Chi2 (5) = 14.57
Pr>chi2 = 0.0124

From the log-rank test above with 5 d.f, the significance level of the test is below (α=0.05) indicating that risk factor is significant and therefore affecting the survival of TB patients. The following risk factors are coded into Stata software for the analysis.

None = 0
High fever = 1
Diagnosis delay = 2
Hypertension = 3
Diabetes = 4
Substance abuse = 5

Table 3.4. Stata output of gamma frailty model with Gompertz distribution

Failure _d: status ==0
Analysis time _t: months
Gompertz regression – log relative hazard form
Gamma frailty

No. of subjects = 255 Number of obs = 255
No. of failures = 15
Time at risk = 4173
LR chi2 (8) = 25.93
Log likelihood = -43.597888 Prob > chi2 = 0.0011

_t	Hazard ratio	Std error	z	P>[z]	[95% conf.	Interval]
drug	0.085779	0.0926379	-2.27	0.023	0.0103304	0.7122693
age	1.090521	0.0568851	1.66	0.097	0.9845385	1.207913
_MaritalStatus	0.5908347	0.7088084	-0.44	0.661	0.0562733	60203394
_SmokongHabit	4.390093	4.513256	1.44	0.150	0.585318	32.92725
_Educationallevel	0.7547019	0.3412641	-0.62	0.534	0.3110826	1.830944
weight	1.04367	0.04504701	0.99	0.322	0.9589694	1.135852
category	0.2650246	0.2046013	-1.72	0.085	0.0583648	1.203431
_riskfactor	1.135659	0.4123552	0.35	0.726	0.5574152	2.313755
_cons	0.000287	0.0012393	-1.89	0.059	6.05e-08	1.360568
/gamma	0.2469574	0.0930287	2.65	0.008	0.0646244	0.4292903
/In_the	2.037338	0.8591983	2.37	0.018	0.3533996	3.721277
theta	7.670168	6.589965			1.4239	41.31714

Likelihood-ratio test of theta=0: chibar2 (01) = 2.66 Prob>=chibar2 = 0.052

. estat ic

Akaike's information criterion and Bayesian information criterion

Model	Obs 11(null)	11(model)	df	AIC	BIC
-	255 -56.56181	-43.59789	10	107.1958	142.6084

Note: N=obs used in calculating BIC; see [R] BIC note

From the gamma frailty model with Gompertz distribution above, 1 unit increase in smoking habit is associated with 4.39 times greater the chance that it affects TB survival. The hazard ratio of drug is 0.08 and has a P value less than (α =0.05) which is significant. We will conclude by saying it reduce the hazard of TB mortality.

Table 3.5. Stata output of cox proportional hazard regression model

failure _d: status ==0
analysis time _t: months

Iteration 0: log likelihood = -67.869635
Iteration 1: log likelihood = -56.98712
Iteration 2: log likelihood = -56.833897
Iteration 3: log likelihood = -56.833616
Refining estimates:
Iteration 0: log likelihood = -56.833616

Cox regression – Breslow method for ties

No. of subjects = 255		Number of obs	= 255
No. of failures = 15			
Time at risk = 4173			
LR chi2 (8)	= 22.07		
Log likelihood = -56.833616		Prob > chi2	= 0.0048

_t	Hazard ratio	Std error	z	P>[z]	[95% conf.	Interval]
drug	0.3102259	0.1340739	-2.71	0.007	0.1329855	0.7236883
age	1.038266	0.0282921	1.38	0.168	0.9842695	1.095225
_MaritalStatus	0.6470829	0.4877549	-0.58	0.564	0.1476881	2.835139
_SmokongHabit	2.960524	2.187533	1.47	0.142	0.6956891	12.59859
_Educationallevel	1.114541	0.2636522	0.46	0.647	0.7010333	1.771958
weight	1.017039	0.0281065	0.61	0.541	0.9634164	1.073646
category	0.4466932	0.1860506	-1.93	0.053	0.1974597	1.010509
_riskfactor	1.184179	0.2260777	0.89	0.376	0.814536	1.72157

. estat ic

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
-	255	-67.86964	-56.83362	8	129.6672	157.9973

Note: N=obs used in calculating BIC; see [R] BIC note

From the above table of Cox proportional hazard ratio drug and category are not significant predictors of mortality.

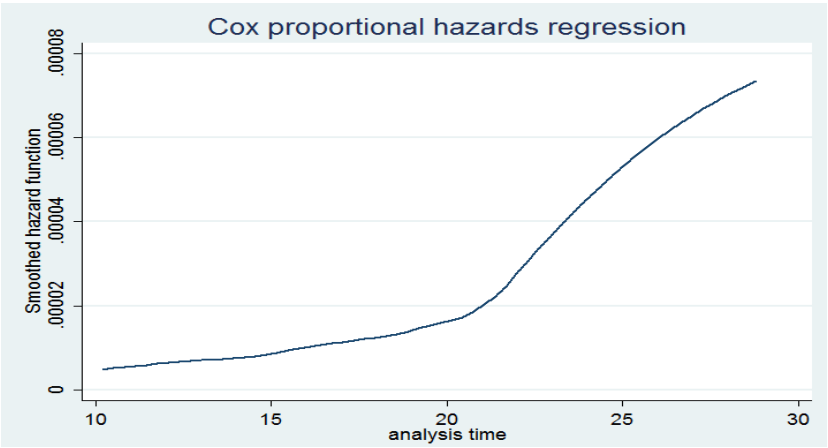

Figure 3.2. Cox hazard plot

The hazard regression plot of Cox model is above zero on the hazard function. It starts to go up from 12 in the analysis time and increases rapidly from 23 showing the hazard rate of the model.

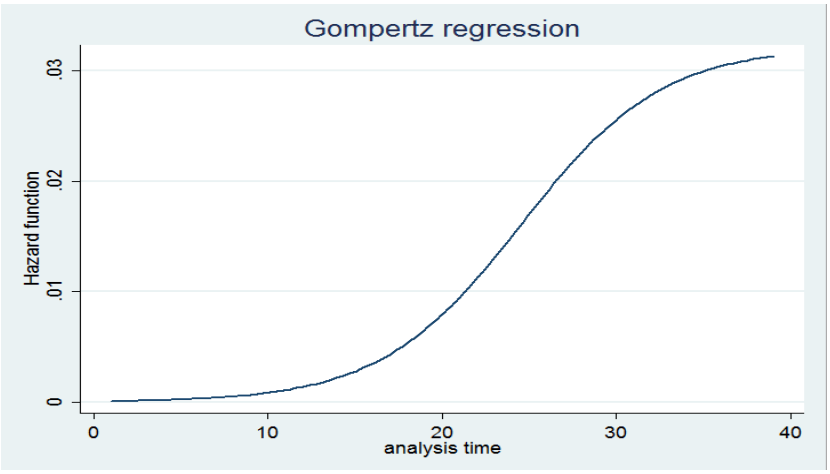

Figure 3.3. Gompertz hazard plot

The hazard regression plot of Gompertz distribution shows a lower hazard when compared to Cox hazard model. The hazard function is on zero. It starts to go up from 8 in the analysis time and increases rapidly from 17 showing the hazard rate of the model. From the hazard plot of the two models Gompertz distribution has a lower hazard rate than Cox proportional hazard model.

4 Discussion of Results

The following covariates were used in the analysis of the data they include drug, age, marital status, smoking habit, educational level, weight, category of TB and risk factor. A total number of 255 HIV negative TB cases were recorded from January 2014 to December 2017. The median followed up period is 18 months. 15 patients died representing (5.88%) of the total number of patients of which 6 (2.35%) are in category I, 8 (3.14%) in category II and 1 (0.39%) in category III. Most of the death occurred within patients who default in taking their drugs and loss to follow-up. 125 patients were lost-to-follow-up during the study period after diagnosis. 220 were censored due to incomplete information.

A total of 94.12% of the TB patients survived the study period, which signifies that TB patients had a high survival rate if they adhere to taking their drugs.

The Kaplan Meier graph in Figure 3.0 indicates that gender is not significant in predicting survival. Figure 3.1 shows that survival is dependent on drug, those with poor adherence end up having complications and dying than those who do not adhere. Drug is significant in the survival of TB patients. From Figure 3.2 the hazard regression plot of Cox model is above zero on the hazard function. It starts to go up from 12 in the analysis time and increases rapidly from 23 showing the hazard rate of the model. Figure 3.3 of hazard regression plot of Gompertz baseline distribution shows a lower hazard when compared to Cox hazard model. The hazard function is on zero. It starts to go up from7 in the analysis time and increases rapidly from 12 From the hazard plot of the two models Gompertz baseline distribution has a lower hazard rate than Cox proportional hazard model.

The Multivariate analysis showed an increased risk of survival for patients with smoking habit [HR=2.960524 SE= 2.187533 (P=0.05) CI (0.6956891 12.59859)], in Cox PH model.

Smoking habit and risk factor are a strong and independent factor affecting the survival of TB patients. From the Gompertz baseline distribution output of Table 3.4, 1 unit increase in smoking habit is associated with 4.39 times greater the chance that it affects TB survival. The hazard ratio of drug is 0.08 and is significant with a probability value of 0.023 less than (α =0.05) we will conclude by saying it reduce the hazard of TB

mortality. The evaluation criteria indicated that parametric model form the best models in multivariate analysis and can lead to more precise result as an alternative to Cox proportional hazard model.

4.1 Model selection based on AIC and BIC

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) is both a model selection tool, they are used in selecting models that have the lowest value from a given set of data. The formular to calculate AIC = -2logL+2k and BIC = -2 * log-likelihood + k * ln (N)

The whole parametric models considered in this study have 2 ancillary model parameter except exponential model that have 1 ancillary model parameter. Cox proportional model doesn't have ancillary model parameter.

Exponential model AIC = -2(-52.212701) + 2(8+1) = 122.425402
BIC = -2(-25.212701) + (8+1) ln 255 = 154.2967739

Exponential regression --- log relative hazard form

No. of subjects	=	255	Number of obs	=	255
No. of failures	=	15			
Time at risk	=	4173			
			LR chi2 (8)	=	15.34
Log likelihood	=	-52.212701	Prob > chi2	=	0.0529

Weibull regression --- log relative hazard form

No. of subjects	=	255	Number of obs	=	255
No. of failures	=	15			
Time at risk	=	4173			
			LR chi2 (8)	=	21.66
Log likelihood	=	-46.029683	Prob > chi2	=	0.0056

Lognormal regression --- accelerated failure-time form

No. of subjects	=	255	Number of obs	=	255
No. of failures	=	15			
Time at risk	=	4173			
			LR chi2 (8)	=	24.45
Log likelihood	=	-45.12726	Prob > chi2	=	0.0019

Loglogistic regression --- accelerated failure-time form

No. of subjects	=	255	Number of obs	=	255
No. of failures	=	15			
Time at risk	=	4173			
			LR chi2 (8)	=	22.29
Log likelihood	=	-45.840609	Prob > chi2	=	0.0044

Gompertz regression --- log relative - hazard form

No. of subjects	=	255	Number of obs	=	255
No. of failures	=	15			
Time at risk	=	4173			
			LR chi2 (8)	=	24.39
Log likelihood	=	-44.365325	Prob > chi2	=	0.0020

Table 4.0. AIC and BIC of the models for non HIV TB patients

Parametric model	Log-likelihood value	AIC	BIC
Exponential	-52.212701	122.4254	154.2968
Log-normal	-45.12726	110.2545	145.6672
Log-logistic	-45.840609	111.6812	147.0939
Weibull	-46.02968	112.0594	147.4720
Generalized gamma shared frailty with Gompertz distribution	-43.59788	107.1958	142.6084
Gompertz	-44.365325	108.7306	144.1433
Cox Prop Hazard	-56.833616	129.6672	157.9973

From the AIC and BIC table above, gamma shared frailty model with Gompertz baseline distribution shows a better goodness of fit among the rest of the models with the lowest value of AIC and BIC [12,13].

5 Conclusion

In this study we have demonstrated that parametric model can be use to model outcome of TB patients, the propose model mixture of gamma shared frailty model with Gompertz baseline distribution having the lowest value of AIC and BIC, shows a better performance and can be used as an alternative to Cox proportional hazard model.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Yong-Jia Ji, Pei-Pei Liang, Jia-Yin Shen, Jian-Jun Sun, Jun-Yang Yang, Jun Chen, Tang-Kai Qi, Zhen-Yan Wang, Wei Song, Yang Tang, Li Liu, Ren-Fang Zhang, Yin-Zhong Shen, Hong-Zhou Lu. Risk factors affecting the mortality of HIV infected patients with pulmonary tuberculosis in the cART era: A retrospective cohort study in China; 2018.

[2] Omari-Sasu AY, Asamoah Owusu D, Boateng MA, Sabogu Z. Survival analysis of tuberculosis patients in Upper West Region of Ghana; 2016.

[3] WHO Publication on TB; 1990.

[4] Global Plan to End Tuberculosis 2016 – 2020 Publication.

[5] Federal Ministry of Health Nigeria (FMOH). National Tuberculosis and Leprosy Control Programme Workers′ Manual. 5th Ed. Abuja: FMOH; 2010.

[6] Ogungbola OO, Akomolafe AA, Musa AZ. Accelerated failure time model with application to data on tuberculosis/HIV co-infected patients in Nigeria; 2018.

[7] Roshany D, Azadi NA, Esmail-Nasab N, Yaghoubi M. Application of parametric, semi-parametric and nonparametric approaches in survival analysis of patients with acute myocardial. J North Khorasan Univ Med Sci. 2011;3:45–51. Google Scholar

[8] Sadegh Kargarian-Marvashi, Shahnaz Rimaz, Jamilen Abolghasemi, Iraj Heydari. Comparing of Cox model and parametric models in analysis of effective factors on event time of neuropathy in patients with type 2 diabetes; 2017.

[9] Samuel J. Wang, Jayashree Kalpathy-Cramer, Jong Sung Kim, David Fuller C, Charles R. Thomas Jr. Parametric survival models for predicting the benefit of adjuvant chemo-radiotherapy in gall bladder cancer; 2010.

[10] Yigzaw Alemu Limenih, Demeke Lakew Workie. Survival analysis of time to cure on multi-drug resistance tuberculosis patients in Amhara Region, Ethiopia; 2019.

[11] Sukhmani Sidhu, Kanchan Jain, Suresh K. Sharma. The generalized gamma shared frailty model under different baseline distributions; 2019.

[12] Akaike H. A new look at the statistical model identification. IEEE Trans Automatic Control; 1974.

[13] Schwarz GE. Estimating the dimension of a model. Annals of Statistics. 1978;6:461–464. (Google Scholar)

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/55283