Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2019;5(2): 110 -119

Original Article

Comparison of Parametric Models: Appication to Hypertensive Patients in a Teaching Hospital, Awka

Ibenegbu Amuche H*, Osuji George A, Umeh Edith U.

Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria.

ARTICLE INFO

Received 25.09.2018
Revised 26.10.2018
Accepted 05.12.2018
Published 01.05.2019

Key words:
Survival analysis;
Censoring;
Parametric models,
Hypertabastic.

ABSTRACT

Introduction: In Nigeria, hypertension is a common sickness among grownups. This research was carried out to determine the best model for predicting survival of hypertensive patients using goodness of fit criteria, Standard Error (SE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Method: A total of 105 patients who were diagnosed with hypertension from January 2013 to July 2018 were considered in which death is the event of interest. Six parametric models such as; exponential, Weibull, Lognormal, Log-logistic, Gompertz and hypertabastic distribution were fitted to the data using goodness of fit such as S.E, AIC and BIC to determine the best model. The parametric models were considered because they are all lifetime distributions.

Results The result shows that the hypertabastic distribution has the lowest AIC and BIC, followed by Gompertz distribution. The standard error also indicates the hypertabastic model is better because it has the least value of standard error. This indicates that in terms of relative efficiency and parameterization the hypertabastic model is the best. The Survival Probability Plot of the six parametric models shows that the Hypertabastic distribution best fitted the data because it shows a clear step function than the other distribution and this justifies the result SE, AIC and BIC presented.

Conclusion: Since hypertabastic distribution has the lowest SE, AIC and BIC it indicates that it is the best parametric model for predicting survival of hypertensive patients in chukwuemeka Odumegwu Ojukwu university teaching hospital Awka, Nigeria.

Introduction

Initially the aim of Survival analysis was solely for investigations of mortality and morbidity on vital registration statistics. Recently, survival analysis has been the major focus of many researchers because of its wide applicability in many fields. Because of the name survival; a lot of people thought that it has to do with only living things. The practice of survival analysis is the use of reason to describe, measure, and analyze features of events for making predictions about not only survival but also 'time - to - event processes' – the length of time until the change of status or the occurrence of an event such as from living to dead, from single to married, or from healthy to sick. Because a life span, genetically, biologically, or mechanically, can be cut short by illness, violence, environment, or other factors, much research in survival analysis involves making comparisons among groups or categories of a population, or examining the variables that influences its survival processes. It is widely applied in the field of medicine and biology; it can also be applied in Engineering, Social Science, Management Science, Mathematical Biology etc. Researchers

* Corresponding Author: Email:amuche4mary@yahoo.com

apply survival analysis to compare the risk of death or recovery from disease between or among population groups receiving different medications or treatments.

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs[1]. There are three purposes of survival analysis which are; Estimate the survivor and hazard functions, compare survivor/hazard functions (e.g. log-rank test) and finally assess the relationship of other variables with survivor and hazard function (e.g. Cox regression).

Censoring is a major problem that arises in surviving analysis. Censoring is when we have little or no information about the survival of an individual[1]. Censoring occurs as a result of this three reasons: the event did not occur at the end of the study, the individual is lost to follow-up during the study; and as a result of death (if not the event of interest), drug reaction and other competing risk the individual withdraws from the study. Censoring is generally divided into three types: right, left and interval. As right censoring occurs far more frequently than other types and its information can be included in the estimation of a survival model, the focus of this research is on the description of right censoring. There are three methods to analyze survival data, namely non-parametric, semi-parametric and parametric. A parametric survival model is one in which survival time (the outcome) is assumed to follow a known distribution such as the Weibull, Exponential, lognormal etc. The works focuses on comparison of six parametric survival models; Exponential, Weibull, lognormal, log logistic, Gompertz and Hypertabastic model and their performance will be assessed using Akaike information Criterion (AIC) and Bayesian Information Criterion (BIC) goodness of fit criteria, the lowest AIC and BIC will be the best model for predicting survival of hypertensive patients in the teaching hospital. In this study, data of hypertensive patients treated in Chukwuemeka Odumegwu Ojukwu University Teaching Hospital Amaku, Awka was analyzed. Patients who were diagnosed with hypertension between 1st January 2013 and 31st July 2018, and received at least one treatment for hypertension in the hospital were included in this study.

The hypertabastic survival model is better preferred to other distribution and the hazard function can help in modeling failure rates in medical, biological and engineering fields[3]. Parametric models such as generalized Gamma, Lognormal, Weibull, Exponential, and Log-logistic were compared using Bayesian Information Criterion (BIC) and the lognormal model emerged the best with the lowest BIC[4]. In their study observed that, according to AIC and area under ROC curves (AUC), the "Lognormal" parametric model, was identified as the best fitted and efficient model in the analysis of the effective factors in the event time of Diabetic Neuropathy. However, parametric models are selected as the best fitted models in survival analysis. As a result, various models show different effectiveness in the analysis of different data sets[5]. Vallinayagam et al applied some parametric models namely; Exponential, Weibull, Gompertz, Lognormal and Log-logistic to Breast cancer data, they compared the models to assess the efficiency of the models and the results showed that lognormal model performs better than other models using likelihood ratio test and standard error as selection method[6]. Studied some parametric models and compared them using log likelihood, AIC and R^2 type statistics. He then concluded that the data is adequately fitted by lognormal accelerated failure time (AFT) models as AIC of lognormal model is less than that of log-logistic model and suggest that results from AFT models are more easier to interpret not only for herpetologists but also for other clinicians for more appropriate explanation of survival[7]. Considered some accelerated failure time models and proportional hazard models all been a parametric model. They observed that based on

Cox-snell residuals and AIC, the exponential and Gompertz models were more efficient than other accelerated failure time models. They concluded by saying that although most cancer researchers tend to use proportional hazard model, accelerated failure time models in analogous conditions as they do not require proportional hazards assumption and consider a parametric statistical distribution for survival time will be credible alternative to proportional hazard models[8]. Applied some selected parametric models to survival of hermadialysis patients and concluded after they compared the performance of models that the Weibull model seemed to show the best fit among the parametric models of the survival of hemadialysis patient with the lowest AIC value[9]. Compared different parametric model; exponential, Weibull, generalized gamma, log-logistic and lognormal using maximum likelihood goodness of fit method and noticed that generalized gamma and Weibull Accelerated failure time (AFT) models were fitted to data too but do to some advantages Weibull model was used[10]. Carried out a study on breast cancer patients in a teaching hospital Osogbo and compared four parametric models that were considered such as Exponential, Weibull, log-Logistic and Lognormal distribution and the result showed that lognormal survival model which has the lowest AIC value was the considerably best model[11]. Based on Akaike's Information Criterion, the Weibull model with lower Akaike value is the most favorable for survival data after comparing five parametric models (exponential, weibull, lognormal, log-logistics and gamma distribution)[12]. Applied some parametric models (Exponential, Lognormal, Gamma, Weibull, Log-logistic and Gompertz) in a credit risk portfolio data to estimate the probability of default which can be used for evaluating the performance of a sample of credit risk portfolio. The results shows that the Gompertz distribution is the best parametric model for predicting the probability of default of the credit portfolio using several goodness of fit criteria namely; MSE, AIC and BIC[13]. Analyzed the survival of patient with CRC (colorectal cancer) admitted to Taleghani hospital by Comparing the 4- parameter log-logistic and log-logistic model to choose the fitted model to the dataand he observed that according to AIC, the 4-parameter log-logistic model was preferred to other models[14]. Studied three parametric survival models; Exponential, Weibull and Log-logistic models using AIC, BIC and log-likelihood value. When the three distributions were compared, Log-logistic distribution has the smallest value in AIC and BIC while it has the biggest value in log-likelihood compared with Exponential and Weibull distribution. Hence they concluded that the log-logistic distribution is the best fitted parametric model for the breast cancer patient's data[15].

Based on the literature reviewed, there has many researches on parametric modeling for the survival of cancer patients and there has been a gap in terms of using Hypertabastic distributions to model survival of patients. As a result of the gap, this work tends to channel its research in another area of health issue. Therefore, the aim of this study is to find a suitable distribution for modeling the survival of hypertensive patients.

Methods

Parametric survival models

Parametric model indicates that the outcome assumed to follow some family of distributions of similar form with unknown parameters. It is only when the value of the parameters is known that the exact distribution is fully specified. For parametric regression models; the data are typically used to estimate the values of the parameters that fully specify that distribution. A parametric survival model is one in which survival time (the outcome) is assumed to follow a known distribution. Survival estimates obtained from parametric survival models typically yields plots more consistent with a theoretical survival curve (Kleinbaum and Klein 2012). If the investigator is comfortable with the underlying distributional assumption, then parameters can be estimated that completely specify the survival and hazard functions. This simplicity and completeness are the main appeals of using a parametric approach. In this research, six survival distribution will be considered namely; Weibull, Exponential, lognormal, log-logistic, Gompertz and Hypertabastic distributions to model survival hypertensive patients.

Suppose the random variable T is the survival time of patients, T can be described in four standard ways; density function, survival function, distribution function and hazard function.

Let a non-negative continuous random variable T denotes the survival time that is measured from the time origin to an event of interest. Suppose that T has a density function that is given by

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t)}{\Delta t} = -\frac{d}{dt} S(t) \qquad 1$$

The survivor function S(t) gives the probability that a person survives longer than some specified time t: that is, S(t) gives the probability that the random variable T exceeds the specified time t.

$$S(t) = P(T > t) = 1 - F(t) \qquad 2$$

The cumulative distribution function (CDF) over the time interval (0, t), denoted by F (t), represents the probability the random variable T takes from time 0 to time t (t = 0,1, . . . , ∞) , given by

$$F(t) = P(T \le t) = \int_0^t f(u)du \qquad 3$$

The probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval, or as the limit of the probability that an individual fails in a very short interval, $t + \Delta t$, given that the individual has survived to time t. The hazard function is denoted by h(t), where Δt denotes a small interval of time. It is also known as instantaneous failure rate, force of mortality, conditional mortality rate, and age-specific failure rate.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t / T \ge t)}{\Delta t} = \frac{f(t)}{S(t)} \qquad 4$$

The cumulative hazard function is given by

$$H(t) = -lnS(t) = \int_0^t h(u)du \qquad 5$$

***Table 1:** Six different parametric models and their Functions*

Model	Parameters	Probability density function $f(t)$	Survival function $S(t)$	Hazard function $h(t)$	Cumulative hazard function $H(t)$
Exponential	λ	$\lambda e^{-\lambda t}$	$e^{-\lambda t}$	λ	λt
Weibull	$\lambda, \tilde{P}$	$\lambda \tilde{P} t^{\tilde{P}-1} exp[-(\lambda t)^{\tilde{P}}]$	$exp[-(\lambda t)^{\tilde{P}}]$	$\lambda \tilde{P} t^{\tilde{P}-1}$	$(\lambda t)^{\tilde{P}}$

Lognormal	μ,σ	$\frac{1}{t\sigma\sqrt{2\pi}}e^{-\frac{1}{2\left[\frac{logt-\mu}{\sigma}\right]^2}}$	$1-\Phi\left(\frac{logt-\mu}{\sigma}\right)$	$\frac{\frac{1}{t\sigma\sqrt{2\pi}}e^{-\frac{1}{2\left[\frac{logt-\mu}{\sigma}\right]^2}}}{1-\Phi\left(\frac{logt-\mu}{\sigma}\right)}$	$-logS(t)$
Log logistic	$\lambda,\hat{P}$	$\frac{(\hat{P}/\lambda)(t/\lambda)^{\hat{P}-1}}{\left[1+(t/\lambda)^{\hat{P}}\right]^2}$	$\frac{1}{1+(t/\lambda)^{\hat{P}}}$	$\frac{(\hat{P}/\lambda)(t/\lambda)^{\hat{P}-1}}{1+(t/\lambda)^{\hat{P}}}$	$-logS(t)$
Gompertz	h_o,r	$h_o\exp(rt)\,exp\left[\frac{h_o}{r}(1-e^{rt})\right]$	$exp\left[\frac{h_o}{r}(1-e^{rt})\right]$	h_oe^{rt}	$-logS(t)$
Hypertabastic		$sech[W(t)]\left(\alpha t^{2\beta-1}csch\left(t^{\beta}\right)^2-\alpha t^{\beta-1}\coth\left(t^{\beta}\right)\right)Tanh[W(t)]$	$\mathrm{sech}[\alpha\left(1-t^{\beta}\coth\left(t^{\beta}\right)\right)/\beta$	$\alpha\left(t^{2\beta-1}csch\left(t^{\beta}\right)^2-t^{\beta-1}\coth\left(t^{\beta}\right)\right)Tanh[W(t)]$	$-logS(t)$
Where Φ is the cumulative distribution function of standard normal distribution and $W(t)=\alpha\left(1-t^{\beta}\coth\left(t^{\beta}\right)\right)/\beta$ and csch is cosecant.					

Model Selection

Some assessing criteria will be considered in this research in order to select the best parametric model for modeling survival of hypertensive patients. These criteria includes; Log likelihood, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Akaike Information Criterion (AIC)

Akaike information criterion (AIC) is based on the maximum likelihood estimates of model parameters but penalizes a sample with large size[13]. The AIC is used to compare and find the best fitted model. The AIC statistic is computed as follows:

$$AIC=-2lnl+2p \qquad 6$$

where lnl is the log-likelihood of the estimated model, p is the number of parameters. Low values of the AIC suggest a better model.

Bayesian Information Criterion (BIC)

The Bayesian information Criterion is based on the maximum likelihood estimates of the model parameters which penalizes a sample with large size and large number of parameters[13]. The BIC statistic is computed as follows:

$$BIC=-2lnl+kp \qquad 7$$

where lnl is the log likelihoodof the estimated model, p is the number of parameters, and $k=logn$ and n observations. The larger the number of parameters p in a distribution, the larger the log-likelihood$l$.

Standard Error (SE)

The standard error (SE) is usually estimated by the sample estimate of standard deviation of population (sample standard deviation) divided by the square root of sample size (assuming independence). The standard error is given by

$$SE = \frac{stddeviation}{\sqrt{n}} or \sqrt{\frac{variance}{n}} \qquad 8$$

Results

Source and Sample size of the Data

The data in this work is a real life data from Chukwuemeka Odumegu Ojukwu University Teaching Hospital (COOUTH) Awka, Anambra State where patients who are diagnosed of hypertension from 1st January 2013 to 31st July 2018 are considered. The data consist of patient's age, sex, status and time of follow-up where death is the event of interest. When there is a death, the status is 1 and when there is lost to follow-up (censored) the status is 0. 105 patients are included which among them 69 has died and 36 are alive.

The descriptive statistics for the data is presented in table 2 which comprises of mean, median, mode, standard deviation, variance and sum.

Table 2: *shows the descriptive statistics of the data and it is displayed below;*

Descriptive	status	Age	sex	Time
Mean	.66	65.01	1.57	42.16
Median	1.00	67.00	2.00	45.00
Mode	1	68	2	31
Std. Deviation	.477	15.083	.497	17.612
Variance	.227	227.510	.247	310.195
Sum	69	6826	165	4427

Table 3 provides the AIC, BIC, SE and log likelihood of the six parametric models and the best model is chosen based on the smallest AIC, BIC and SE.

Table 3: *Goodness-of-fit criteria for parametric models*

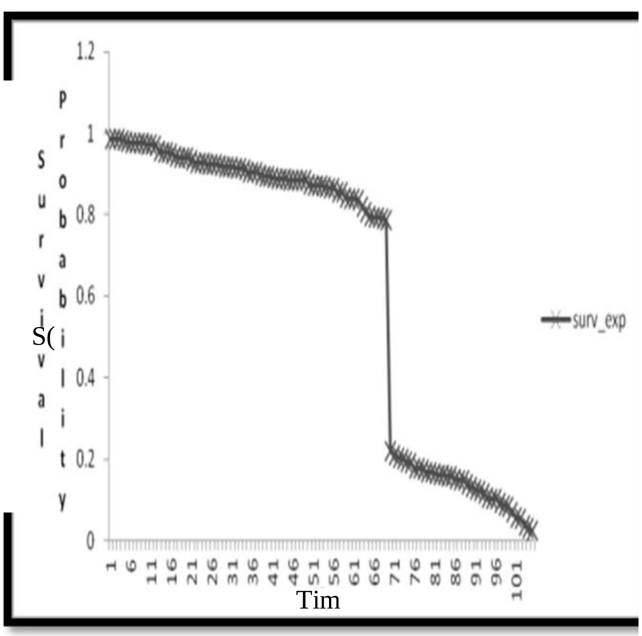
Models	Log likehood	AIC	BIC	S.E
Exponential	-355.7188	715.4377	720.7456	0.2410
Weibull	-333.1176	672.2352	680.1971	0.1069
Lognormal	-347.8798	701.7596	709.7215	0.1672
Log logistic	-339.3508	684.7016	692.6634	0.1330
Gompertz	-327.8516	661.7031	669.6650	0.2411
Hypertabastic	-253.4563	512.9126	520.8745	0.0637

Based on the analysis, the hypertabastic distribution has the lowest AIC = 512.9126 and BIC = 520.8745. The Gompertz distribution is the second with AIC = 661.7031 and BIC = 669.6650, followed by Weibull distribution with AIC =672.2352 and BIC = 680.1971, and then Log logistic and Lognormal distribution. The Exponential distribution turns out to be the least of all the distribution. The standard error also indicates the hypertabastic model is better because it has the least value of standard error. This indicates that in terms of relative efficiency and parameterization the hypertabastic model is the best. From the analysis, we can say that the hypertabastic model is the best parametric model for predicting the survival of hypertensive patient's data.

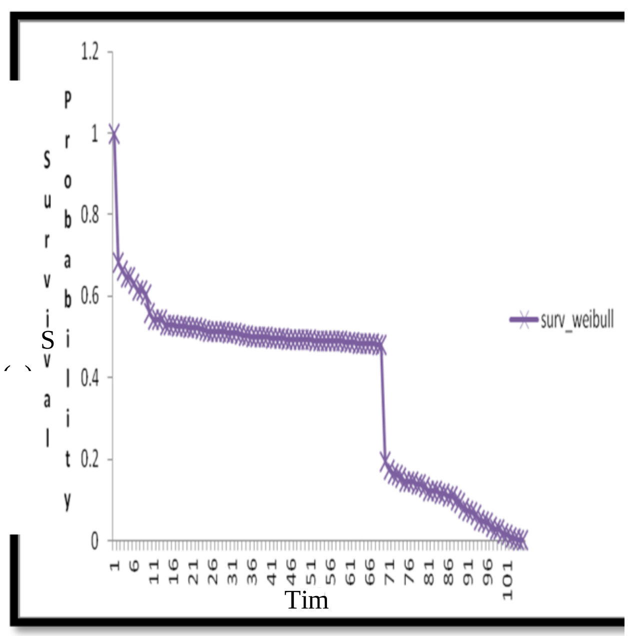
Graphical Representation of the Survival probability of the Six Models

Fitted survival probability distribution for parametric models. a) Exponential distribution, b) Weibull distribution, c) Lognormal distribution, d) Log logistic distribution, e) Gompertz distribution, f) hypertabastic distribution. The survival plot of the six parametric models is presented in figure 1.

(a)

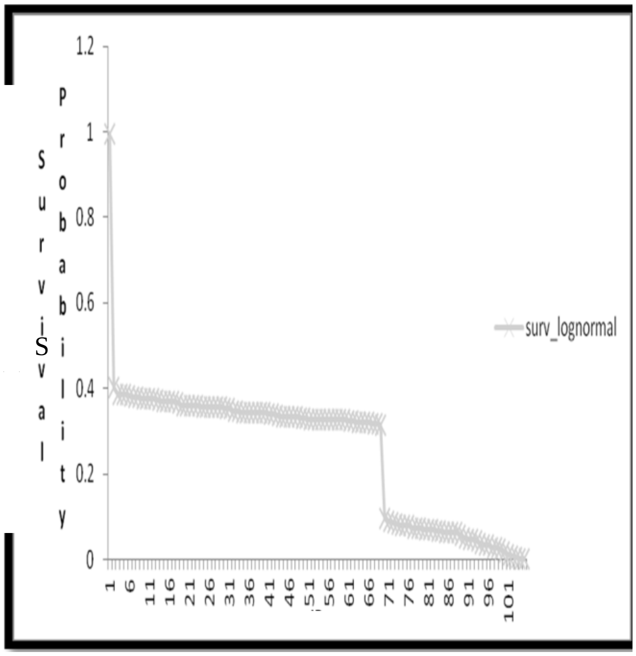


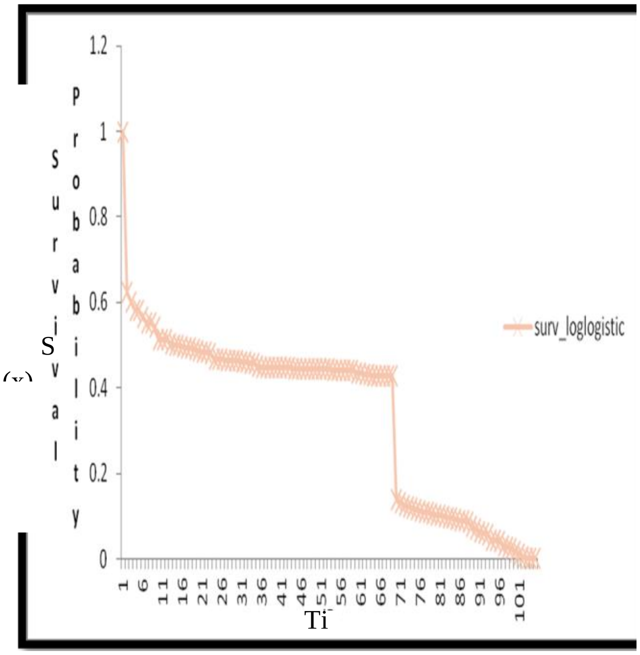
(b)



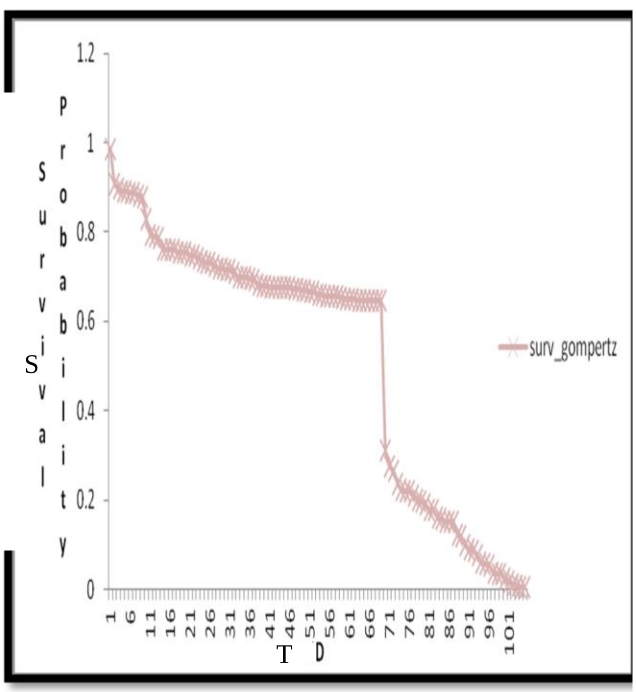
(c)

(d)





(e) (f)



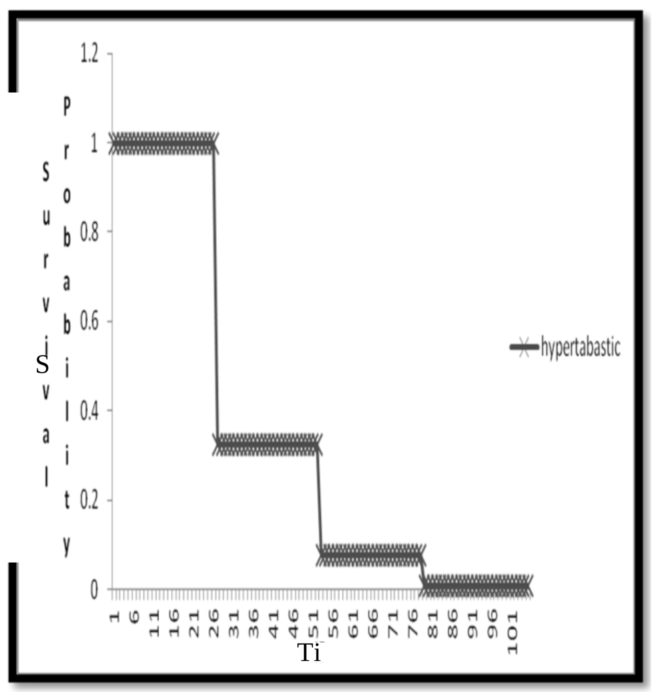


Figure 1: *survival plot of the six parametric models*

The Survival Probability Plot of the six parametric models presented above shows that the Hypertabastic distribution performs better because it shows a clear step function than the other distribution and this justifies the result AIC and BIC presented.

Conclusion

This paper focused on comparison of six different parametric models; exponential, weibull, lognormal, log logistic, gompertz and hypertabastic which was applied in hypertensive patients in a teaching hospital located in Awka. The analysis was carried out using R software in which the hypertabastic model turns out to be better than other models with the lowest AIC and BIC values. Also This indicates that the hypertabastic model is the best parametric model for predicting the survival of hypertensive patients.

Acknowledgement

Thanks to Almighty God and to all those who contributed in one way or the other for the progress of this research. This work was supported and approved by Nnamdi Azikiwe University, Awka, Nigeria.

Conflict of Interest

None

Reference

1.KleinbaumD.G and Klein M. Survival Analysis: *A Self Learning Text, Third Edition.* New York: Springer-Verlag, 2012.
2. Collett D.Modeling Survival Data in Medical Research, first Edition.London, UK. Chapman Hall, 1994
3. Tabatabai M.A, Zoran B, David K.W and Karam S.P.*Hypertabastic Survival Model.*Theoretical Biology and Medical Modeling, 2007, 4: 40, pp 1-13.
4. Gardiner J. Survival Analysis: *Overview of Parametric, Nonparametric and Semi-parametric approaches and New Developments.*SAS Global Forum 2010.Statistics and Data Analysis. Pp 1-25..
5. Sadegh K.M, Shahnaz R, Jamileh A and Iraj H Comparing of Cox Model and Parametric Models In Analysis of Effective factors on Event Time of Neuropathy in Patients with type 2 Diabetes. Journal of Research in Medical sciences, 2007, 22:115, pp 1-8.
6. Vallinayagam V, Prathap S, Venkatesa P. *Parametric Regression Models in the Analysis of Breast Cancer Survival Data.* International Journal of Science and Technology, 2014, Vol. 3, No. 3, ISSN 2049-7318, pp 163-167.
7. Shankar P.K, Sreenivas V,Subrat K.A. Accelerated Failure Time Models: *An Application in the Survival of Acute Liver Failure of Patients in India.*International Journal of Science and Research, ISSN 2319-7064, Vol.3, Issue 6, 2014, pp 161-166.
8. Ali Z, Mostafa H, Mahmood M, Kazem M, Hojjat Z and Kouroush H.N. A comparison Between Accelerated Failure Time and Cox Proportional Hazard Models in Analyzing the Survival of Gastric Cancer Patients.Iran Journal of Public Health, 2015, Vol. 44, No. 8. PP 1095-1102.
9. Maryam M, Jamshid Y.C and Fateme E. Application of Parametric Models to a Survival Analysis of Hemadialysis Patients.Nephrourol, 2016, 8(6): e28738, pp 1-6.
10. Asrin K, Ali D, Kourosh S. Application of Accelerated Failure Time Models for Breast Cancer of Patients' Survival in Kurdistan Province of Iran.Journal of Cancer Research and Therapeutics.2016, Vol. 12, Issue 3, pp 1184-1188.
11. Philip O.A, Oladapo A.K, Oluwatosin R. *Parametric Modeling of Survival Analysis among Breast Cancer Patients in a Teaching Hospital, Osogbo.*Journal of Cancer Treatment and Research, 2017, Vol. 5, No.5, pp. 81-85.
12. Abolfrazl N, Jamshid Y.C, Iradj M, Hosien R and Alireza K. *Parametric Model Evaluation in*

*Examining the Survival of Gastric Cancer Patients Andits Influencing Factors.*Global Journal Health Science, 2017, Vol.9, No. 3, ISSN 1916-1736, pp 260-265.
13. Jaber J. Credit Risk Assessment Using Survival Analysis for Progressive Right – Censored Data: *A Case Study in Jordan.*Journal of Internet Banking and Commerce, 2017, Vol. 22, No. 1, pp 1-18
14. Soraya M, Ahmad R.B, Mohamad A.P and Ali A.K. Application of the Parametric Regression Model with the Four Parameter Log-logistic Distribution for Determining of the Effecting Factors on the Survival Rate of Colorectal Cancer Patients in the Presence of Competing Risks. Iran Red Crescent Medical Journal, 2017, Vol. 19 (6): e55609, pp 1-8.
15. Syahila E.A, Abdullah M, Kek S.L and Muhammad J. *Analysis of Survival in Breast Cancer Patients by Using Different Parametric Models.*journal of Physics: Conference Series 2017, 890, 012169, pp 1-7.