FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 7 No. 5, October, 2023, pp 375 - 379
DOI: https://doi.org/10.33003/fjs-2023-0705-1980

COMPARATIVE ANALYSIS OF COX MIXED CURE MODEL WITH PARAMETRIC MODELS USING TOXIC DATA

[1]Usman, U., [2]Suleiman, S., [3]Dogondaji, A. M, *[1]Usman, M.

[1]Department of Statistics, Usmanu Danfodiyo University, Sokoto, Nigeria
[2]Department of Mathematics and Statistics, Federal University Dutsinma
[3]Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria

*Corresponding authors' email: mustystats@gmail.com

ABSTRACT

The investigation focused on examining of the survival analysis which entails the design and study of the occurrence and time of survival events. The study aimed to compare the result of the analysis using semi parametric model; cox mixed cure model and two parametric models; weibull and lognormal model to determine the model that fit the toxic data. The data was obtained from State Special Hospital Maiduguri (SSH) from 2016-2020. Akaike information criterion (AIC) was used to compare and evaluate the models. Results show that Cox mixed cure model has 13.65 with least AIC value, Weibull with 18.29 and lognormal with 18.30 with highest AIC value. The study concludes that semi parametric cox mixed cure model is the suitable to fit the toxic data.

Keywords: cox missed cure model, censoring, survival, Akaike Information Criterion

INTRODUCTION

Graves' disease is an autoimmune disorder that causes hyperthyroidism, or overactive thyroid (NIDDK 2017). With this disease, the immune system attacks the thyroid and causes it to make more thyroid hormone than the body needs. Thyroid hormones regulate the body's energy utilization, which has an impact on all bodily organs and even the heartbeat. Hyperthyroidism can lead to severe issues with the heart, bones, muscles, menstrual cycle, and fertility if treatment is not received. Untreated hyperthyroidism during pregnancy can cause health issues for both the mother and the unborn child. Graves' disease can also affect eyes and skin NIDDK (2017). Thus, this research focused on Graves' disease (toxic Goiter) and some variable that contribute towards the increase of the disease in Nigeria. It is reported that 3% of female and 0.05% of male were experiencing Graves in their lifetime (Adeleye *et al.,* 2020). Weight loss despite an increase in appetite, heat sensitivity, irritability, sleeplessness, sweating, hyperdfaecation, palpitations, muscle weakness, and irregular menstruation are among the basic signs of hyperthyroidism. Diffuse goiter, fine resting tremor, tachycardia, hyperreflexia, lid lag, warm, smooth skin, and proximal myopathy are examples of clinical symptoms. Depending on the situation, antithyroid medications, radioablative therapy, or thyroidectomy are among the treatment options typically provided to individuals with Graves' disease (Adeleye *et al.,* 2020). According to current studies, there are one or two cases of Graves' illness for every 1,000 people in England each year. Compared to the previously reported rate of roughly 0.3 occurrences per 1,000, this rate is significantly higher. Compared to men, women experience it far more frequently (Feingold *et al.,* 2015). Classical signs of hyperthyroidism include weight loss of greater than 6%, which is about equal to the combined effects of idiopathic hypothyroidism, Hashimoto's thyroiditis, and Graves' disease. (Feingold *et al.* 2015) It was discovered that 2.7% of women and.23% of males had either a history or a present case of Graves' illness. Additionally, 15% of women reported having goiter, 10.3% had antithyroid antibodies, and around two-thirds of women had hypothyroidism compared to Graves' illness. According to a recent report in this area, the annual incidence of 80 cases per 100,000 women is still present. According to the "HANES" study, the prevalence of hyperthyroidism is 0.7% at the sub-clinical level and 0.5% at the clinical level. Information confirms that there is a > 6% lifetime incidence of autoimmune thyroid disease, which is about equal to the sum of Hashimoto's thyroiditis, idiopathic hypothyroidism, and Graves' disease (Feingold, *et.al*. 2015). An growth of the thyroid gland is called a goiter. That is the gland located right below the Adam's apple at the front of the neck. It can just be a passing issue that goes away on its own, or it might be a sign of another thyroid illness that could be more serious and need medical attention. Since many things can make a thyroid swell, there are lots of types of Goiters. A few of them are: Simple Goiters, Endemic Goiters, Sporadic or nontoxic Goiters and Multinodular Goiters.

A Goiter is described as "toxic" when it's linked to hyperthyroidism. That means the thyroid makes too much thyroid hormone. A "nontoxic" Goiter does not cause ether hyperthyroidism or hypothyroidism (Nazario, 2021).

Elvan *et al*, (2010) compared five models (Gampertz, Gamma Log-logistics, Weibull and Log-normal) using breast cancer data. The result of the study showed that Gampertz model was the most suitable model to fit their data. Anet and Nestor (2011) carried out research on the application of hazard model for patient with breast cancer in Cuba using 6381 patients to compare six models (Gampertz, Exponential, Weibull, Log-logistics, Log-normal and Generalize Gamma). Their result showed that Gampertz model fit data well. Log-logistics distribution is an important survival parametric model that is used in the field of science, actuarial hydrology, survival analysis, reliability and the Economics (Al shorani, *et al* 2016). According to Bannett, (1983) although the structure of the log-logistics distribution and the log-normal distribution are extremely similar, the log-logistic distribution is more suited for survival analysis when data contains censored observations than the log-normal distribution. The lognormal model gives a fully stated probability distribution for the observations as well as a reasonable estimate of the variance explained by the model, a number that the Cox model is debatable about. The Cox and lognormal models' results are compared, and they appear to differ to some extent, it is concluded that, if the lognormal model correctly fits the data, it may be a useful method for analyzing censored

survival data. (Royston, 2001). Hui (2011) evaluated Weibull and Cox models in a study of stomach cancer patients and found that the Weibull model suited the data better than the Cox model.

All these above-mentioned authors have a difference in either using the models or the data in the diseases they used for the analysis compared to the present study. Thus, the study focused on semi-parametric, and parametric models for analysis and data used for research on Graves' disease (toxic). Toxic thyroid carcinoma should be suspected especially in male patients older than 50 years presenting with a typical, recurrent, highly symptomatic thyrotoxicosis. Patients with ToxTc seemed clinical resistant, therefore requiring higher amount of therapeutic radioiodine and presenting a higher clinical morbidity (Als *et al* 2002).

The two methods that contributed significantly to the development of the survival analysis. The first is Kaplan and Meier who introduced an estimator for survival probabilities. Second is Cox who introduced what is now called the Cox Proportional Hazard Model (CPHM), which is a regression model. Both models are heavily used to date and belong to the toolbox of every data scientist. Initially presented by Boag (1949) and Berkson & Gage (1952), the mixture cure model is a usually used tool for assessing both the treatment's cure rate and the survival rate of uncured patients.

The aim is to investigate the proportion of cured and uncured patients that have an impact on the Graves disease (toxic Goiter) that will be achieved through the Compare the Cox mixed cured models with two parametric models, that is; Weibull and log-normal models to determine the effect of individual characteristics on the cure rate of patients with Graves' disease (Toxic Goiter).

MATERIAL AND METHODS

This section presents the theoretical background of the models used in this research, it also explained the procedure for data collection and variables of the study, study area and method of data analysis will be discussed. In an attempt to achieve the main goal of this research, the researcher used four different models and R software to determine their efficiency. Model selection criteria provide a Useful tool, in this regard Akaike Information Criterion (AIC) is a technique that measures the goodness of an estimated statistical model and selects a model from a set of candidate models. AIC was introduced in 1973 by Hirotogu Akaike as an extension to the maximum likelihood principle. is given by the formula:

AIC = -2{ln (likelihood)} + 2k

Where,

Likelihood = the probability of the data in a given a model.

K = the number of independently adjusted parameters within the model.

Cox Mixed Cure Model

Let T represent the failure time of interest, $1 - \pi(z)$ the probability of a patient being cured depending on z, and S(t|x) the survival probability of the uncured patients depending on x. The observe values of two covariate vectors that may have an impact on the survival function are $x_1, x_2, \ldots, x_p$ and $z_1, z_2, \ldots, z_q$ The expression for the mixed cure model is as follows:

$$S_{POP}(t/x, z) = \pi(z)\, S(t/x) + 1 - \pi(z) \quad (1)$$

Usually, $\pi(z)$ is refer to as incidence and S(t|x) is refer to as latency. If the PH model is used to model the latency part, the cure model is called the PH mixture cure model.

If the proportion hazard model is used in S(t/x) estimation, the mixed cure model is described as the proportional hazard mixed model, and if accelerated failure model is used in S(t/x) estimation, the model described as the accelerated failure mixed cure model.

(Boag 1949) Studied the case when S(t/x) is modeled as log normal., (Farewell, 1982) studied the latency part (S(t/x)) is modeled by Weibull distribution., (Denham *et al*, 1996) studied another case with log normal assumed S(t/x)., The mixture cure model by assuming a proportion hazard model for S(t/x) with unspecified baseline hazard function (Peng & Dear, 2000) and (Sy & Taylor, 2000).

The advantage of the mixed cure model is that it enables the separate modelling of the cure distribution of both the cured and uncured individuals. Generally, using the logit function, the effect of z covariant is modelled with

$$\pi(z) = \frac{\exp(bz)}{1+\exp(bz)} \quad (2)$$

The b in the formula is the vector of unknown parameters.

The effects of z covariates modelled with $\log(-\log(1-\pi(z))) = bz$ by using log-log. When probity function is used, it is modelled as $(\pi(z)) = bz$. The term $\Phi(.)$ is the cumulative probability function of the standard normal distribution.

In computational method, Let $O = (t_i, \delta_i, z_i, x_i)$ denote the observed data for the *i*th individual $i = 1, \cdots, n$, where z_i, x_i are the possible covariates in the incidence and latency par respectively. We consider the censorship to be impartial land non educational. It is important to note that even though we use distinct covariate notations for the incidence and delay components, the same let

$\boldsymbol{\Theta} = (b, \beta, S_o, (t))$ denote the unknown parameters. To use the EM algorithm to estimate unknown parameters in this PH mixture cure model, let yi be an indicator of cure status of the *i*th patient, namely, $yi = 1$ if the patient is uncured and 0 otherwise, $i = 1, 2, \cdots, n$.. Obviously, if $\delta_i = 1$, $y_i = 1$; if $\delta_i = 0$, y_i is not observable and it can be one or zero. Note that $\pi(z) = P(yi = 1/\mathbf{z})$. Let $\mathbf{y} = (y1, y2, \cdots, yn)$ Therefore, **y** is partially missing information which will be employed in the EM algorithm. Given $\mathbf{y} = (y1, y2, \cdots, yn)$ and **O**, the complete likelihood function can be expressed as;

$$\Pi_{i=1}^{n}[1-\pi(z_i)]^{1-y_i}\ \pi\ (z_i)^{y_i} h(t_i|Y=1,x_i)^{\delta_i y_i}\ \mathrm{S}\ (t_i|Y=1,x_I)^{y_I} \quad (3)$$

Where $h(\cdot)$ is the hazard function corresponding to $S(\cdot)$. The logarithm of the complete

Likelihood function can be written as $lc(\mathbf{b}, \boldsymbol{\beta}; \mathbf{O}, \mathbf{y}) = lc1(\mathbf{b}; \mathbf{O}, \mathbf{y}) + lc2(\beta; \mathbf{O}, \mathbf{y})$.

Where,

$$lc1(\mathbf{b}; \mathbf{O}, \mathbf{y}) = \sum_{i=1}^{n} y_i log[\pi(z_i)] + (1-y_i)\log[1-\pi(z_i)] \quad (4)$$

$$lc2(\beta; \mathbf{O}, \mathbf{y}) = \sum_{i=1}^{n} y_i \delta_i log[h(t_i|Y=1,x_i)] + y_i \log[S(t_i|Y=1,x_i)] \quad (5)$$

Given the observed data O and current estimations of parameters $\Theta^{(m)} = (b^{(m)}, \beta^{(m)}, S_o{}^{(m)}(t)$ the E-step in the EM method computes the conditional expectation of the entire log-likelihood with regard to $y_i's$, This phase can be finished using the conditional expectation of y_i because (1.4) and (1.5) are both linear functions of y_i.The following describes the predicted value of $E(y_I, O, \Theta^{(m)})$.

$$w_i^{(m)} = E\left(y_I, O, \Theta^{(m)}\right)$$

$$= \delta_I + (1-\delta_I)\frac{\pi(Z_I)S(t_i|Y=1,x_i)}{1-\pi(z_i)+\pi(z_i)S(t_i|Y=1,x_i)} | (O, \Theta^{(m)} \quad (6)$$

It is easy to see that $w_i{}^{(m)} = 1$ if $\delta_i = 1$ and $w_i{}^{(m)}$ is the probability of uncured patients if $\delta_i = 0$. Thus, the another part of $E(y_i|\mathbf{O}, \Theta^{(m)})$ can be interpreted as the conditional probability of the ith individual remaining uncured. Because

$\delta_i \log w_i^{(m)} = 0$ and $w_i^{(m)} = \delta_i$, the expectations of (1.4) and (1.5) can be written as

$E(I_{C1}) = \sum_{i=1}^{n} w_i^{(m)} \log [\pi(z_i)] + (1 - w_i^{(m)}) \log[1 - \pi(z_i)]$ (7)

$E(I_{c2}) = \sum_{i=1}^{n} \delta_i \log[w_i^{(m)} h(t_i|Y = 1, x_i)] + w_i^{(m)} \log [S(t_i|Y = 1, x_i)]$ (8)

The Weibull Distribution

The Weibull model (invented by Waloddi Weibull in 1939) is a common descriptive two-dimensional model. The model's second parameter gives it more flexibility and a different risk function. The Weibull model's usefulness for performance work is determined in part by its flexibility, and in part by the ease with which risk and survival functions may be calculated. Weibull Distribution is useful in a wide range of situations. The most usefull methods which are considered to be the traditional methods are maximum likelihood and the moment estimation (Cohen and Whitten, 1982).

Survivals function

$s(t) = \exp(-\lambda t)^p$ Lawless(2011) (9)

The hazard function is $f(t) = \lambda p(\lambda t)^{p-1}$ (10)

Where $\lambda = \exp(x, \beta)$ (11)

Probability density function

$f(t) = \lambda p(\lambda t)^{p-1} \exp(-\lambda t)^p$ (12)

The expected duration from Weibull is

$E(T) = (\frac{1}{\lambda})^{\frac{1}{P}} \Gamma(1+\frac{1}{P})$ (13)

Where Γ denotes gamma function.

Where p and λ are the shape and scale parameters respectively.

Log-normal Distribution

Continuous probability distribution of a random variable whose logarithm is normally distributed is known as a log-normal distribution in probability theory. By the definition of Lognormal, if σ ln(X) has normal distribution X has Log-normal distribution. That is, if X is normally distributed exp (X) is log normally distributed. If Y is normally distributed with mean 0 and variance σ , then the random variable X defined by the relationship Y = log(X) is distributed as Log-normal, and is denoted as lognormal $(0 , \sigma^2)$

Y = ln(X) has a normal distribution if the random variable X is log-normal distributed (Eric W 2020).

$S(t) = 1 - \rho(\propto \ln(\lambda t))$ (14)

$h(t) = \frac{\propto}{\sqrt{2\pi 2}} \exp(\frac{-\propto^2 ln(\lambda t)^2}{2})^2 (1 - \rho(\propto \ln(\lambda t)))^{-1}$ (15)

$f(t) = \frac{\propto}{\sqrt{2\pi 2}} \exp(\frac{-\propto^2 ln(\lambda t)^2}{2})^2$ (16)

The expected life time and its are given by;

$E(T) = \exp(\mu + \frac{\delta^2}{2})$ (17)

where μ and ρ denote this distribution's scale and shape respectively.

RESULTS AND DISCUSSION

Comparison of the Cox mixed cure model with the Weibull and Lognormal

The result obtained from semi parametric model (Cox mixed cure model) and parametric models (weibull and lognormal) are compared using Akaike information criterion (AIC). The results for the models are shown below

Table 1: Weibull Model Result On toxic Goiter

Covariate-	HR	Standard Error	P-Value	95% C.I Min	95% C.I Max
Sex	0.7394	0.4662	0.0410	0.2211	1.2200
Age at diag	0.5237	0.0204	0.0148	-1.380	0.1666
T3	0.6275	0.4660	0.3810	0.2211	1.2200
T4	0.6679	0.0059	0.0164	0.3700	0.7125
TSH	0.4833	-0.0446	0.3333	-1.3400	0.1796

Log likelihood = -209.30311, Scale= 0.581

Loglik(model)= -63 Loglik(intercept only)= -65.9

Chisq= 5.75 on 5 degrees of freedom, p= 0.33

Table 2: Lognormal Model Result on Toxic Goiter

Covariate	HR	Standard Er	P-value	95% C.Imin	95%C.Imax
Age at. Diag	0.0146	0.0122	0.0155	0.433	0.780
Sex	0.2983	0.4059	0.3818	0.288	1.060
T3	0.0528	0.0456	0.0462	0.324	0.990
T4	0.0359	0.0271	0.0192	0.155	1.42
TSH	0.0027	0.0207	0.0358	0.564	0.580

(Intercept) 1.5764, 0.7277, 2.17 0.030

Log(scale) -0.3509, 0.1429 -2.46 0.014,

Scale=0.704

Table 3: Cox cure Model Result On toxic Goiter

Covariate	HR	Standard Error	P-Values	95%C.I.min	95%C.I.max
Age at diag	0.0330	1.0336	0,0258	0.1970	1.2900
Sex	0.8523	0.4264	0.6462	0.1870	1.3190
T3	0.0559	0.9455	0.0783	0.7140	0.4750
T4	0.9750	0.0252	0.0293	-0.859	0.3900
TSH	1.0695	0,0672	0.0581	1.1560	1.1560

Likelihood = 6.21

df, p=0.28

Table 4: Summary of the Results of the Analysis and Akaike Information Criterion (AIC) for Cox Mixed Cure Model, Weibull and Lognormal models

Covariate	Cox mixed M. HR (95% C. I.)	Weibull HR (95% C. I.)	Lognormal HR (95% C. I.)
Age at diagnosis	0.03(0.18-1.30)	0.74(0.21-1.22)	0.02(0.32-1.00)
Sex	0.85(0.19-1.32)	0.52(1.38-1.17)	0.33(0.25-0.26)
T3	0.06(0.71-0.46)	0.63(0.22-1.22)	0.03(0.52-0.65)
T4	0.98(-.86-0.39)	0.67(0.37-0.71)	0.03(0.24-1.12)
TSH	1.07(1.16-1.66)	0.48(1.34-0.18)	0.01(0.50-0.67)
AIC	**13.65**	**18.29**	**18.30**

Our investigation in table 1, (Weibull model) revealed that sex, age at diagnosis and thyroxine (T4) showed it is statistically significant which related to the finding of Bijan et al. (2008). Except Triiodothyronine (T3) showed it is insignificant. Also found that sex is insignificant in Cox mixed model and Log-normal models.

CONCLUSION

Based on the investigation of Comparative Analysis of the Cox Mixed Cure model with Parametric Models, it concluded that Cox mixed cure model is the suitable model for estimating survival of toxic goiter. It is also the efficient model to be used to minimize toxic goiter using semi parametric and parametric models. Cox-mixed cure models should be used to carryout Grave's Disease survival analysis, due to the fact that it gives the best fit for survival analysis

REFERENCES

Adeleye, J. O., Emuze, M. E., Azeez, T. A., Esan, A. Balagun, W. O. & Akande, T. O. (2020). Clinical Profile of Males with Graves'Disease: A Two Year Review in Tirtiary Hospital in Nigeria. *Journal of Clinical and Biomedical*, 1-3.

Als, C., Gedeon, P., Rosler, H., Minder, C,. Netzer, P. & Laissue J. A. (2015). Survival Analysis of 19 Patients with Toxic. *Medical Statistics*, 2.

Amico, M; keilegom, I. & Legran, C. (2018). The Single Index/Cox Mixed Cure Model. *medstats, 75*, 3.

Berkson, J. & Gage, R. P. (1952). Survival Curve for Cancer Patients Following Treatment. *American Statistical Association*, 501-512.

Boag, J. (1949). Maximum Likilihood Estimates of the proportion of patients curved by Cancer Therapy. *Royal Statistical Society, 11(1)*, 15-40.

Cohen, A. C. & Whitten, B. (1983). Modified Maximun Likelihood and Modified Momment Estimation for Three Parametric Weibull ditribution. *Comm. Stat Theory Method 2*, 128-150.

Cox, D. R. (1972). Regression Model and Life Tables. *Journal of Royal Statistical Socirty, B 34*, 187-220.

Elvan, A. H., Asli, s., Burak, U., Omar, D., Mehmet, N.O., & Gui, K. (2010). Comparison of Five Survival Models:Breast Cancer Registry, Data from Ege University Cancer Registry Centre . *Journal Med. 30(5)*, 78-81.

Farewell, V. (1982). The use of mixeture model for the analysis of Survival data with long-term survivors . *Biometrics 38*, 10411046.

Feingold, K. R., Anawait, B. & Boyce, A. (2015). Graves' Disease and Manifestation of Thyrotoxicois. *medstatistics*.

Frank, E-S. & Matthias, D. (2019). Introduction to Survival Analysis in Practice. *Medstatistics*, 20-25.

Gien, S. (2017). "Semi Parametric Model: Simple Defination and Example". *Statistics How to. Com*.

Golub, J. (2007). Survival Analysi and European Under Decision Making. *Euripean Union Politics*, 155-159.

Henry, B. B. & David, S. C. (2015). Management of Graves' Disease. *Clinical Review & Education, 314*, 1-3.

Hui, P. Z., Xin, X., Chuan, H. Y., Ahmed, A., Shun, F. L. & Yu, K. D. (2011). *Application of Weibull Regression Model for Survival with Gatric Cancer* (3RD ed.). John Wiley.

Judy, P. S., & Jeremy, T. (2000). Estimation in a Cox Proportional Hazards Cure Model. *pubmed*, 13-14.

Kaplan, E. L. & Meier, P. L. (1958). Non Parametric Estimation from Incomplete Observation *Journal of American Statistical Association*, 212-219.

Lawless, J. F. (2011). *Statistical Model for Lifetime Data*. John Waley & Sons.

Lee, H. (1982). On Clinical Trial and Survival Analysis. Singapore Medical Journal 23: 164-167.

Nazario, B. (2021, September 22). *WebMD*. Retrieved December 4, 2021

NIDDK (2017). National Institute of Diabetics Digestive and Kidney, Retrieved from 09-o1 2022, https://www.niddk.nih.gov/about-niddk

Nor, A. A., Wan Razita, M., Nor, A. M., Zainudin, M. A., Lailano, I., Saleha, N. I. T. Nasiru, M. & Muhammad, K. . (2013). Survival Rate of Breast Cancer Patient in Malaysia A Population-Based Study . *Asian Parcific Journal of Cancer Prevention*, 1-2.

Peng, Y. & Dear, K. B. G. (2000). A non Parametric Mixture Modelfor Cure Rate Estimation . *Biometric 56* , 237-243.

Peng, Y. W. (2003). Fetting Semiparametric Cure Model, Competational Statistics and Data Analysis. *Medicine statistics*, 481.

Perperoglou, A., Keramopoulos, A & Van Houwelingin, H. C. (2007). An Application to Breast Cancer. *Statistics in Medicine*, 26.

Rao, N. & C. R. (2004). *Handbook of Statistics 23: Advances in Survival Analysis.* Elseview.

Royston, P. (2001). The Lognormal Distribution as a Model for Survival Time In Cancer. With an Emphasis on Prognastics Factors. *Statistica Neerlandica 55(1)*, 89-104. doi:101111/1467-tesh9574. 00168

Sy, J. P. & Tailor,J.M.G. (2000). Estimation in a Cox Proportion Hazard Cure Model. *Biometric 56*, 237-243.

Umar, U. & Marafa, H. M. (2020). Comparative Analysis of the Cox Semi-parametric and Weibull Parametric Models on Colorectal. *Internation Journal of Data Science Analysis, 6, (1)*, 41-47. doi:10.11648/l.ijdsa.20200601.15

Weibull, W. (1951). Statistical Distribution of Wide pplicability: . *Journal of Applied Mechanics 18:*, 393-297.

Weisteir, E. W. (2020). "Log-normal Distribution". *In Mathwolrd A Wolfrom*. https://mathwolrd.wolfrom.com/LogNormal Distribution.html

You and Your Hormone (2021). Society for Endocrinology https://www.yo urhormones.info/about/