Semiconductor Optoelectronics, *Vol. 43 No. 1 (2024), 377-389*
https://bdtgd.cn/

Semiconductor Optoelectronics
ISSN: 1001-5868

A STUDY ON FITTING DIFFERENT FORMS REGRESSION MODELS IN CASE OF SURVIVAL DATA

Dr. Rinku Saikia
Assistant Professor, Department of Mathematics & Statistics, Golaghat Commerce College, Golaghat, Assam, India, Email. saikiarinku25@gmail.com

Abstract
Survival analysis is the statistical methodology which is used in case of censored observation. Censored means incomplete information of the study subject. In survival analysis, it is considered that the outcome variable of interest is time until an event occurs. In this study, an attempt has been made to fit Cox Proportional hazard (PH) model and compare the estimated value with Accelerated Failure Time (AFT) models having some probability distributions considering as exponential, weibull, log-normal, log-logistic etc. in the survival data of esophagus cancer patients. After fitting the models by using model selection criterion, the best fitted model is identified. The survival behaviour of esophagus cancer patients are observed by considering various demographic, socio-economic and disease factors by using the best fitted model. A sample of the esophagus cancer with survival data is collected from hospitals records. In survival analysis, the model comparison process which is also known as model selection process is mostly used to find the best fitted model. Some mostly used criteria are Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC), Cox-Snell residual plot etc. From the observation it is seen that Weibull AFT model is better fitted than other models. Different factors such as stage at the time of diagnosis, cancer directed treatment taken, socio-economic status of the patients etc. are found to contribute to survival time of esophagus cancer patients. The patients diagnosed in an early stage survive much more than the patients diagnosed at a later stage. The patients undergo cancer directed treatment other than surgery have lower survival time than surgery patients. The death risk is more in patients who are from lower and middle socio-economic group as compared to higher socio-economic group of patients. In this study, though age is not a significant factor in case of esophagus cancer, the patients belonging to older age groups have a higher risk of dying in comparison to the younger age group.
Keywords: Cox Snell residual, Cox PH model, survival, AIC, weibull etc.

Introduction:

Survival analysis is the statistical methodology which is used in case of censored observation. Censored means incomplete information of the study subject. In survival analysis, it is considered that the outcome variable of interest is time until an event occurs. Survival analysis is also introduced as the time to event analysis. In survival analysis, the term time indicates survival time, which is a non-negative random variable that measures the follow-up time from a well-defined starting point to the occurrence of a given event. Here, the term event means the occurrence of a particular incident to the study subjects in a statistical study. The event may be referred to as death, disease, job change, falling in love, marriage, divorce, cancer diagnosis, etc.

There are several areas such as medicine, engineering, health, social science, marketing, etc. where survival analysis can be used. According to its fields, it has various names, such as, in sociology, it is known as event history analysis; in engineering, it is known as failure time analysis; in economics, it is known as duration analysis or transition analysis. In industrial life testing and biomedical studies failure time analysis or survival analysis is mostly applied.

The study of survival analysis started from mortality tables centuries ago. The statistical methodology in the field of survival analysis was not properly developed until World War II. At the end of the war, various statistical methodologies were developed, such as, non-parametric methods, semi-parametric methods, parametric methods etc. The life table estimation method is the oldest non-parametric method for analyzing survival data, which was proposed by Berkson and Garge[1] in 1952. Cox Proportional Hazard (PH) model is the semi-parametric method which is now widely popular in survival analysis [2]. In the Cox PH model the covariates act multiplicatively on the hazard function. This Cox PH model is a widely popular model due to its ease of usage and there are no assumptions about the survival distribution. This Cox PH model presumes that the underlying hazard rate is a function of the independent covariates, but no assumptions are made about the shape or nature of the hazard function. There are some situations where this semi-parametric Cox PH model may not be appropriate. In that situation, Parametric models are much more reliable or appropriate than the semi-parametric and non-parametric method [3][4][5]. The parametric models are preferable if the functional form of the parametric model is completely known. The Accelerated failure time (AFT) model is the parametric model that can be used as another alternative of non-parametric or semi-parametric Cox PH model. The AFT models are most popular in industrial life testing but are rarely used in survival analysis.

Cancer is a leading cause of death and an important barrier to increasing life expectancy worldwide. The GLOBOCAN [6] 2020 estimates that there were 19.3 million new cases of cancer and nearly 10 million deaths, or nearly one in six deaths due to cancer in 2020. The most common cancer (in terms of new cases of cancer) are breast cancer, lung cancer, colon and rectum cancer, prostate cancer, skin cancer, stomach cancer and esophagus cancer. Among other cancers, esophagus cancer is also a very serious type of cancer. Esophagus cancer (EC) is ranked as the seventh most common cancer worldwide with over 570,000 new cases in 2018[7]. The American Cancer Society's estimates for esophagus cancer in the United States for 2022 are about 20,640 new esophagus cancer cases diagnosed (16,510 in men and 4,130 in women) and about 16,410 deaths from esophagus cancer (13,250 in men and 3,160 in women). Esophagus cancer is more common among men than among women. The lifetime risk of esophagus cancer in the United States is about 1 in 125 in men and about 1 in 417 in women [8]. According to Global cancer statistics 2020, in esophagus cancer over 604,100 (3.1%) new cases were reported and registered number of deaths was 544,076 (5.5%) of all sites. It is expected that by 2025, incidence of esophagus cancer is expected to rise by 140% [9].

The main objectives of this paper are: (i) to fit Cox PH model in case of esophagus cancer patients in North Eastern Region of India and compare the estimated value with various forms of AFT models. (ii) After fitting the best fitted model, the survival behaviour of esophagus cancer patients are observed by considering various factors such as sex, age at the time of diagnosis, location, socio economic status, stage of the cancer at the time of diagnosis, and different types of treatments taken by the patients.

Material and Methods:

In Assam Medical College Hospital (AMCH) Dibrugarh, Assam, North-East, this study was conducted from the medical charts of cancer patients. The study period of the research work was three years i.e., for the survival study, randomly selected esophagus cancer patients diagnosed between 1st January 2007 to 31st December 2008 were enrolled, and they were followed up to 31st December 2009 i.e., all the patients diagnosed with esophagus cancer during the first two years were included in the study. The third year was kept for follow-up of the patients.A total number of 178 patients were diagnosed with esophagus cancer during this study time in AMCH. For the collection of data, there is a pre-designed and pre-tested questionnaire. The demographic, treatment and disease profile of esophagus cancer patients were collected from hospitals records. It was treated as censored observation if the patients were alive beyond the follow-up period i.e.; 31st December 2009, or loss to follow up during the follow-up period and the patients were withdrawn from the study during the follow-up period or died due to other causes. After collecting the hospital records from AMCH, a re-verification of the information was conducted during the household visits to patients. During this study, the survival status of patients was also considered such as continuation of treatment taken by patients, socio-economic profile of patients and extension of the disease, that is, stage of the patients is also included. Survival time (in months) was estimated from the month of diagnosis until death, loss to follow-up, or the end of study period.

2.1 The Accelerated Failure Time (AFT) Model: Accelerated failure time (AFT) model is a class of parametric models used to analyze time to event data or failure time data. According to Lawless [10], accelerated failure time models (AFT) are useful in reliability theory and industrial experiments (which usually study the failure of equipments) and it is rarely used to analyze the survival data. The AFT model was first developed by Pike[11] in case of survival data. In case of AFT model, covariate effect is proportional (multiplicative) with respect to the survival time, while in Cox PH model effect of covariate is proportional (multiplicative) with reference to the hazard function. To estimate the impact of explanatory variables on survival data, maximum likelihood estimation method is used. The survival time of AFT model follows some specific probability distributions such as Exponential, Weibull, Lognormal, Log-logistic, Gamma, and many more.

The AFT model for ith study subject is expressed as

$$\log T_i = \mu + \beta_1 x_1 + \cdots + \beta_p x_p + \sigma\varepsilon_i \quad , \qquad (1)$$

Where $\log T_i$ represents the log-transformed survival time, $x_1, \dots\dots.. x_p$ are the explanatory variables with the coefficients $\beta_1, \dots\dots.. \beta_p$, ε_i is the residual term and it presumes a specific probability distribution and μ is the intercept and σ is the scale parameters respectively.

In fitting the AFT model, it is considered that for each error term i.e.; ε_i, there is a corresponding distribution for survival time T_i. If the error term ε_i, follows extreme value distribution then the survival time T_i follows the weibull distribution. Again, if the error term

ε_i follows logistic distribution, then the survival time T_i follows log-logistic distribution etc.

2.2 Exponential AFT model & Weibull AFT model: The exponential AFT model is a linear association between the logarithmic of the covariates and the logarithmic of hazard. If the ε_i, follows an extreme value distribution which is also known as Gumbel distribution then the survival time T_i follows the exponential distribution.

The Survival function of Weibull AFT model is expressed as

$$S_i(t) = \exp\left[-\exp\left(\frac{\log t - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right)\right] \ \ldots (2)$$

The cumulative hazard function and hazard function of Weibull AFT model can be determined directly from the survival function.

The cumulative hazard function of Weibull AFT model is presented as

$$H_i(t) = \log S_i(t) = \exp\left(\frac{\log t - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right) \ldots (3)$$

The hazard function of Weibull AFT model is given by

$$h_i(t) = \frac{1}{\sigma t}\exp\left(\frac{\log t - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right) \ldots \qquad (4)$$

Or $\quad h_i(t) = \lambda_i \sigma^{-1} t^{\sigma^{-1}-1}$

2.3 Log-normal AFT model:

If the error term(ε_i), of the regression model in the equation (1) follows the standard normal distribution then T_i has the log-normal distribution.

The survival function of log-normal AFT model is given by

$$S_i(t) = \ 1 - \phi\left(\frac{\log t - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right) \ldots \qquad (5)$$

The cumulative hazard function of Log-normal AFT model is

$$H_i(t) = -\log S_i(t)$$

$$= -\log\left(1 - \phi\left(\frac{\log t - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right)\right) \ldots \qquad (6)$$

2.4 Log-logistic AFT model:

If error term(ε_i), of the regression model in the equation (1) follows logistic distribution then T_i considers the log-logistic distribution.

The survival function of log-logistic AFT model is presented by

$$S_i(t) = \left\{\frac{1}{1 + e^{\left(\frac{\log t - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right)}}\right\} \ldots. \qquad (7)$$

The cumulative hazard function of log-logistic AFT model is given by

$$H_i(t) = -\log S_i(t)$$
$$= \log\left(1 + \exp\frac{\log t - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right) \ldots. \qquad (8)$$

In this study, to compare the models Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC), are used.

2.5 Cox Proportional Hazard (PH) Model: In 1972, D.R. Cox introduced a model which is known as the Cox PH model or relative risk model[2].

The Cox PH model is given by

$$\lambda_i(t/X_i) = \lambda_0(t)\exp(X_i\beta_i) \ldots \quad (9)$$

Where $X_i = x_1 \ldots\ldots..x_p$ are the p covariates and $\beta_i = \beta_1 \ldots\ldots\ldots\beta_p$ are the $p \times 1$ vector of regression coefficients, $\lambda_0(t)$ is the baseline hazard function which describes the risk for an individual with $x_i = 0$.

2.6 Akaike Information Criterion (AIC):

Akaike Information Criterion (AIC) is used to compare various models. The AIC was formulated by the statistician Hirotugu Akaike[12] in 1974. It is a measure of goodness of fit test of an estimated statistical model. The formula for the AIC is analysed as

$\mathrm{AIC} = -2(\mathrm{loglikelihood}) + 2(\mathrm{P} + \mathrm{K}) \ldots (10)$

Where P is the number of parameters in the model and K is the number of regression coefficients. The model which considers the smallest AIC value is said to be the best fitted model.

2.6. Bayesian Information Criteria (BIC): The Bayesian Information Criteria (BIC) is also used to measure the goodness of fit test and it is proposed by Schwarz [13] in 1978. The formula for BIC is given by

$$\mathrm{BIC} = -2(\mathrm{loglikelihood}) + (\mathrm{P} + \mathrm{K}) * \log(\mathrm{n}), \quad (11)$$

Where P is the number of parameters in the distribution, K is the number of coefficients and log(n) is the number of observations. The model which has the smallest BIC value is said to be the best fitted model.

2.7 Cox- Snell Residuals: Cox-Snell residual [14] plot checks the overall goodness of fit of the survival models. Cox-Snell residuals for a survival model is defined as,

$$\mathrm{r}_{\mathrm{c_i}} = \widehat{\mathrm{H}}_\mathrm{i}(\mathrm{t_i}) = -\log \widehat{\mathrm{S}}_\mathrm{i}(\mathrm{t_i}), \quad (12)$$

Where $\widehat{\mathrm{H}}_\mathrm{i}(\mathrm{t_i})$ and $\widehat{\mathrm{S}}_\mathrm{i}(\mathrm{t_i})$ are the estimated values of the cumulative hazard and survival function of the i^{th} individual at $\mathrm{t_i}$. Cox and Snell residual plot is a plot of estimated cumulative hazard function (based on Cox and Snell residual and the censored data) versus the Cox and Snell residual. Or we may fit a Nelson-Aalen curve to Cox and Snell residual and compare it with the standard exponential curve. A straight line with unit slope and zero intercept will then point out that the fitted survival model is adequate. On the other hand, if the plot displays a systematic departure from a straight line, or yields a line that does not have approximately unit slope or zero intercept, it might imply that the model requires some modification.

1. **Results and Discussion:**

In the study of survival of esophagus cancer patients, a total of 178 cancer patients diagnosed between 1st January 2007 to 31st December 2008 were enrolled, and they were followed up to 31st December 2009. Among the patients, it is seen that the number of persons suffering from esophagus cancer from both areas (urban and rural) is almost same, which is 52.8% in urban areas and 47.3% in rural areas. The outline of demographic, disease and treatment of the esophagus cancer patients are presented in the following table 1.

Table 1: Demographic, Disease and treatment profile of the esophagus cancer patients

Characteristics	Frequency (%)	Characteristics	Frequency(%)

Location Rural Urban	 84 (47.2) 94 (52.8)	**Caste** General OBC ST SC TG Unknown	 66(37.08) 57(32.02) 37(20.79) 9 (5.06) 8 (4.50) 1 (0.562)
Sex Male Female	 120 (67.4) 58 (32.6)	**Cancer Directed Treatment** Surgery & others Other than Surgery No treatment	 49 (27.5) 106(59.5) 23(12.9)
Age Less than 50 50 to70 Above 70	 33(18.5) 118 (66.3) 27(15.2)	**Continuation of Treatment** Continue Discontinue No treatment	 137 (76.97) 18 (10.11) 23 (12.92)
Marital Status Unmarried Married Widow	 5 (2.80) 152 (85.39) 21(11.79)	**Socio-economic status** Lower Middle Higher	 24 (13.5) 133 (74.7) 21 (11.8)
Stage Localized Regional Distant Unknown	 34 (19.1) 67 (37.6) 49 (27.5) 28 (15.7)		

The Cox PH model and parametric AFT models including Exponential, Weibull, Log-logistic, Lognormal, are fitted and by using the best fitted model, the effect of different demographic, treatment and disease characteristics such as location, age, sex, socio economic status, type of cancer directed treatments taken, stage at the time of diagnosis on the survival of esophagus cancer patients are analysed. With the help of various goodness of fit test such as AIC, BIC and Cox-Snell residual, the best fitted model is identified and it is shown in table 2.

Table 2: AIC and BIC values of Cox PH model and AFT models

Models	AIC	BIC
Exponential AFT model	490.82	529.00
Weibull AFT model	482.52	523.87
Log-Normal AFT model	488.83	530.20
Log-Logistic AFT model	483.32	524.69
Cox PH model	1214.45	1233.54

In table 2, a comparison is made between various forms of AFT models with Cox PH model and it is found that Weibull AFT model fits better than other AFT models. The AIC and the BIC values of Cox PH model are very much higher than other models and the values are 1214.45 and 1233.54 respectively.

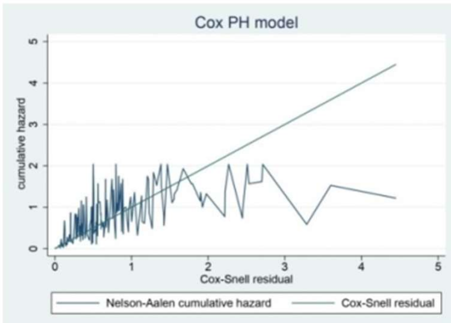

Fig.1. Cox Snell residual plot for Cox PH model

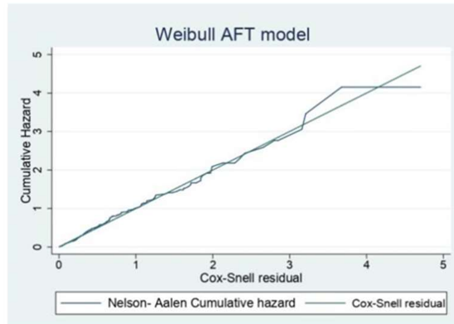

Fig.2. Cox-Snell residual plot for Weibull AFT model

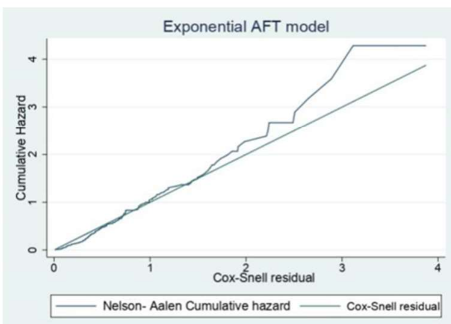

Fig.3. Cox-Snell residual plot for Exponential AFT model

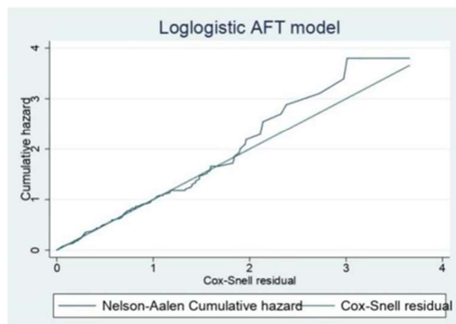

Fig.4. Cox-Snell residual plot for Loglogistic AFT model

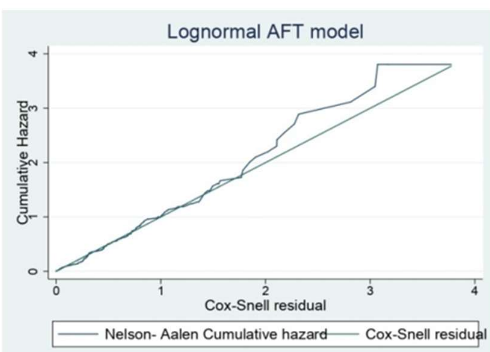

Fig.5. Cox-Snell residual plot for Lognormal AFT model

From the Cox-Snell residual plot, it is seen that Weibull AFT model fits better than other models. Since Weibull AFT model is better than the other models, so by using Weibull AFT model the survival behaviours of esophagus cancer patients are observed by considering various demographic, treatment and disease factors. The outcome of the Weibull AFT model is shown in the following table 3.

Table3. The outcome of Weibull AFT model:

Variable	Coefficient	SE	z value	P value	AF	95% CI of AF
Location						
Rural					Reference	
Urban	-0.21	0.14	-1.47	0.142	0.81	1.68 2.98
Age						
Less than 50					Reference	
50 to 70	0.029	0.19	0.15	0.877	1.03	1.92 4.07
70 and above	-0.39	0.23	-1.64	0.101	0.67	1.23 3.13
Sex						
Male					Reference	
Female	0.019	0.148	0.13	0.894	1.019	2.07 3.71
Socio Economic status						
Lower					Reference	
Middle	0.63	0.198	3.18	0.001	1.88	4.44 9.67
Higher	0.86	0.305	2.85	0.004	2.38	5.97 19.78
Cancer Treatment						
Surgery & others					Reference	
Other than Surgery	-0.407	0.185	-2.20	0.028	0.66	1.35 2.79
No treatment	-1.026	0.25	-4.01	0.000	0.35	0.866 2.36
Stage						
Localized					Reference	
Regional	-0.53	0.22	-2.40	0.017	0.58	1.15 2.78
Distant	-1.20	0.23	-5.03	0.000	0.30	0.84 2.16
Unknown	-0.83	0.257	-3.22	0.001	0.44	0.94 2.56
Intercept term	6.58	0.321	20.51	0.000		5.95 7.21
log(p)	0.22	0.069	3.24	0.001		0.08 0.35

p(shape parameter)	1.25	0.086				1.092 1.43
1/p (shape parameter)	0.799	0.055				0.69 0.91

*AF: Acceleration factor, CI= Class Interval, SE: Standard Error

The table 3, shows that in fitting AFT model with weibull distribution it get three forms of shape parameters and they are log(p), p and 1/p. The estimated value of p is 1.25.
Here, the null hypothesis is considered as $H_0:\log(p)=0$
And in this study, it is found that the value of the shape parameter is 0.22 and its p value is 0.001 so the null hypothesis is rejected.
The estimated acceleration factor for all the coefficients are shown in the table 3. The survival time is increased by acceleration factor of 1.88 and 2.38 for middle socio economic and higher socio-economic status of patients respectively compared to lower socio-economic status of the patients. The patients with high and moderate individual socio-economic status have lower risk for mortality than lower socio-economic group of people [28]. The survival time of the patients diagnosed in regional and distant stage are decreasing by the acceleration factors 0.58 and 0.30 respectively in comparison to the patients diagnosed in localized stage. The patients whose stage at the time of diagnosis remains unspecified is also the decreasing survival time by acceleration factor of 0.44 in comparison to the patients with localized stage. Patients diagnosed at an early cancer stage shows lower survival time than the other stages [27]. The survival time of patients who have taken the cancer directed treatment other than surgery and its combinations are decreasing by acceleration factor 0.66 in comparison to the patients who have taken surgery and other treatments. The patients who do not undergo any of the treatment have lower survival time in comparison to surgery and other treatments by acceleration factor 0.35. Similar studies indicates that the patients who received either chemotherapy, radiotherapy or surgery showed a better survival time compared with those who did not receive any treatment [27].
Though the factors – location, age and sex – of patients are not statistically significant, its acceleration factors have some impact on survival of esophagus cancer patients. In case of location of the patients, it indicates that urban patients have lower survival time than the patients from rural areas by the acceleration of 0.81. The survival time of patients from the age group 50 to 70 are increasing by the acceleration of 1.03 and the patients from age 70 and above are decreasing survival time with acceleration factor of 0.67. According to the gender of the patients, survival time accelerated for female patients than male patients by acceleration of 1.019.
From this analysis, it is seen that the factors such as stage at the time of diagnosis, cancer directed treatment taken, socio-economic status are found to contribute to survival time of esophagus cancer patients of Assam, North-East, India.
The AFT model is applicable for a comparison of survival time but the Cox PH model is applicable for the comparison of hazard function. In AFT model, the acceleration factor is the

key measure of relationship and in Cox PH model hazard ratio is the key measure of relationship. From the acceleration factor, one may be able to know how a change in covariate values changes in time scale from the baseline time scale. With the help of acceleration factor, one can determine the responsible factor for which the rate of failure is increasing or decreasing. It means the acceleration factor allows evaluation of the effect of predictor variables on survival time. But, in Cox PH model, the hazard ratio allows the evaluation of the effect of predictor variables on hazard. In AFT model, to estimate the impact of explanatory variables on survival data, maximum likelihood estimation method is used but in Cox PH model, partial likelihood method of estimation is used. It is known as partial because it considers the probability for those who fail and does not consider the probabilities for censored observation. In AFT model, if the acceleration factor is greater than 1, the effect of an exposure indicates that the exposure is beneficial to survival time but if the hazard ratio is greater than 1, the indication is that exposure is dangerous to survival. The AFT models are more trustworthy alternatives to Cox PH model. So, on the basis of the asymptotic results of AFT models, one can conclude that it leads to more efficient parameters than Cox PH model.

Conclusion:

This paper is an attempt to interpret the survival data of esophagus cancer patients who hail from the North Eastern part of India. The statistical technique Cox proportional hazard model is applied and it is compared to Parametric Accelerated failure time (AFT) models.

Stute [15],[16] suggested a new methodology for AFT models which are dealt with censored observation .Orbe et al.[5] studied the performance of Stute's model with respect to PH and AFT parametric models and found that Stute's method can be successfully applied where the assumption of Cox PH model also holds. Nardi and Scheme [17] compared Cox PH and parametric models in clinical trial studies and found that the Weibull model was better than other parametric model.

A simple linear regression considering log scale is more natural and gives better estimators for uncensored data [18]. AFT model gives possible and understandable estimates of the effect of important covariates on survival time. AFT model provides better prediction than the Cox PH model [19]. AFT models give easier explanation not only for herpetologists but also for clinicians [20]. Tolosie and Sharma [21] in their work studied Tuberculosis (TB) Patients of Ethiopia by applying Cox PH model. The main objective of the study was to identify the factors which influenced the survival of TB patients and found that the covariate age, TB patients' category, HIV, and age by HIV interaction are significant risk factors associated with death status of TB patients. Kargarian-Marvasti et al. [22] ;investigated the comparative performance of Cox PH model and parametric models in the survival analysis of factors affecting the event time of neuropathy in patients with type 2 diabetes. In their study, it was found that the log-normal model was the most efficient and fitted model. Wuryandari et al[23]; analyzes the duration of birth process using Cox PH model for durational data or survival data and found that the duration of birth process with gentle birth method is faster than the other method. Bustan et al. [24]; studied the inpatient Breast cancer data by using Cox PH model. Faruk[25] compared Cox PH model with AFT models in case of the 1st birth interval survival data and found that log-normal AFT model fitted better than other models. Goerdten et al. [26] compared Cox proportional hazards model and generalized Cox regression models applied in dementia risk prediction. It was seen that Generalized Cox PH model performed better than

Cox PH model in predicting dementia risk.

In this study, it is seen that the factor location that is the residence of the patients' sex does not have any influence on the survival of esophagus cancer patients. Though age is not a significant factor in case of esophagus cancer, the patients belonging to older age groups have a higher risk of dying in comparison to the younger age group. Here, the socio-economic status factor has a significant role in patients' survival. The death risk is higher in patients who are from lower and middle socio-economic group. The stage at the time of diagnosis is also an important factor in this study. The patients diagnosed in an early stage survive much more than the patients diagnosed at a later stage. The cancer directed treatments which are taken by the patients are also found to be significant factors. The patients undergo the cancer directed treatment other than surgery and the patients who have not taken any treatment experience a significantly higher risk of dying than patients who take treatments. After finding the factors which influence the survival of esophagus cancer patients by fitting Cox PH model, the analysis is compared with Parametric AFT model. In comparison, it is seen that the Weibull AFT model fits better than other models. So, from this paper, it is seen that if the parametric model is identified correctly, these parametric AFT models give a trustworthy result in comparison to the widely used Cox proportional hazard model.

Researcher in future may try to fit these models by using some other compatible distribution such as Generalized exponential, Skew-normal etc. which may provide better fit to the data. In accelerated failure time model the dependent variable is logT (logarithm of the survival time). There is immense scope for future study for fitting accelerated failure time model by using different strictly increasing function (exponential, sine etc.) as the dependent variable.

Reference:

[1] Berkson, Joseph and Gage, P.: Survival Curve for Cancer Patients Following Treatment, Journal of the American Statistical Association, 1952, 47 (259), 501- 515.

[2] Cox, D.R.: Regression Models and Life Tables, Journal of the Royal Statistical Society,Series B, (Methodological), 1972, 34 (2), 187-220.

[3] Wei, L.J.: The Accelerated Failure Time Model: A Useful Alternative to The Cox Regression Model In Survival Analysis, Statistics in Medicine, 1992 ,11 (14-15), 1871-1879.

[4] Kalbfleisch, J. D., & Prentice, R. L.: The Statistical Analysis of Failure Time Data, 2nd edition, John Wiley & Sons, 2002, New-York.

[5] Orbe,Jesus, F. Eva, & Nunez- Anton, Vicente: Comparing Proportional Hazards and Accelerated Failure Time models for Survival Analysis, Statistics in Medicine, 2002, 21(22), 3493-3510.

[6] World Health Organization (WHO). Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019, WHO, 2020. Accessed December 11, 2020.

[7] Yang J, Lin X, Cao Sai, Dong X, Rao S, Cai, K.: Understanding Esophageal Cancer: The Challenges and Opportunities for the Next Decade, Frontiers in Oncology, 2020, https://doi.org/10.3389/fonc.2020.01727.

[8] American Cancer Society. Facts & Figures, 2022, American Cancer Society, Atlanta, Ga. , 2022.

[9] Deepjyoti Kalita, Srinivas Bannoth, Joydeep Purkayastha, Abhijit Talukdar, Gaurav Das,

Pritesh Singh : A Study of Hormonal Receptors in Esophageal Carcinoma: Northeast Indian Tertiary Cancer Center Study, South Asian J Cancer, 2021, 9, 222–226.
[10] Lawless, J. F.: Statistical Models And Methods For Lifetime Data Analysis,1982,1st edition, Wiley, New York.
[11] Saikia, Rinku & Barman, Manash Pratim : A Review on Accelerated Failure Time Models, International Journal of Statistics and Systems, 2017, 12(2), 311-322.
[12] Akaike H : A New Look at the Statistical Model Identification, IEEE Transaction and Automatic Control, 1974, 19 (6), 716-23.
[13] Schwarz, G.: Estimating the Dimension of a Model, Ann.Statist., 1978, 6(2), 461-464.
[14] Cox D.R., and Snell E.J.: A General Definition of Residuals (with discussion), Journal of the Royal Statistical Society, Series B (Methodological), 1968, 30 (2), 248-275.
[15] Stute, W.: Consistent Estimation under Random Censorship when Covariables are Present, Journal of Multivariate Analysis, 1993, 45, 89-103.
[16] Stute, W.: Distributional Convergence Under Random Censorship when Covariables are Present , Journal of Statistics, 1996, 23, 461- 471.
[17] Nardi, Alessandra and Schemper, Michael : Comparing Cox and Parametric models in Clinical Studies, Statistics in Medicine, 2003, 22, 3597-3610.
[18] Mandal, Micha & Ritov, Ya'akov : The Accelerated Failure Time Model Under Biased Sampling, Biometrics, 2010, 66 (4), 1306-1308.
[19] Ponnuraja, C., & Venkatesan, P.: Survival Models for Exploring Tuberculosis Clinical Trial data-an empirical
Comparison, Indian Journal of Science and Technology, 2010, 3(7),755-758.
[20] Khanal, Shankhar P., Sreenivas, V., & Acharya, S. K.: Accelerated Failure Time Models: An Application in the Survival of Acute Liver Failure Patients in India, International Journal of Science of Research (IJSR), 2012, 3 (6), 161-166.
[21] Tolosie, K., & Sharma, M. K. : Application of Cox Proportional Hazards Model in Case of Tuberculosis Patients in Selected Addis Ababa Health Centres, 2014, Ethiopia, Tuberculosis Research and Treatment .
[22] Kargarian-Marvasti, S, Rimaz, S, Ablghasemi J, Heydari I : Comparing of Cox Model and Parametric Models in analysis of Effective Factors on Event Time of Neuropathy in Patients with type 2 diabetes, J Res Med Sci., 2007, 22- 115.
[23] Wuryandari, T., Kartiko, S. H., & Danardono : The Cox Proportional Hazard Model on Duration of Birth Process, (Coference presentation), (2018, 17th October). Semarang, Indonesia: Journal of Physics: Conference Series (JPCS).
[24] Bustan, M. Nadjib , Kasim Aidid, A.M., Gobel, F. A., & Syamsidar : Cox Proportional Hazard Survival Analysis to Inpatient Breast Cancer Cases, 2nd International Conference on Statistics, Mathematics, Teaching, and Research, Journal Of Physics:Conference Series,(2018, October 9-10) ,1028,Makassar, Indonesia.
[25] Faruk Alfensi : The Comparison of Proportional Hazards and Accelerated Failure Time models in Analyzing the First Birth interval Survival Data, Journal of Physics: Conference Series, 2018.
[26] Goerdten, Jantje, Carriere Isabelle, Muniz-Terrera, Graciela: Comparison of Coxproportional Hazards Regression and Generalized Cox Regression models Applied in Dementia Risk Prediction, Translational Research Clinical Interventions, 2020.

[27] Hassen, Y, Hamid, Teka, A. Mohammed & Addisse, Adamu: Survival Status of Esophageal Cancer Patients and its Determinants in Ethiopia: A Facility Based Retrospective Cohort Study, Frontiers in Oncology, 10, 2021.
[28] Chen,Hao-Yun, Chen, Chief, Chen, Yi-Huei, Chen, Chi-Chang, Chuang, Cheng –Yen & Lin, Ching-Heng: The Influence of Socioeconomic Status on Esophageal Cancer in Taiwan: A Population-Based Study, Journal of Personalized Medicine, 12, 2022.