An Analysis of the Survival of Gall Bladder Patients in a Tertiary Cancer Center in India using Accelerated Failure Time Models

Anurag Sharma[1,*] and Komal[2]

[1]*Department of Statistics, Ram Lal Anand College, University of Delhi, Delhi, India*

[2]*Department of Statistics, University of Delhi, Delhi, India*

Abstract: *Objective*: Accelerated Failure Time (AFT) models are an useful alternative of Cox- PH model to determine the significant predictors affecting the survival of the patients. This article aims to determine the significant prognostic factors of hospitalized Gall Bladder Cancer patients in Rajiv Gandhi Cancer Institute and Research Center, New Delhi, India by applying AFT Models. To the best of our knowledge, this is the first study to be carried out in India identifying the factors of Gall bladder patients using AFTM.

Materials and Methods: The data are taken from original proformae of 652 hospital admitted Gall Bladder patients from a tertiary care hospital from Delhi from the period January 2012 to December 2016. These models take the logarithm of survival time, S(t) as dependent variable and prognostic factors as independent variables. Thereby, effect of these prognostic factors is multiplicative and therefore these models can be easily interpreted. AFTM demonstrates the predictor's effect in terms of time ratio (TR). Analysis was implemented on R software version 3.5.1.

Results and Conclusions: In the Gall Bladder data considered in this article, shape of hazard function, H(t) and the exploratory data analysis falls in line with the Lognormal AFT model. AFT models give an estimate of Time Ratio which helps doctors, clinicians, epidemiologists etc. to determine the effect of treatment in terms of an increasing/decreasing survival time.

Keywords: Accelerated failure time models; Gall Bladder Cancer; time ratio; time to event data.

1. INTRODUCTION

Survival Analysis is defined as statistical method for data analysis where the outcome variable of is defined as the time till the occurrence of an event. The event may be death, disease incidence, recovery or any designated experience of interest that may happen to an individual in clinical trials [1-12]. Survival analysis is used in a number of fields for analyzing data involving the duration between two events. It is also known as event history analysis, lifetime data analysis, reliability analysis or life to event analysis [13].

Some examples of Survival analysis' problems include "the study of leukemia patients in remission over several weeks to see how long they stay in remission" or "how long patients survive after receiving a hair transplant" [14]. So, survival analysis deals with statistical methods for analyzing survival data derived from laboratory studies of animals, clinical and epidemiologic studies of humans and other appropriate applications, medicine, public health, social science, engineering etc. [15].

A very popular model in survival analysis which works well with these problems is the Cox- PH model. The Cox- PH model is used for the analysis of data in the presence of covariates or prognostic factors. The most specific feature of this model is that it does not depend on any assumption about the distribution of survival data but it assumes that the hazard function of the data is a function of the independent covariates [16].

Though, proportional hazard models are quite popular in analyzing and estimating the survival data, the proportionality assumption for these models is seldom met. There are very few regression models for which the parametric forms of PH models are defined such as exponential, Gompertz and Weibull distribution. Also, hazard functions of some distributions are very difficult to obtain as compared to survival function. So, Cox- proportional hazard models can be used with relatively few probability distributions [17].

In such situations, the AFT models can be used as a substitute of the PH model to analyze the survival data. Under this model, the effect of the explanatory variables is measured on the survival time instead of that of hazard function in the Cox Proportional hazard model. This allows to easily interpret the results as the effect of the corresponding prognostic factors on the mean survival time is measured [18]. Unlike, proportional hazard models, AFT models lies on the assumption that greater hazard rates are observed due to acceleration in the survival time whereas deceleration in the survival time is due to low hazard rates. Parametric AFT models can be defined for many distributions such as Exponential, Weibull, Log-Logistic, Log- normal, generalized gamma models etc. Some models such as Log-Logistic models, Lognormal and Generalized Gamma are the models which are defined only in AFT framework [19]

*Address correspondence to this author at the Department of Statistics, Ram Lal Anand College, University of Delhi, Delhi, India;
E-mail: anuragsharma532@gmail.com

The formulation of the AFTM permits the derivation of the acceleration factor which is also known as time ratio (T.R), which is easy to interpret as compared to a hazard rate [20, 21]. In this article, different parametric AFT models are fitted and best fitted model is chosen by performing two exploratory analysis, one using AIC and other by plotting Cox- Snell residual plots for Gall Bladder cancer. In our knowledge, no such study focusing on the determining significant risk factors of Gall Bladder cancer using Accelerated Failure Models has been conducted in India.

2. MATERIALS AND METHODS

The data of 652 hospital Gall Bladder patients admitted to Rajiv Gandhi Cancer Institute and Research Center from the period January 2012 to December 2016 are taken. Overall survival is defined as the duration from the date of diagnosis to the death of death/last contact and last follows up. Alive patients and lost to follow up patients are considered censored. Different AFT models are then applied and the model with least AIC is chosen as the best fitted model. Time ratios (T.R) and 95% confidence interval are presented for each predictor. P-value less than 0.05 are considered significant. Analysis was implemented on R software version 3.5.1.

3. SELECTION OF APPROPRIATE SURVIVAL MODEL

To determine the suitable survival AFT model, two highly used exploratory data analysis methods are used. The first method makes the use of baseline hazard function for the identification of the appropriate survival model. It is observed that the plotted baseline hazard function initially increases and then decreases (Figure **1**) of Gall Bladder data which is identical with the shape of H(t) of Log-Normal survival model. Thus, it can be considered that Gall Bladder data is best modeled by Lognormal Survival distribution.
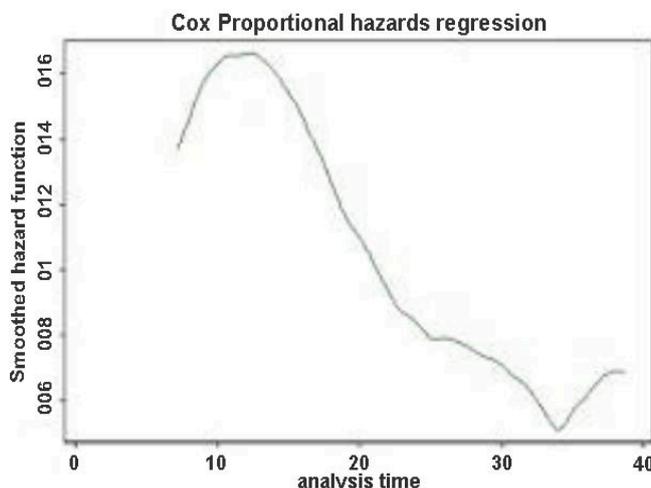

Figure 1: Baseline hazard function.

Second exploratory technique to determine the appropriate survival model is by plotting the adequately transformed survival function, S(t) with log S(t) as discussed below. Suppose, T is a random variable representing survival time following Log-Logistic distribution with scale parameter *θ,* shape parameter *k.* Then, PDF will be given by:

$$f_x(x) = \frac{1}{x}\frac{1}{\sigma\sqrt{2\pi}}exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), x > 0$$

The fitting of Log-normal survival model is furthermore investigated by making the survival function linear using the transformation as follows:

$$\Phi^{-1}\{1 - S(t)\} = \frac{\log t - \mu}{\sigma} = -\mu\sigma^{-1} + \sigma^{-1}\log t$$

Where, $S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$, and $\Phi(.)$ is the Standard Normal distribution.

Log-Normal survival model will be considered appropriate if a straight line is obtained by plotting the log-odds of S(t) against $\log_e(t)$.

Thus, Gall bladder survival time can be best fitted using Log Normal survival distribution as the baseline hazard plot is identical to the shape of hazard plot of Log-normal survival model. However, final choice is dependent on the goodness of fit using residual plots, AIC [22], R^2 statistic, log-likelihood etc.

4. THE AFT MODEL

AFT model determines the impact of different risk factors on deceleration or acceleration of survival time.

The survival function, S(t) for a class of patients with predictors $(x_1, x_2, x_3, \ldots \ldots \ldots x_n)$ can be written in the form of baseline survival function, $S_o(t)$ as :- $S(t) = S_o(\Phi t)$

Where,

$$\Phi = \exp\{(\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots \ldots \ldots \beta_p x_{pi})\}$$

is defined as the acceleration factor.

In order for better explanation of this concept, let us consider an example. Assume, a data set with one prognostic factor at two levels, "0" signifies the absence of predictor and "1" signifies the presence of predictor. The proportion of individuals who survived in the group with this predictor at the time point t_i is equal to the proportion of individuals surviving in the group without the predictor at any time $t_2 = \Phi t_1$,i.e, t_1/t_2 is constant. Φ is called acceleration factor.

In regression framework, this acceleration factor Φ can be parameterized as $\exp(\alpha)$, where α is a variable to be established from the data.

The relation between the log of S(t) and the set of prognostic factors is given by the log- linear form of the Accelerated Failure Time model as follows:-

$$\log(T_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots \ldots \ldots \beta_p x_{pi} + \sigma \varepsilon_i$$

where,

β_0 is the intercept of the model

$\beta_i's$ are the coefficients of "p" explainatory variables for i^{th} individual.

Scale parameter is represented by σ

ε_i is a random variable modelling the deviation of $\log_e(T_i)$ from the linear part of model.

For simplification and easy interpretation, $\exp(\beta_i)$ known as time ratio (T.R) should be reported on the same lines as hazard ratio is reported in the Cox P.H model. Time Ratio greater than 1 indicates that the prognostic factor increases the survival time whereas TR<1 represents that the presence of prognostic factor may cause early occurrence of event.

Two most commonly used method to fit the AFT model are MLE method and Newton- Raphson procedure [23, 24].

The distribution of ε_i and the corresponding distributions of T_i can be summarized as:-

Distribution of ε	Distribution of T
Extreme value (1 parameter)	Exponential
Extreme value (2 parameter)	Weibull
Logistic	Log- Logistic
Normal	Log- Normal
Log- Gamma	Gamma

[13, 23, 25] further explained the application of AFT models with practical examples.

Table 1:   Multivariate Analysis using Lognormal AFT Model

Variable		TR	95% CI	SE
Age	<30	1		
	>=30	0.878	1.104-1.256	0.03
Sex	Male	1		
	Female	1.345	0.931-1.049	0.08
Comorbidity	Nil	1		
	DM	0.745	0.455-1.346	0.05
	HT	0.954	0.515-1.441	0.02
	TB	0.873		
Marital Status	Unmarried	1		
	Married	1.126	0.831-1.215	0.06
Stage	I	1		
	II	0.761	1.052-1.244	0.09
	III	0.639	0.639-0.942	0.05
	IV	0.542	0.472-2.698	0.02
Treatment	CT	0.824	1.009-1.432	0.11
	RT	1		
	Surgery	0.776	0.605-0.940	0.04
T Category	T1	1		
	T2	0.932	1.676-2.148	0.06
	T3-T4	0.821	0.447-2.675	0.02
N Category	N0	1		
	N1-N2	0.678	0.567-3.199	0.1
	Unknown	0.821	0.965-1.239	0.07
M Category	M0	1		
	M1	0.783	0.792-1.035	0.05
	Unknown	0.641	0.080-4.233	0.03

4.1. Multivariate Analysis

Survival time was determined from the date of diagnosis to the date of death/last follow up/ Alive. Alive and lost to follow up patients were considered as censored. Different prognostic factors, viz., Age, Sex, Co-morbidities, Marital Status, Stage of the cancer, Treatment give and TNM stages are recorded. Multivariate analysis was conducted using R software. Table **1** shows the Time Ratio (TR) along with the standard errors and 95% confidence intervals of each of the prognostic factors.

4.2. Goodness of Fit of the Model

Cox- Snell residual plot is used to determine the goodness of fit of the AFT model (1). These residuals are determined by using cumulative hazard function and standardized residual as:

$$rs_i = \frac{log t_i - (\hat{\beta}_0 + \hat{\beta}_i x_i)}{\hat{\sigma}}$$

Where $\hat{\beta}_0, \hat{\beta}_i$ and $\hat{\sigma}$ are the MLE's of β_0, β and σ respectively.

Cox-Snell residuals for Log-Normal AFT models (25) is given by:-

$$rc_i = log\{1 + \Phi(rs_i)\}$$

Where $\Phi(.)$ is the CDF of the SND, i.e, Standard Normal Distribution.

Figure **2** represents the plot of Cox-Snell residuals for Lognormal AFT model. It can be seen that the plotted points follow a referent line. On the basis of the plot, it can be interpreted that Lognormal AFT model is a good fit to the Gall Bladder data.
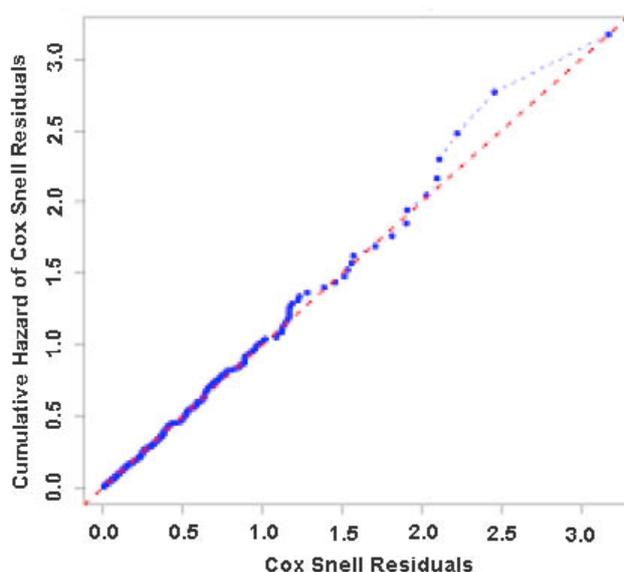

Figure 2: Plot representing the Cox Snell Residuals for Lognormal AFT model (Original).

Other methods to assess the goodness of fit of this AFT model comprise of Akaike's information criterion (AIC) and R^2 type statistic.

R^2 type statistic is determined as:

$$R_p{}^2 = 1 - \left\{ \exp\left[\frac{2}{n}(L_0 - L_1) \right] \right\}$$

Where

L_1 = log likelihood for the fitted model in presence of covariates,

L_0 =log likelihood for model without covariates.

The value of R^2 for Lognormal AFT model is 0.39.

Two survival models can be compared using their AIC values where AIC value is given by :-

$$AIC = -2L + 2(x + z)$$

where L represents the Log-likelihood value of the chosen model, x represents the No. of parameters of the fitted probability distribution & z are the No. of coefficients excluding constant in final model. Lesser the AIC, better the fitted model in comparison to other AFT models under consideration [22]. The AIC value for fitted Lognormal AFT model of the data under consideration is 2133.09.

5. DISCUSSION

In the fields of medical research, Cox-PH model is the foremost choice among different regression models for the estimation of survival data due to its convenience and familiarity [3, 26]. AFT models are traditionally used in reliability theory. Proportional hazards (PH) model is not necessarily a priori choice to AFT models [27]. AFT models are an useful alternate to Cox-PH model option when the primary interest is either the estimation of survival time or relative time to event are the interested association measures [28]. Clinicians may also benefit by the easy interpretation of acceleration factor. Furthermore, AFT models are of avid regard as they can specify a relation between logarithm of survival time S(t) and the prognostic factors which is in line to a multiple linear regression [29]. However, these models are estimated by assuming the survival time distribution. If the survival distribution of time is not known, then estimating these AFT models becomes questionable.

In our Gall Bladder data set, baseline hazard function plot and the exploratory analysis matches with the plot of Lognormal AFT model.

AFT models give the estimation of time ratio which helps clinicians/medical researchers to determine the benefit of treatment in terms of the effect survival time.

If the impact of predictors on S(t) is the primary objective of the study, AFT models is the best alternative if the user is able to recognize the distribution of S(t), survival time. To apply Accelerated Failure model, it exploration of Cox PH model is not necessary. Only thing one need to keep in mind is the identification of appropriate survival distribution to avoid miss classification. After determining AIC and R^2 statistics and Cox- Snell Residual plot, it can be concluded that the data is best fitted by Lognormal AFT survival model.

Therefore, the results from AFT models can be easily interpreted than Cox-PH model not only by clinicians but also by other hepatologists as it provides more appropriate justification and explanation of time to event or survival data. We hope that this study would motivate the use of AFT models among medical statisticians.

FUNDING SOURCE

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1]   Cox DR, Snell EJ. A general definition of residuals. J R Stat Soc Ser B 1968; 30(2): 248-65.
https://doi.org/10.1111/j.2517-6161.1968.tb00724.x

[2]   Crowley J, Hu M. Covariance analysis of heart transplant survival data. J Am Stat Assoc 1977; 72(357): 27-36.
https://doi.org/10.1080/01621459.1977.10479903

[3]   Andersen PK. Survival analysis 1982-1991: the second decade of the proportional hazards regression model. Stat Med 1991; 10(12): 1931-41.
https://doi.org/10.1002/sim.4780101208

[4]   Aalen OO. Medical statistics--no time for complacency. Stat Methods Med Res 2000 Feb; 9(1): 31-40.
https://doi.org/10.1191/096228000671850673

[5]   Kalbfleisch JD, Prentice RL. Survival models and data analysis. New York, NY: John Wiley; 1980.

[6]   Jacquillat C, Weil M, Auclerc MF, Schaison G, Chastang C, Harousseau JL, *et al*. Application of the study of prognostic factors to the treatment of childhood (less than 20 years old) acute lymphoblastic leukemia. Bull Cancer 1980; 67(4): 458-69.

[7]   Lawless JF. Statistical models and methods for lifetime data. Vol. 362. John Wiley & Sons; 2011.

[8]   Cox DR, Oakes D. Analysis of survival data. Vol. 21. CRC Press; 1984.

[9]   Urken ML, Weinberg H, Vickery C, Aviv JE, Buchbinder D, Lawson W, *et al*. The combined sensate radial forearm and iliac crest free flaps for reconstruction of significant glossectomy-mandibulectomy defects. Laryngoscope 1992; 102(5): 543-58.
https://doi.org/10.1288/00005537-199205000-00014

[10]   Van der Burg E, De Leeuw J, Dijksterhuis G. OVERALS: Nonlinear canonical correlation with k sets of variables. Comput Stat Data Anal 1994; 18(1): 141-63.
https://doi.org/10.1016/0167-9473(94)90136-8

[11]   Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B 1996; 58(1): 267-88.
https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[12]   Makridakis S, Hibon M. ARMA models and the Box-Jenkins methodology. J Forecast 1997; 16(3): 147-63.
https://doi.org/10.1002/(SICI)1099-131X(199705)16:3<147::AID-FOR652>3.0.CO;2-X

[13]   Hosmer Jr DW, Lemeshow S. Applied survival analysis: time-to-Event. Vol. 317. Wiley-Interscience; 1999.

[14]   Emmert-Streib F, Dehmer M. Introduction to survival analysis in practice. Mach Learn Knowl Extr 2019; 1(3): 1013-38.
https://doi.org/10.3390/make1030058

[15]   Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. Br J Cancer 2003 Jul; 89(2): 232-8.
https://doi.org/10.1038/sj.bjc.6601118

[16]   Chen C, Liu Y, Wang S, Sun X, Di Cairano-Gilfedder C, Titmus S, *et al*. Predictive maintenance using cox proportional hazard deep learning. Adv Eng Informatics 2020; 44: 101054.
https://doi.org/10.1016/j.aei.2020.101054

[17]   Stensrud MJ, Aalen JM, Aalen OO, Valberg M. Limitations of hazard ratios in clinical trials. Eur Heart J 2019.
https://doi.org/10.1093/eurheartj/ehy770

[18]   Parsa M, Van Keilegom I. Accelerated failure time vs Cox proportional hazards mixture cure models: David vs Goliath? Stat Pap 2022; 1-21.
https://doi.org/10.1007/s00362-022-01345-5

[19]   Khanal SP, Sreenivas V, Acharya SK. Comparison of cox proportional hazards model and lognormal accelerated failure time model: application in time to event analysis of acute liver failure patients in India. Nepal J Stat 2019; 3: 21-40.
https://doi.org/10.3126/njs.v3i0.25576

[20]   Patel K, Kay R, Rowell L. Comparing proportional hazards and accelerated failure time models: an application in influenza. Pharm Stat [Internet] 2006 Jul 1; 5(3): 213-24. A
https://doi.org/10.1002/pst.213

[21]   Kay R, Kinnersley N. On the Use of the Accelerated Failure Time Model As An Alternative to the Proportional Hazards Model in the Treatment of Time to Event Data: A Case Study in Influenza. Drug Inf J DIJ / Drug Inf Assoc [Internet] 2002; 36(3): 571-9.
https://doi.org/10.1177/009286150203600312

[22]   Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974; 19(6): 716-23.
https://doi.org/10.1109/TAC.1974.1100705

[23]   Collett D. Modelling survival data in medical research, third edition. Modelling Survival Data in Medical Research, Third Edition 2015; pp. 1-521.
https://doi.org/10.1201/b18041

[24]   Lemeshow S, May S, Hosmer Jr DW. Applied survival analysis: regression modeling of time-to-event data. John Wiley & Sons; 2011.

[25]   Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data Springer. New York 1997;

[26]   Cox DR. Regression Models and Life-Tables. J R Stat Soc Ser B [Internet] 1972; 34(2): 187-202.
https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

[27]   Kwong GPS, Hutton JL. Choice of parametric models in survival analysis: applications to monotherapy for epilepsy and cerebral palsy. J R Stat Soc Ser C (Applied Stat [Internet] 2003 May 1; 52(2): 153-68.
https://doi.org/10.1111/1467-9876.00395

[28]   Cox C, Chu H, Schneider MF, Munoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. Stat Med 2007; 26(23): 4352-74.
https://doi.org/10.1002/sim.2836

[29]   Orbe J, Ferreira E, Núñez-Antón V. Comparing proportional hazards and accelerated failure time models for survival analysis. Stat Med 2002; 21(22): 3493-510.
https://doi.org/10.1002/sim.1251