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

Anurag Sharma¹,* and Komal²

¹Department of Statistics, Ram Lal Anand College, University of Delhi, Delhi, India
²Department 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.

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 $\theta$, shape parameter $k$. Then, PDF will be given by:

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

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$_{10}(t)$.

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 $\{x_1, x_2, x_3, \ldots, x_n\}$ can be written in the form of baseline survival function, $S_0(t)$ as $S(t) = S_0(\Phi t)$

Where,

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

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 $t_1$ 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. $\Phi$ is called acceleration factor.

In regression framework, this acceleration factor $\Phi$ can be parameterized as $\exp(\alpha)$, where $\alpha$ 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:

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

where,

$\beta_0$ is the intercept of the model

$\beta_i$’s are the coefficients of “$p$” explanatory variables for $i^{th}$ individual.

Scale parameter is represented by $\sigma$

$\epsilon_i$ 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 methods to fit the AFT model are MLE method and Newton-Raphson procedure [23, 24].

The distribution of $\epsilon_i$ and the corresponding distributions of $T_i$ can be summarized as:

<table>
<thead>
<tr>
<th>Distribution of $\epsilon_i$</th>
<th>Distribution of T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extreme value (1 parameter)</td>
<td>Exponential</td>
</tr>
<tr>
<td>Extreme value (2 parameter)</td>
<td>Weibull</td>
</tr>
<tr>
<td>Logistic</td>
<td>Log-Logistic</td>
</tr>
<tr>
<td>Normal</td>
<td>Log-Normal</td>
</tr>
<tr>
<td>Log-Gamma</td>
<td>Gamma</td>
</tr>
</tbody>
</table>

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

Table 1: Multivariate Analysis using Lognormal AFT Model

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

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

Where \( \hat{\beta}_0, \hat{\beta} \) and \( \hat{\sigma} \) are the MLE’s of \( \beta_0, \beta \) and \( \sigma \) respectively.

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

\[ r_{c_i} = \log \{1 + \Phi(rs_i)\} \]

Where \( \Phi(\cdot) \) 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).](image)

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

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

Where

\( L_1 = \log \text{likelihood for the fitted model in presence of covariates}, \)
\( L_0 = \log \text{likelihood for model without covariates}. \)

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 :-

\[ \text{AIC} = -2L + 2(x + z) \]

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.
The impact of predictors on S(t) is the primary objective of the study. AFT models are 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 R² 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 epidemiologists 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.

CONFLICT OF INTEREST
None.

REFERENCES


© 2022 Sharma and Komal; Licensee Lifescience Global. This is an open access article licensed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution and reproduction in any medium, provided the work is properly cited.