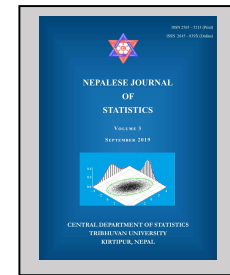# Comparison of Cox Proportional Hazards Model and Lognormal Accelerated Failure Time Model: Application in Time to Event Analysis of Acute Liver Failure Patients in India

Shankar Prasad Khanal[1]*, V. Sreenivas[2] and Subrat K. Acharya[3]

ABSTRACT

Background: Different survival analysis techniques such as nonparametric, semi-parametric, parametric Accelerated Failure Time (AFT) models have been generally applied to analyze time to event data. In order to identify the prognostic factors for survival of Acute Liver Failure (ALF) patients, previous studies applied Cox Proportional hazards (CPH) model, Lognormal AFT and Log-Logistic AFT model satisfying respective model's assumptions and goodness of fit of each model. However, comparison of CPH model and AFT model has not been reported so far for ALF data with short follow up time.

Objective: To compare CPH model and Lognormal AFT model based on different parameters for assessing the model performance and prospective validation of the finally selected model.

Materials and Methods: Altogether 1099 ALF patients' data from liver clinic of All India Institute of Medical Sciences, New Delhi India were analyzed based on the retrospective cohort study design. For validating the final model, a separate data set of 138 ALF patients from the same clinic was used. CPH model and Lognormal model's performance was assessed through selection of variables in the final model, R^2 type statistic, goodness of fit of the model, visual assessment of Cox-Snell's residuals plot and robustness of the model. The prospective validation of the over scored CPH model was done by comparing overall survival, regression coefficients, observed and predicted survival curves between original and validation data set.

Results: It is found that 60% of variation in the partial log-likelihood is explained by the CPH model whereas 39% of variation in full log-likelihood is explained by Lognormal AFT model. Cox-Snell residuals plot for CPH model seems less deviated from the line of ideal fit, replications of variables measured through bootstrapping resampling technique in CPH model are on the higher side, model predicted and observed survival curves in each risk stratum were closer than that of Lognormal model. The survival experience of original data and validation data set for CPH model does not seem to be very different (p = 0.07) at 5% level of significance.

Conclusion: Both CPH and Lognormal AFT model are found well fitted and can be applied either of them for this ALF data. While comparing the model performance, the CPH model for the identification of prognostic factors for the survival of ALF patients is found comparatively better.

Keywords: Acute liver failure, bootstrapping, Cox regression, goodness of fit, lognormal, robustness, validation.

_____

Address correspondence to the author: Central Department of Statistics, Institute of Science and Technology, Tribhuvan University, Kirtipur, Kathmandu, Nepal.
Email: drshankarcds@gmail.com (Corresponding author)[1*]; Department of Biostatistics, All India Institute of Medical Sciences (AIIMS), New Delhi India. Email: sreevishnubhatla@gmail.com[2]; Department of Gastroenterology, All India Institute of Medical Sciences (AIIMS), New Delhi India. Email: subratacharya@gmail.com[3]

_____

INTRODUCTION

In order to analyze time to even data, different survival analysis techniques such as Kaplan & Meier(K-M), log rank test, Cox Proportional Hazards (CPH) model, parametric hazards models such as exponential, Weibul, Gompertz; parametric survival models such as Lognormal, Log-Logistic, Gamma, etc have been generally used. Previously nonparametric techniques such as K-M estimate of survival function for estimating the survival probability, and to assess difference of the survival experience across different groups, log-rank test (Peto et al.; 1976, 1977) were found frequently used in clinical set up since these techniques do not require any distributional assumptions. CPH model is also most popular regression for the analysis of time to event data since the baseline hazard function embedded in this model, does not require to follow any probability distribution. It means this model does not require to specify any probability distribution in order to quantify the effects of independent variables. If one is able to recognize the probability distribution of the baseline hazard function in the CPH model, then it becomes parametric proportional hazards model. Then, the maximum likelihood estimates and inferences based on the parametric proportional hazards model are expected to be more precise and worthy statistically.

However, the proportional hazards (PH) models are based on PH assumption and this assumption may not be realistic in all situations. If this assumption does not hold, the application of standard PH models may entail serious bias and loss of power when estimating or making inference about the effect of a given prognostic factor on mortality (Abrahamowicz, MacKenzie, & Esdaile, 1996; Hess, 1994). One survival analysis method that overcomes the non-proportionality of hazards is stratified PH model. In this method, the variable for which the hazards of the event are not proportional, the analysis is carried out in each strata of this variable and combined to have overall picture. However, the effect of the stratification variable itself cannot be assessed in this approach. Another method of overcoming the non-proportional hazards is the Accelerated

Failure Time (AFT) model in which the relation between the logarithm of the survival time and the explanatory variables is expressed just like a multiple linear regression. Usually the estimation of these models requires knowledge on the distribution of survival times and if the distribution could be guessed correctly, the estimates obtained through AFT models are expected to be more precise as compared to those of non-parametric or semi-parametric survival techniques.

Literature search indicated that AFT models have been found applied frequently in industrial research. Though they are equally applicable in clinical research, application of AFT modeling is not frequent (Kay & Kinnersley, 2002). Further, comparison of different survival models such as semi-parametric (CPH model), parametric hazards and AFT models is limited, more so in clinical settings. Till date, in most of the studies related to survival of Acute Liver Failure (ALF) patients has been analyzed by using K-M method, logistic regression model and CPH model. Khanal, Sreenivas, and Acharya (2018) has also applied CPH model to identify the prognostic factors of ALF patients in India. With the best of our knowledge, for the first time, the survival analysis of ALF data had been found analyzed by applying AFT models and their comparison was also reported by Khanal, Sreenivas, and Acharya (2014). While comparing Log-Logistic and Lognormal AFT models in the survival analysis of ALF patients with the occurrence of outcome in short gestation period, it was little to choose between these two models. Nonetheless, Lognormal AFT model seems to be slightly better since AIC of lognormal AFT model is lesser than that of Log-Logistic AFT model for the analysis of ALF data (Khanal et al., 2014). In this communication, attempt has been made to compare the CPH model and Lognormal AFT model on the basis of variable selection in the final model, goodness of fit, R^2 type statistic and frequency of the variable selection in the model through bootstrapping method. Further, we have also attempted to validate the finally selected model with another separate set of ALF data.

Acute liver failure is a kind of dangerous rare liver injury among all liver diseases. No comprehensive information is available on ALF globally. Almost all reports available talk of liver disease in general or specific conditions such as cirrhosis of liver and hepatocellular carcinoma. The frequency of occurrence of ALF seems to be relatively less but the mortality is very high. There are 2,300 to 2,800 patients with ALF in a year in the United States (Kim, Brown, Terrault, El-Serag, 2002). Viral hepatitis, caused by hepatitis A through E, is a major public health problem in India. Based on the data obtained from Indian hospitals, it is found that about 250,000 people die annually due to viral hepatitis or its sequelae (Acharya, Madan, Dattagupta, & Panda, 2006). In India, hepatitis E virus (HEV) infection is responsible for 30%-70% of cases of acute sporadic hepatitis and is the major cause of ALF (Acharya et al., 2006).

MATERIALS AND METHODS
Data and study design
Altogether 1099 ALF patients' data from liver clinic of All India Institute of medical Sciences (AIIMS) New Delhi India admitted during the period of 25 May 1986 to 31 December 2005 were

analyzed based on the retrospective cohort study design. The outcome of interest variable in the analysis is the survival status of the ALF patients (coded 0 for survived and 1 for experiencing death). Variables were selected from the total 15 independent covariates for each final model under consideration. The details of the variable selection, data quality assessment, etc have been explained (Please see Khanal et al.; 2014, 2018). For validating the finally over scored model, prospectively collected separate data set of 138 ALF patients from the same clinic during the period 2006 to May 2009 was used.

Statistical models

The functional form of Cox regression model (Cox, 1972) for the ith individual with a set of $x_1, x_2, \ldots, x_p$ explanatory variables is:

$$h_i(t) = h_0(t) \, \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}) \quad \ldots\ldots\ldots\ldots\ldots (1)$$

where $\beta_1, \beta_2, \ldots, \beta_p$ are unknown regression coefficients and $h_0(t)$ the baseline hazard function.

For Lognormal AFT model, suppose T_i is a random variable representing the survival time for i^{th} individual. The general log-linear form of the AFT model shows the mathematical relation between the log of time and the set of covariates expressed as follows:

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

where β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_p$ are unknown regression coefficients of the values of p explanatory variables, σ is the scale parameter, and the quantity ε_i is a random variable used to model the deviation of values of $Log(T_i)$ from the linear part of the model. ε_i is assumed to have a particular probability distribution supposed to be followed by the survival time under study. The maximum likelihood estimates of the parameters for both the models are estimated by using iterative Newton-Raphson procedure (Collett, 2003; Hosmer & Lemeshow, 1999). The details about the model specification, goodness of fit of the model, assessment of proportionality of hazards assumption, residuals analysis for CPH model for this data has already been explained (Khanal et al., 2018), and the required tests for selection of Lognormal AFT model, goodness of fit of the model, residual analysis, etc has also been already described for this ALF data (Khanal et al., 2014). Both the considered models have satisfied the goodness of fit of the model for this ALF data (Khanal et al.; 2014, 2018).

CPH model and Lognormal AFT model has been compared on the basis of selection of variables in the final model though the coefficients cannot be compared directly since CPH model yields hazards ratio (HR) whereas Lognormal AFT model yields Time Ratio (TR). The models have also been compared on the basis of goodness of fit tests such as R^2 type statistic (Hosmer, & Lemeshow, 1999), Cox-Snell (CS) residuals (Cox & Snell, 1968) plot, and comparison of observed versus predicted survival curves based on the risk score namely the prognostic index (*PI*). The

formula used to compute CS residuals, R^2 type statistic, etc in the analysis of ALF data were discussed elsewhere (Khanal et al.; 2014, 2018). The mathematical expression for computation of survival probability in different risk strata and *PI* for CPH model had been explained by Khanal et al.(2018). The mathematical expression for computation of survival probability in different risk strata and *PI* in Lognormal AFT model has been provided as follows.

Goodness of fit in different risk strata

The survival probability of Lognormal AFT model for given set of covariates $x_1, x_2, \ldots, x_p$ is as follows (Cleves, Gould, & Gutierrez, 2004).

$$S(t) = S_0 \left\{ \exp(-x_1\beta_1 - x_2\beta_2 - ..... - x_p\beta_p)t \right\}$$

$$= 1 - \Phi \left[\frac{\log \left\{ \exp(-x_1\beta_1 - x_2\beta_2 - .... - x_p\beta_p)t \right\} - \beta_0}{\sigma} \right]$$

$$= 1 - \Phi \left\{ \frac{\log(t) - (\beta_0 + x_1\beta_1 + x_2\beta_2 + ... + x_p\beta_p)}{\sigma} \right\} \qquad \text{................................ (3)}$$

Where, σ is the shape parameter and other notations have usual meaning as already defined. The risk score *PI* for Lognormal AFT model is given by:

$$PI = \left(-\beta_0 - x_1\beta_1 - x_2\beta_2 - ..... - x_p\beta_p \right) \qquad \text{................................ (4)}$$

The *PI* values computed from the model are ranked and divided into number of strata in such a way that there are approximately equal numbers of events in each stratum in a similar manner as done in the case of Cox regression model (Khanal et al., 2018). Curves of model predicted survival probability and observed survival probability (by K-M method) for each risk stratum can be compared for visual assessment of good fit of the model. Besides these, two models have been compared on the basis of robustness of the model.

Robustness of the models

For any regression model, the validity of the assumptions specific to the model and goodness of the model fitted are important before recommending it to others. The validity of such a model can be checked in different situations/set ups, once if the assumptions are met and goodness of the model is satisfactory. In addition, the robustness is also an important component in assessing a statistical model. It is desirable that the model built is stable and reproducible over repeated studies. The term stability refers to the inclusion of same variables in the regression model. If the model built is stable, only those covariates should be included in the model, which can exert strong influence on the outcome variable, every time. This aspect can be studied by a technique called bootstrapping. Chen and George (1985), and Altman and Andersen (1989) applied bootstrap method to investigate the stability of Cox regression model and to study the prognostic implications for individual patients. Later, generalizing the method proposed by Chen & George (1985), Sauerbrei and Schumacher (1992) applied bootstrapping method with stepwise selection

technique to assess the stability of Cox regression model. However these methods are yet to be readily available in all statistical software. The bootstrapping procedure as described by Sauerbrei and Schumacher (1992) is as follows.

A random sample of size n drawn with replacement from the original observations, the vectors $T_j, \delta_j, X_{j1}, \ldots, X_{jp}$, where T_j, the survival time, δ_j censoring indicator variable as defined obviously and $X_{j1}, \ldots, X_{jp}$ are covariates of the jth patient for all j = 1,2,......, n is called bootstrap replication. Suppose M, a large number of bootstrap replications, is considered and treated as M independent samples. Stepwise selection procedure is applied for each bootstrap replication in order to identify significant variables. The results of M bootstrap replications can be summarized in a tabular form showing the percentage inclusion of each variable in bootstrap replications. If the frequency of variables in the model built is high in the bootstrap replications also, the model can be considered as stable. Those variables which repeat at least 50% times in bootstrap replications at selection level of α = 0.05 indicates strong prognostic factors. The procedure suggested by Sauerbrei and Schumacher (1992) was implemented in order to assess the stability of Cox regression model and Lognormal AFT model in ALF data.

Prospective validation of Cox model

Cox model based on the original ALF data from the period 1986 - 2005 has been validated with the prospectively collected data. The validation of a model involves comparison of overall survival in the two data sets and comparison of regression coefficients of the two models based on the development data set and validation data set. In addition, a visual assessment of the closeness of observed and predicted survival curves among different risk strata in the validation data set can also be used. All these methods have been attempted in this study, and are explained briefly as follows.

Comparison of overall survival: original data set vs. validation data set

The overall survival is calculated by the K-M method and the survival curves of the original data and new data sets are compared by using log-rank test.

Comparison of regression coefficients: original data set vs. validation data set

For each study variable in the final model, regression coefficients for original data and prospective data are compared by using two-tailed Z statistic (Liu et al., 2004). The detail computation procedure is explained as follows:

$$Z = \frac{b[O] - b[P]}{SE}$$

……………………………….. (5)

where $b[O]$ = Regression coefficient of each risk factor of the developed model from the original data set and $b[P]$ = Regression coefficient of each risk factor of the model of the prospective data set. SE is the standard error of the difference in coefficients, and is given by:

$$SE = \sqrt{[SE(O)]^2 + [SE(P)]^2}$$ …………………………… (6)

The level of significance is considered as 10% instead of the conventional 5% to test the difference of regression coefficients.

Visual assessment of observed and predicted survival

The predicted survival probability of the subjects in the validation data are calculated by the developed model with the covariate values from the validation data. However, the baseline survival is estimated from the new validation data (Liu et al., 2004). Hence, the survival probability is computed as:

$$S(t) = \left[S_0(t) \right]^{\exp(\beta_1 x_1 + \beta_2 x_2 + \dots\dots + \beta_p x_p)}$$ ………………………….. (7)

where $S_0(t)$ is the estimate of baseline survival function of the prospective data, $\beta_1, \beta_2, \dots\dots, \beta_p$ are the regression coefficients of risk factors of the original Cox model, and $x_1, x_2, \dots\dots, x_p$ are corresponding covariates of the prospective data. The observed survival in each risk strata as estimated by K-M method is compared with the predicted survival in that stratum. Most of the statistical analysis has been carried out using STATA 9.0 except assessing the robustness of the models. The later has been performed by using S-plus.

RESULTS

Comparison of CPH model and Lognormal AFT model on the basis of variable selection

There were altogether 15 independent variables, among which only six same variables were selected by each model either through backward selection or forward selection or stepwise selection method. The hazards ratios with 95% confidence interval generated through final CPH model and time ratios with 95% confidence interval yielded by Lognormal AFT model are presented in Table 1. The comparison of Log-Logistic AFT model and Lognormal AFT model (Khanal et al., 2014) has indicated that there is little to choose between these two AFTs. However, on the basis of AIC, the Lognormal model seems to be relatively better than Log-Logistic AFT model. Each model has satisfied its own model assumptions, the goodness of fit of the model through recommended standard statistical methods (Khanal et al.; 2014, 2018). CPH model yields hazards ratio (HR) whereas AFT model yields time ratio (TR). Hazards ratio measures the effect of covariate in terms of hazards of death of ALF patients whereas time ratio measures the effect of covariate to accelerate or to decelerate survival time. Just for interpretation, HR for the variable total serum bilirubin 1.49 indicates that the hazards of death of ALF is 1.49 times more in patients with total serum bilirubin ≥ 15 mg/dl as compared with the patients having total serum bilirubin < 15 mg/dl. The TR for this variable 0.72 indicates that the survival times for ALF patients with total

serum bilirubin ≥ 15 mg/dl are estimated to be 72 percent of those for ALF patients with total serum bilirubin < 15 mg/dl. In other words, same can be interpreted as the survival time for subjects with total serum bilirubin ≥ 15 mg/dl is estimated to be 28 percent shorter than for subjects with total serum bilirubin < 15 mg/dl.

Table 1. Comparison of CPH model and Lognormal AFT model on the basis of variable selection.

Variable	Cox PH model		Lognormal AFT model	
	HR (95% CI)	SE	*TR* (95% C.I.)	SE
Cerebral edema				
Absent	1.00		1.00	
Present	2.38 (1.99 2.85)	0.22	0.47 (0.41 0.54)	0.03
Total S.bilirubin (mg/dl)				
< 15	1.00		1.00	
≥ 15	1.49 (1.27 1.76)	0.12	0.72 (0.64 0.82)	0.05
Prothrombin time (seconds)				
≤ 25	1.00		1.00	
≥ 25	1.66 (1.41 1.96)	0.14	0.68 (0.59 0.78)	0.04
Age (years)				
< 40	1.00		1.00	
≥ 40	1.41 (1.15 1.72)	0.14	0.74 (0.63 0.87)	0.06
S. creatinine (mg %)				
≤ 1	1.00		1.00	
> 1	1.32 (1.10 1.57)	0.12	0.76 (0.66 0.87)	0.05
Etiology				
Hep. E	1.00		1.00	
Non E	1.33 (1.09 1.63)	0.14	0.81 (0.69 0.94)	0.06

Comparison based on R^2 type statistic: CPH model vs. Lognormal AFT model

As far as R^2 type statistic is concerned, 60% of variation in the partial log-likelihood is explained by the CPH model whereas 39% of variation in full log-likelihood is explained by Lognormal AFT model. On the basis of this statistic, CPH model showed as a better performing model, relatively, in comparison with Lognormal AFT model.

Comparison based on Cox-Snell's residuals plot: CPH model vs. Lognormal AFT model

The plots of Cox-Snell (CS) residuals show that the fit of both the models appear to be good since most of the points of the plot fall on straight line passing through the origin. However CS residuals plot for CPH model seems to be better comparatively (Figure 1) with lesser deviation from the line of ideal fit. The deviation for the CPH model is not considerable up to a value of 4 of the CS residuals.   However, the same is observed up to a value of 3 of the CS residuals in Lognormal AFT model. Though, the scales of measurement of CPH model and AFT model are

different, on the basis of this visual assessment, it can be argued that the CPH model performed better than the Lognormal AFT model.

Comparison of models based on observed and predicted survival curves: CPH model vs. Lognormal AFT model

Another visual assessment of the goodness of fit of the model is the comparison of observed and predicted survival curves in different risk strata (Figure 2). We have made four risk strata in each model (Table 2) and the model predicted survival curves were close to the observed curves in each risk stratum for the CPH model. But in the Lognormal AFT model, the curves are close only in the highest risk stratum. This assessment too indicates that the CPH model may be a better fit than the Lognormal AFT model.

Table 2. Distribution of *PI* in 4 strata with approximately equal number of events for CPH model and Lognormal AFT model.

Risk stratum	CPH Model			Lognormal AFT Model		
	Prognostic Index (PI)	Total	Events observed	Prognostic Index (PI)	Total	Events Observed
I	< 2.3665	430	143	< -2.1223	449	158
II	2.3665 – 3.5129	234	147	-2.1223 to -1.7389	248	156
III	3.5129 – 5.2784	185	145	-1.7389 to -1.3884	184	151
IV	≥ 5.2784	177	165	≥ -1.3884	145	135
Total		1026	600		1026	600

Robustness of the models: bootstrap resampling technique: CPH model vs. Lognormal AFT model

In order to assess the robustness or stability of the developed model, bootstrap resampling technique as suggested by Sauerbrei and Schumacher (1992) was applied for Cox regression and Lognormal AFT model on ALF data. The methodological aspects of bootstrapping procedure for the stability of the model were already described in materials and methods section of this article. We have run 1000 bootstrap replications treating them as 1000 independent samples. Stepwise selection procedure was applied for each bootstrap replication in order to identify significant variables. Percentage inclusion of each variable in bootstrap replications is considered as the criteria for prognostic importance of variable. Summary of bootstrap replication frequency for each covariate for Cox PH model and Lognormal AFT model is presented in Table 3.
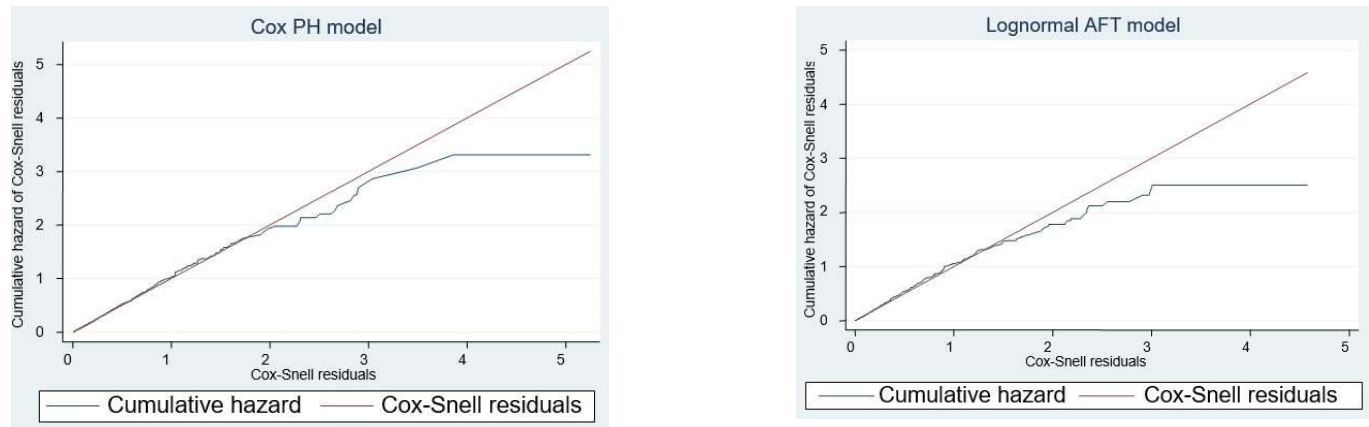
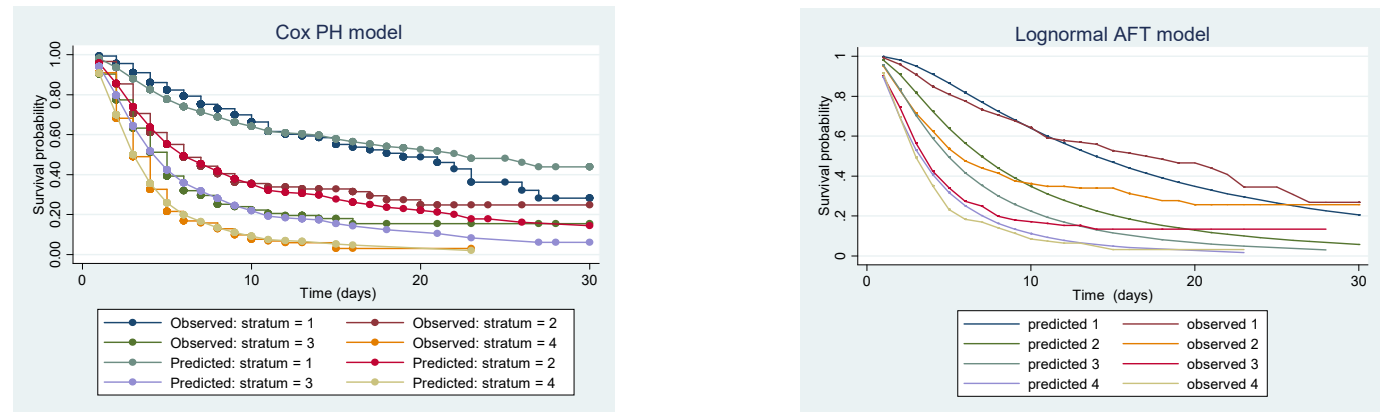**Fig. 1.** Cox-Snell residuals plot for Cox PH model and Lognormal AFT model.

Fig. 2. Observed versus predicted survival curves in 4 risk strata for CPH model and Lognormal AFT model.

Table 3. Summary of bootstrap replications for Cox PH model and Lognormal AFT model.

Variable	Cox PH model		Lognormal AFT model	
	Replication (%)	Rank	Replication (%)	Rank
Cerebral edema	**100.0**	**1.5**	**100.0**	**1.5**
Prothrombin time (Seconds)	**100.0**	**1.5**	**100.0**	**1.5**
Total serum bilirubin (mg/dl)	**96.1**	**3**	**90.6**	**4**
Age(yrs)	**92.0**	**4**	**87.3**	**5**
Serum creatinine (mg %)	**87.6**	**5**	**97.9**	**3**
Etiology	**83.2**	**6**	**72.2**	**6**
Aspartate aminotransferase (AST)	41.8	7	43.6	7
Pregnancy status	19.8	8	21.7	9
Hepatic encephalopathy grade	18.7	9	26.6	8
Alanine aminotransferase (ALT)	8.9	10	9.4	10

It can be seen that the same six variables are selected in each model with a replication frequency ≥ 50%. Both CPH and Lognormal AFT models seem to be stable in the sense that same variables are picked up with at least fifty percent of replication. However, replications of variables in CPH model are on the higher side as compared to those of Lognormal AFT model except for variable serum creatinine. This may indicate that CPH model seems to be more stable comparatively. On the basis of the performance of each model through different measures, and the stability, CPH model can be considered to be the best model for ALF data.

Validation of Cox PH model

We have described the comparison of two models on the basis of variables selected, model performance and the stability of the developed models in the previous sections of this paper. On the basis of these comparisons, we have come to the conclusion that CPH model would be the best one for the mortality due to Acute Liver Failure. An attempt has been made to validate the developed Cox model through a prospectively collected data set of 138 ALF patients from the same clinic from the period 2006 to May 2009. The mortality of these 138 ALF patients was 55%. The male female ratio was 63:75. Among females, 21(28%) were pregnant women. However some of the observations are missing for some variables and complete information of 119 ALF subjects was considered for regression analysis. CPH model developed with original data was validated with the validation data set by comparing overall survival curves of the original model and model based on validation data set, the regression coefficient of each covariate between two models and comparing curves of observed and predicted survival probabilities on the validation data set.

Comparison of overall survival curves: original model vs. model based on validation data

The overall survival curves estimated by using K-M method for the original data set and the validation data set(Figure 3) has indicated that the survival experience of these two data sets does not seem to be considerably different (p = 0.07) at 5% level of significance as assessed by the log-rank test. It is also noted that two curves are crossed with each other towards the end of the survival time which might be because of the smaller number of subjects in each data set.
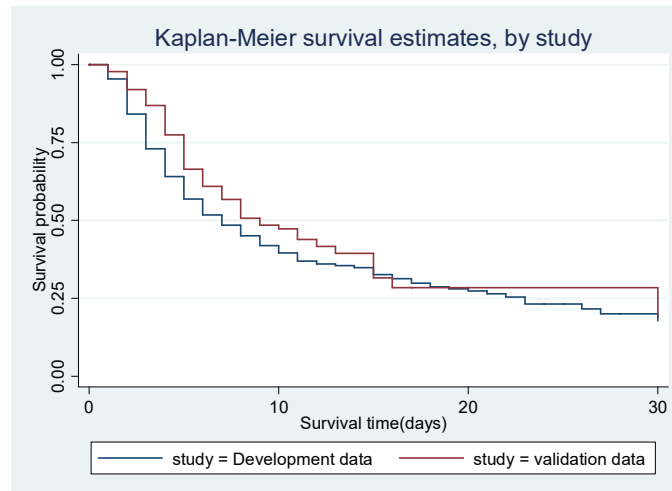

Fig. 3. Comparison of overall survival curves (K-M) between original data and validation data.

Comparison of regression coefficients: Original model vs. model based on validation data set

Six variables, namely, age, cerebral edema, total serum bilirubin, prothrombin time, serum creatinine and etiology were independent predictors for the mortality of ALF patients in the original model. The regression coefficient of each predictor of the original model was compared with that of validation data set by using Z-test at significance level of 0.10 (Table 4).

The notations β_o and β_v used in Table 4 are regression coefficients based on original model and validation data, respectively. The magnitude of the regression coefficient does not seem to be exactly same for each predictor in the original model and in the validation model. It might be because of smaller data set in the validation set as compared to the original data but the exact reason behind this could not be known. Nonetheless, regression coefficients are not significantly different (p > 0.10) between the Cox models based on original data and the validation data set except for one predictor - serum creatinine.

Table 4. Comparison of regression coefficients: Model based on original and validation data.

Variable	Original data (N = 1026)		Validation data (N = 119)		Comparison of β_0 and β_v	
	β_o	SE	β_v	SE	Z	P
Cerebral Edema	0.8686	0.0919	0.6996	0.2850	0.5644	0.57
Total S.bilirubin (mg/dl)	0.3999	0.0832	0.9374	0.3212	- 1.6199	0.11
Prothrombine time(seconds)	0.5091	0.0846	1.0000	0.4063	- 1.1828	0.24
Age	0.3416	0.1015	0.5398	0.4175	- 0.4613	0.64
S. creatinine (mg %)	0.2747	0.0902	1.1420	0.3194	- 2.6131	0.01
Etiology	0 .2859	0.1037	0.5569	0.3610	- 0.7215	0.47

Comparison of observed and predicted survival probability

PI was calculated plugging regression coefficient (β) from the original model into the new model based on prospective data. However the baseline survival function $\left(S_0\right)$ was used from the validation data set. The computed *PI* values varied from 1 to 14.58 and the distribution of *PI* is provided in Table 5.

Table 5. Distribution of PI.

PI	Number	%
< 2	22	18.5
2 – 5	58	48.7
≥ 5	39	32.8
Total	119	100.0

Further, three strata (Table 6) were made in such a way that there was approximately equal number of events. The observed survival probability estimates (using K-M method) and predicted model survival probability were compared through survival curves in each stratum (Figure 4).

Table 6. Distribution of *PI* considering approximately equal number of events in each stratum.

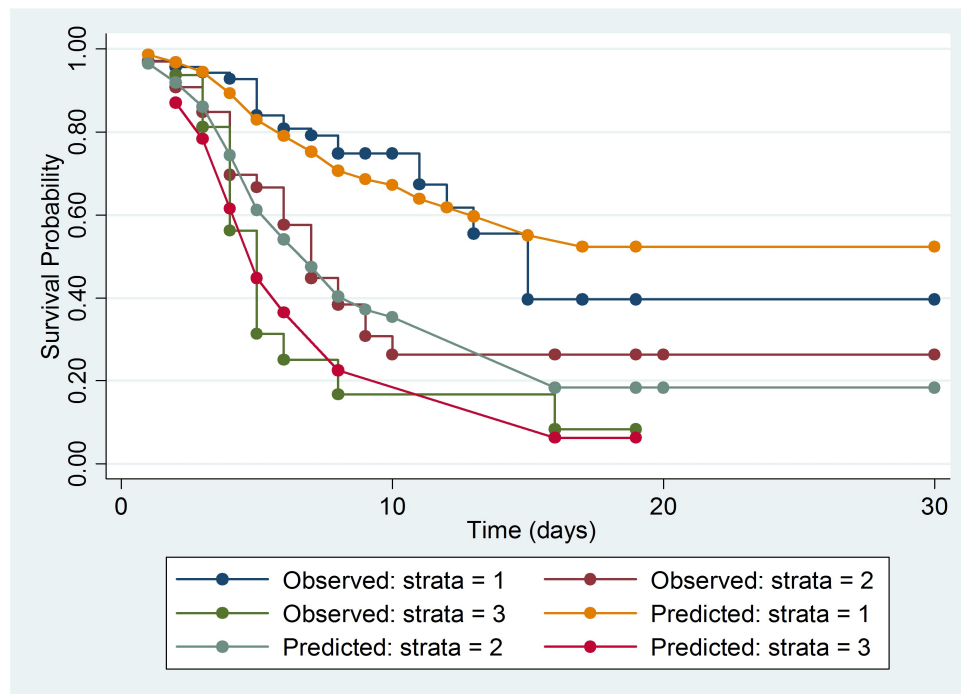| Stratum	Prognostic Index (PI)	Total	Events observed
1	< 3.55	70	22
2	3.55 – 7.87	33	23
3	≥ 7.87	16	14
Total		119	59

Fig. 4. Comparison of observed and predicted survival probability (K-M method) based on CPH model.

Observed vs. predicted survival curves (Figure 4) in each stratum appear to be overlapping up to 18 days. The deviation beyond 18 days may be due to small number of events in the validation data set. Thus, it may indicate that the original Cox model is valid.

DISCUSSION

After satisfying all the model assumptions of the proposed model, the model has been run and the required goodness of fit of the model has also been evaluated. Three survival models had been applied to identify the prognostic factors of ALF patients. Among these three applied models, CPH model measured in proportionality of hazards metric and other two models namely- Lognormal AFT and Log-Logistic AFT models measured in AFT metric were applied. Each of these three models has satisfied model assumption, different tests of goodness of fit of the model though the Cox regression yields the hazards ratio whereas later two AFTs yield time ratio (Khanal et al. 2014; 2018). It is clearly indicated that the suitability of any of these three models has been justified for this ALF data with short follow up time. It is very difficult to choose either Lognormal AFT model or Log-Logistic AFT model particularly for the analysis of this ALF data since both the models have satisfied all the model performance tests. However, comparing AIC between these two AFTs, Lognormal model seems to be slightly better because the AIC of this is smaller than that of Log-Logistic AFT model (Khanal et al., 2014).

Though both CPH model and Lognormal AFT model seems to be suitable for this ALF data, these two models were compared on the basis of variables selection in the final model and the goodness of the fit of each model. Both the models identified the same set of predictors of survival but the fit of CPH model appeared to be better than that of Lognormal AFT model. The Cox-Snell residuals plots showed more deviation in Lognormal AFT model than in the CPH model. The deviation for residuals of size more than 4 was notable in the CPH model while the same is seen for residuals of size more than 2 in the Lognormal AFT model. Numerically, observations with a CS residual of more than 3 amounted to only 0.6% of the total 1026 observations in the CPH model while they formed 3.5% in Lognormal AFT model. So, based on CS residuals plots, CPH model showed a better fit than the Lognormal AFT model.

The observed and model predicted survival curves for these two models showed that the fit of the CPH model is good across the four risk strata while it is good only in the highest risk stratum in the Lognormal AFT model. Thus, again, the CPH model is indicated as a better choice among the models considered. Apart from the different measures for assessing the goodness of the fitted models, the stability or the robustness of each developed model was also assessed by bootstrapping technique. Both the models seem to be stable with respect to the replication of variables. However, the replication frequency of each variable in final CPH model is higher than that of Lognormal AFT model, which means that the CPH model is more stable than Lognormal AFT model.

Studies related to comparison of CPH model with AFT models in health/clinical setup is found relatively less. Some of them are Folorunso and Osanyintupin (2018), Faruk (2018), and Zare et al. (2018) has compared CPH model and different AFT models in the analysis of neonatal jaundice data of Nigeria, birth interval data and survival of gastric cancer data respectively. Obre, Ferreira, and Nun˜ez-Anton (2002) has also explained different aspects of CPH model and AFT models and their comparison with sufficient theoretical aspects and applications.

Application of parametric survival models on clinical data was mainly in the area of survival of cancer patients such as female breast cancer. Such applications involve long duration of follow-up. Very few such applications are found in scenarios of events of short gestation. The present study is an attempt to assess the suitability of the parametric survival models in clinical data with events of short gestation. Our results indicated the semi-parametric CPH model is the best suited for the survival of Acute Liver Failure. This supports the suggestion that Lognormal and Log-Logistic survival models are good in the analysis of time to event data for events with long gestation (Gamel & Vogel, 1997; McCready et al., 2000; Smaletz et al., 2002).

Cox regression does not require any distributional assumption for baseline hazard function. However, it is able to provide good estimates of regression coefficients, hazard ratios of interest

and survival curves for a wide variety of data situations. It can be considered as a robust model in the sense that the results from using the Cox model will closely approximate the results for the correct parametric model. In other words, Cox PH model adopts the baseline hazard function whatever appropriate distribution suits for the data under study (Kleinbaum & Klein, 2005). The observed shape of the baseline hazard after smoothening, though indicative of Log-logistic or Lognormal distribution, may not be a perfect indication. Under such situation, Cox model might have adopted the appropriate distribution and hence scored better than other models. After satisfying the model assumptions, recommended goodness of fit of the model and other necessary requirements, one can apply either of the models. Nonetheless, different authors preferred one model instead of another depending on researchers' choice for ease of interpretation of the estimates.

The parameters in the AFT models are interpreted in the time scale and some authors consider it as an advantage for the interpretation as compared to the interpretation of parameters in semi-parametric models (Bradburn, Clark, Love, & Altman, 2003[a,b]; Patel et al., 2006). Wei (1992) argued that the interpretation of parameters in AFT models may be more easily understood than hazard ratios by clinical investigators, especially those unfamiliar with survival time analyses. However, AFT models are not frequently used in clinical research. Further, CPH model has already been accepted as the standard method for regression analysis of survival times not only in clinical research but also in many applied settings, without any doubt. Most research-oriented clinicians have little or no trouble in understanding the proportional hazards model or the hazard ratio. The proportional hazards model has, in fact, been used by many investigators for years. It is so embedded in the statistical practice of some fields that it is unlikely that it will be replaced by another model in the foreseeable future (Fisher, 1992). Hence it can be argued that Cox regression model is the most popular technique in survival analysis. Validation of the final Cox PH model showed that the regression coefficient for each predictor did not change significantly, except for the variable serum creatinine. The reasons for this are not immediately understood. If the developed model is not valid, it should be reflected in other five variables too. Because of this, it may be interpreted that the Cox model developed based on the large data set is valid.

The ALF data used in this study is one of the largest data sets in the world. The identification of prognostic factors for the mortality of ALF patients in the Indian situation applying suitable statistical model can be considered as the major epidemiological strength of this study. It is very important to have sufficiently large sample size for adequate power of the study in survival analysis. The rule of thumb is that there should be at least 10 events for one predictor variable. In our study, we had 15 predictor variables on 1099 ALF patients with 647 events, which provided 43 events on the average for each predictor. Thus we considered large number of events providing adequate power for the study. Statistically, large number of events itself can be regarded as a strength to fit a good survival model.

CONCLUSION

After satisfying the model assumptions, model adequacy and the goodness of fit of the model, either CPH or parametric model is equally applicable. In this data set both CPH and Lognormal AFT model are found well fitted and can be applied either of them. While comparing between them based on the goodness of fit of the model, plot of the CS residuals, plot of the observed and predicted survival curves and robustness of the model through the repetition of the predictor variables in the final model, the performance of CPH model for the identification of prognostic factors for the survival of ALF patients is found better comparatively. The validity of the final Cox model is also checked with different tests using separate validation data set of ALF patients.

CONFLICT OF INTEREST

Authors declare that there is not any conflict of interest among them.

ACKNOWLEDGMENTS

REFERENCES

Abrahamowicz, M., MacKenzie, T., & Esdaile, J. M. (1996). Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. *Journal of the American Statistical Association*, *91*, 1432-1439.

Acharya, S. K., Madan, K., Dattagupta, S., & Panda, S. K. (2006). Viral hepatitis in India. *The National Medical Journal of India*, *19*, 203-217.

Altman, D. G., & Andersen, P. K. (1989). Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine*, *8,* 771-783.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003[a]). Survival analysis part III: multivariate data analysis-choosing a model and assessing its adequacy and fit. *British Journal of Cancer*, *89*, 605-611.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003[b]). Survival analysis part II: Multivariate data analysis - An introduction to concepts and methods. *British Journal of Cancer*, *89,* 431-436.

Chen, C-H, & George, S. L. (1985). The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. Statistics in Medicine, *4*. 39–46.

Cleves, M. A., Gould W. W., & Gutierrez R. G. (2004). *An Introduction to Survival Analysis Using Stata*. A Stata Press Publication, STATA Corporation, Texas.

Collett, D. (2003). *Modeling survival data in medical research*. Boca Raton Florida: Chapman & Hall/CRC.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, Series B, *34*. 187-220.

Cox, D. R., & Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society*, Series B, *30*, 248-275.

Faruk, A. (2018). The comparison of proportional hazards and accelerated failure time models in analyzing the first birth interval survival data. *International Conference on Mathematics: Pure, Applied and Computation. IOP Conf. Series: Journal of Physics: Conf. Series* 974 (2018) 012008. DOI:10.1088/1742-6596/974/1/012008

Fisher, L. L. (1992). Discussion of session 1: Clinical trials, survival analysis. *Statistics in Medicine*, *11*, 1881–1885.

Folorunso, S. A., & Osanyintupin, O. D. (2018). Comparison of Cox proportional hazard model and accelerated failure time (Aft) models: An application to neonatal jaundice. *Int J Pub Health & Safety, 3*, 171.

Gamel, J. W., & Vogel, R. L. (1997). Comparison of parametric and nonparametric survival methods using simulated clinical data. *Statistics in Medicine*, *16,* 1629-1643.

Hess, K. R. (1994). Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in Medicine*, *13,* 1045-1062.

Hosmer, D. W., Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modelling of Time to Event Data,* New York: A Wiley-Inter Science Publication, John Wiley & Sons.

Kay, R, & Kinnersley N. (2002). 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 Infection Journal*, *36,* 571-579.

Khanal, S. P., Sreenivas, V., & Acharya, S. K. (2014). Accelerated failure time models: An application in the survival of acute liver failure patients in India. *International Journal of Science of Research (IJSR)*, *3*(6), 161-166.

Khanal, S. P., Sreenivas, V., & Acharya, S. K. (2018). Cox proportional hazards model for identification of the prognostic factors in the survival of acute liver failure patients in India. *Nep. J. Stat.*, *2*, 53-74.

Kim, W. R., Brown, J. R. S., Terrault, N. A., El-Serag, H. (2002). Burden of liver disease in the United States: Summary of a Workshop. *Hepatology*, *36*, 227-242.

Kleinbaum, D.G., & Klein, M. (2005). *Survival Analysis: A Self – Learning Text.* 2nd Edition, Springer, U.S.A.

Liu, J., Hong, Y., D'Agostino, R. B. Sr., Wu, Z., Wang, W., Sun. J., Wilson, P. W., Kannel, W. B., & Zhao, D. (2004). Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. *The Journal of the American Medical Association*, *291*, 2591-2599.

McCready, D. R., Chapman, J. A., Hanna, W. M., Kahn, H. J., Murray, D., Fish, E. B., Trudeau, M. E., Andrulis, I. L., Lickley, H. L. (2000). Factors affecting distant disease-free survival for primary invasive breast cancer: use of a log-normal survival model. *Annals of Surgical*

Oncology, *7,* 416-426.

Obre, J., Ferreira, E., & Nun˜ez-Anton,   V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine*. *21*, 3493–3510. doi: 10.1002/sim.1251

Patel, K., Kay, R., & Rowell, L. (2006). Comparing proportional hazards and accelerated failure time models: an application in influenza. *Pharmaceutical Statistics*, *5,* 213-224.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J., & Smith, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer*, *34,* 585-612.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J., & Smith, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *British Journal of Cancer*, *35,* 1-39.

Sauerbrei, W., & Schumacher, M. (1992). A bootstrap resampling procedure for model building: application to the Cox regression model. *Statistics in Medicine*, *11*, 2093-2109.

Smaletz, O., Scher, H. I., Small, E. J., Verbel, D. A., McMillan, A., Regan, K., Kelly, W. K., & Kattan, M. W. (2002). Nomogram for overall survival of patients with progressive metastatic prostate cancer after castration. *Journal of Clinical Oncology*, *20,* 3972-3982.

Wei, L. J. (1992). The accelerated failure time: A useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, *11,* 1871-1879.

Zare, A., Hosseini, M., Mahmoodi, M., Mohammad, K., Zeraati, H., & Naieni, K. H. (2015). A comparison between accelerated failure-time and Cox proportional hazard models in analyzing the survival of gastric cancer patients. *Iran J Public Health*, *44*(8), 1095-1102. http://ijph.tums.ac.ir

_____

Reference to this paper should be made as follows:

Khanal, S. P., Sreenivas, V., & Acharya, S. K. (2019). 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**.** *Nep. J. Stat., 3*, 21-40.

_____