Volume : 2 | Issue : 5 | May 2013 • ISSN No 2277 - 8160

Research Paper | Medical Science

Modelling of Time to Event Breast Cancer Data Using Accelerated Failure Time (Aft) in South India Women

Pari Dayal L	DRBCCC Hindu College, Chennai,
Leo Alexander T	Loyola College,Chennai
Ponnuraja C	NIRT(ICMR), Chennai,
Ranganathan Rama	Cancer institute(WIA), Chennai.
Venkatesan P	NIRT(ICMR), Chennai

ABSTRACT

Most of the studies have been widely studied between breast cancer and risk factor using the classical way of statistical methods. This paper aims to implement a class of flexible parametric survival models through accelerated failure time models for identifying risk factors for breast cancer time to death among South India women. This study also attempts to explore the survival experience of breast cancer patients. Since the death due to severity of the stages varies with age; the age is broadly classified into two groups <50 years & 50 years. The survival experiences between these age groups are presented using Kaplan-Meier survival curves and there is no significant difference between groups. However, the stages differ significantly in each group. The accelerated failure time (AFT) models using Exponential, Weibull, Gamma, Log logistic and lognormal were explored and compared by using the AIC and deviance. The Gamma and log normal models produced similar results.

KEYWORDS: survival analysis, AFT modes, Kaplan-Meier, Log rank test

1. Introduction

There are two types of regression models that have been developed for time to event survival data. The first model is based on the hazard function in patient groups compared to a baseline population by means of a multiplicative effect on hazards scale. The multiplicative factor is assumed to be constant over time, in which case the model forces the hazards in the different patient groups to be proportional(Cox 1972). The second model is applied for modeling the survival time directly along with covariates assumed to act multiplicatively on the time scale. The accelerated failure time (AFT) model is of the second type and it is a class of linear regression model in which the response variable is the logarithm or a known monotone transformation of a failure time (Kalbfleisch and Prentice, 1980). When using semi-parametric models for the analysis of time to event data, it is needed to provide a reduced set of assumption for forming the hazard ratio from the coefficients that can be easily interrupted and clinically meaningful (Richard and Nelson, 2002).

The AFT model is not frequently used model to analyze survivorship data, but it offers a potentially useful statistical approach which is based upon the survival curve rather than the hazard function. Wei (1992) suggested that since the parameters in the AFT models are interpreted as effects on the time scale, they may be more easily understood than the hazard ratios. We desire to emphasize that all AFT models are named for the distribution of T rather than the distribution of $\log T$. The reason for allowing a different distribution assumption is that they have different implications for the shape of hazard function. Three parametric regression models, namely the exponential, gamma and Weibull are used to compare a contrast the analysis of right censored cancer trail data with covariate effects, through a proportional hazards interpretation (Hayat et al., 2010). The other two parametric survival models, the log-logistic and log normal are to be described in the other way to proportional hazards.

2. Statistical Methods

Parametric Survival time model have been viewed excessively in many texts (Cox and Oakes,1984; Wei, 1992; Lee, 1992; Anderson et al., 1993; Klein and Moeschberger, 1997 and Hosmer and Lemeshow, 1999). Hosmer and Lemeshow cautioned that while using parametric form of the hazard instead of using semi parametric form of the hazard. A recent paper by Orbe et al., (2002) compares both the Cox and AFT models and discussed the advantages of AFT models with their limitations.

Let T_i be a random variable denoting the failure time for the i^{th} subject, $x_{i1}, x_{i2}, \ldots x_p$ and let be the values of p covariates for that same subject. The model is then

$$\log T_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_p + \sigma_i \quad \varepsilon_i \overset{iid}{\sim} S_0(.) \text{ -(1)}$$

where ε_t is a random disturbance term, $\beta_0, \beta_1 \ldots, \beta_p$, and σ are parameters to be estimated, $S_0(.)$ is a known baseline survival, $T_i's$ are actual survival time some times observed, σ is a scale parameter and $x_i's$ are fixed $p \times 1$ vector of covariates. The σ can be omitted, which requires that the variance of ε_i be allowed to be different from 1. But it is simpler to fix the variance of ε_i at 1 and let σ change. All AFT models are named for the distribution of T rather than the distribution of ε or log T. The reason for allowing different distribution assumptions is that they have different implications for the shapes of hazard function.

In this application to Cancer data, the age and stages are considered as important covariates based on previous studies. The stages of the disease are classified based on the five point scale (Engel et al., 2003; Michaelson et al., 2002). The age is classified into less than 50 years and more than 50 years as per Robb et al., (2007). Moreover, in breast cancer age is an important factor and 70% of the women diagnosed were over the age 50 (Robb et al., 2007). In order to determine the most suitable distribution to the survival periods, the AFT form of the Exponential, Gamma, Weibull, Lognormal, and Log-logistic distributions were used. The models obtained through these five were compared using Akaike Information Criteria (AIC), and -2 log likelihood (-2LL).

Application to Breast Cancer Data

The database consists of 368 breast cancer women patients diagnosed at Cancer Institute (WIA), Chennai, India and follow-up period up to 120 months. The event of interest was time to death. The demographic and disease characteristics of the patients are given in table 1

Table 1: Classification of death according to Stages and Age group

	Stages			Age groups	
	Stage2B N (%)	Stage3A N (%)	Stage3B N (%)	Age <50 years N (%)	Age ³ 50 years N (%)

Death Status Alive Dead	61 (55) 49 (45)	72 (56) 56 (44)	48 (37) 82 (63)	115 (53) 103 (47)	66 (44) 84 (56)
Total	110	128	130	218	150

From the table1, we see that death increases with the severity of stage and age.

Figure 1: Kaplan-Meier graph of the survival curves for different age groups

In the Kaplan-Meier graph, is seen that the estimated survival probability of the less than 50 group patients was higher than the one estimated in the 50-and-above group (Figure 1). The difference was not significant using the log rank test.

The mean estimates (Table 2) of stage2B and 3A are related with each other than compared to stage 3B. Though the Stages 2B and 3A are similar with respect to the mean in estimates, they are statistically significant differences based on the log rank test (p =0.001). The median estimates are also presented for all the three stages in the Table 2 and this is also imitating the same message when compared to mean estimates of stages.

Table 2. Means for Survival Time for different stages				
	Mean			
	Estimate	SE	95% CI	
Stage 2B	106.729	5.998	94.974	118.485
Stage 3A	110.488	5.915	98.895	122.081
Stage 3B	79.655	5.776	68.335	90.975
Overall	99.488	3.559	92.512	106.463
log rank test (p =0.001)				

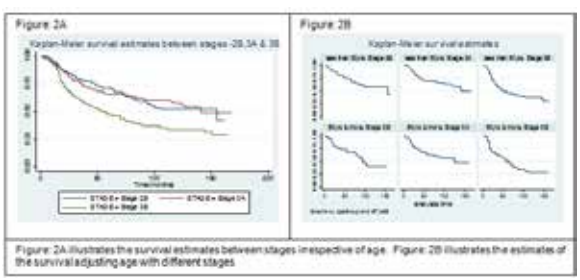

Figure: 2A illustrates the survival estimates between stages irrespective of age. Figure: 2B illustrates the estimates of the survival adjusting age with different stages

From the above Kaplan-Meier curves (Figure 2A&B), we note that substantial differences between the stages exists.

The parameter estimates, standard errors, and indicator for *p*-values are given in Table 3. Here, we see that patients with stage 3B and stage3A disease do significantly worse than patients with stage 2B disease.

When analyzing the survival periods using parametric models, the age and stage variables are taken as the covariate. The parameter obtained as a result of the analysis, the -2LL and AIC values are presented in Table 3.

Table 3. Parameter Estimates (with SE) for the Effects of Stage and Age on Survival for Breast Cancer Patients, Modelling Time Directly Assuming the following Distributions										
Variable	Exponential		Weibull		log Normal		Log Logistic		Gamma	
Age	-0.006	0.008	-0.006	0.009	-0.003	0.009	-0.002	0.009	-0.001	0.008
Stage 3A	-0.291*	0.092	-0.308*	0.108	-0.392*	0.122	-0.393*	0.114	-0.857	0.194
Stage 3B	0.358*	0.161	0.413*	0.206	0.430*	0.226	0.447*	0.213	-0.093	0.491
Constant	5.480	0.384	5.536	0.429	5.051	0.449	4.960	0.446	5.019	0.507
-2LL	950.83		946.41		918.34		928.88		922.05	
AIC	957.74		955.23		929.76		939.12		928.05	
* significant at 5% level										

As lower values of AIC and -2LL suggest a better model and the model obtained by using lognormal and Gamma model are almost similar. These models have lower AIC and -2LL for both age groups as well as stages.

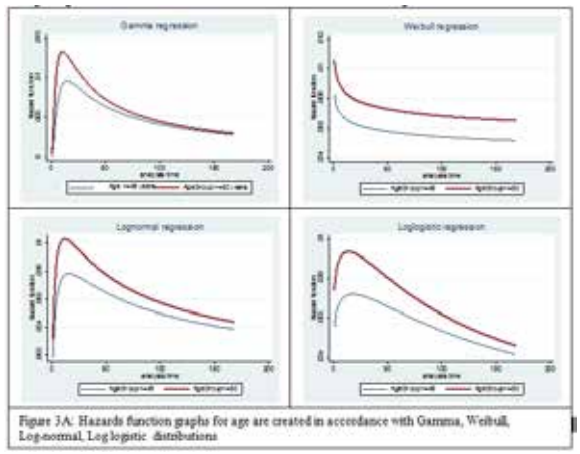

Figure 3A: Hazards function graphs for age are created in accordance with Gamma, Weibull, Log-normal, Log logistic distributions

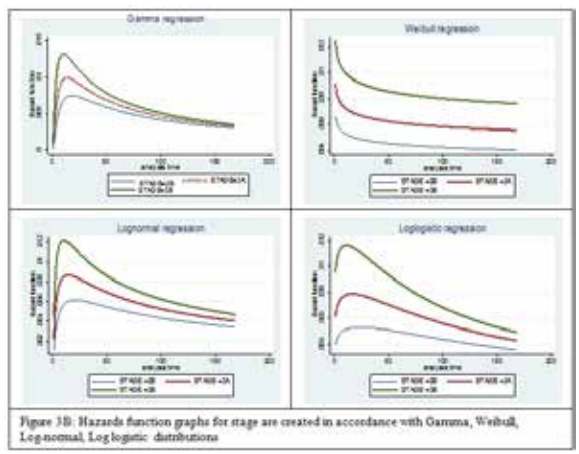

Figure 3B: Hazards function graphs for stage are created in accordance with Gamma, Weibull, Log-normal, Log logistic distributions

The Gamma model is the more suitable model and the other way decision based on -2LL, the log normal is the most suitable model.

Some of the supportive evidence of hazards function graphs (Figure 3A) for age and stage are created in accordance with Lognormal, Log logistic, Weibull, Gamma distributions and are given in figure 3A & 3B respectively. When the hazard function graphs are examined, it is seen that the mortality risk in the <50 age group is lesser than ^{3} 50 age group(Figure 3A).

Also the hazard function graphs obtained as a result of the analyses, it is seen that the mortality risk of the stages 2B and 3A group patients are lower than the stage3B group patients (Figure 3B).

The time ratio (TR) values are calculated as a result of the AFT function and their confidence intervals are given (Table 4). The time ratio values obtained for all distributions related to the age are not statistically significant (p>0.05). This shows that the age does not make-believe as a risk factor for breast cancer patients.

Table 4. Time Ratio (with SE) for the Effects of Age on Survival for Breast Cancer Patients.					
Distribution	Time Ratio	SE	Pvalue	95% Confidence Interval	
Gamma	0.994	0.009	0.464	0.976	1.011
Exponential	0.990	0.007	0.221	0.975	1.005
Weibull	0.989	0.009	0.245	0.973	1.007
log Normal	0.993	0.009	0.422	0.975	1.010
Log Logistic	0.993	0.009	0.433	0.975	1.011

Discussion

The AFT models are compared for time to death in breast cancer data using Exponential, Weibull, Gamma, Log logistic and lognormal models. These models illustrate age is not a risk factor but stages are found to be a risk factor. It is concluded that Gamma and log normal models are identified as suitable models based on the lower values of AIC and deviance. The findings closely relate with other findings like (Robb et al., 2007; Hayat, et al., 2010).

REFERENCES

1. Anderson, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). Statistical Model Based on Counting Process. Springer-Verlag, New York. | 2. Cox, D.R and Oakes, D. (1984), Analysis of Survival Data. London Chapman and Hall. | 3. Engel J, Eckel R, Kerr J, Schmidt M, Furstenberger G, Richter R, Sauer H, Senn HJ, Holzel D. (2003). The process of metastasisation for breast cancer, Eur J Cancer, 39:1794–1806. | 4. Hayat, E. A., Suner, A., Burak, U. Y. A. R., Dursun, Ö., Orman, M. N., & Kitapçioğlu, G. (2010). Comparison of Five Survival Models: Breast Cancer Registry Data from Ege University Cancer Research Center. Turkiye Klinikleri Journal of Medical Sciences, 30(5), 1665. | 5. Hosmer, D. W. and Lemeshow, S. (1999). Applied Survival Analysis; Regression Modeling of Time to Event Data. New York: John Wiley & Sons. | 6. Klein, J. P. and Moeschberger, M. L. (1997). Survival Analysis Techniques for Censored and truncated data, New York: Springer-Verlag. | 7. Lee, E. T. (1992). Statistical Methods for Survival Data Analysis, 2nd Edition, John Wiley , New York. | 8. Limpert E, Stahel WA, Abbt M (2001). Log-normal distributions across the sciences: Keys and clues. BioScience, 51(5):341-52. | 9. Michaelson J.S., Silverstein M., Wyatt J., Weber G., Moore R., Halpern E., Kopans D.B., Hughes K. (2002). Predicting the survival of patients with breast carcinoma using tumor size, Cancer, 95:713–23. | 10. Orbe, J., Ferreira, E., Nunez-Anton, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. Statistics in Medicine, 21:3493- 3510. | 11. Richard K., Nelson K. (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 Information Journal, 36: 571–579. | 12. Robb C, Haley WE, Balducci L, Extermann M, Perkins EA, Small BJ, et al. (2007). Impact of breast cancer survivorship on quality of life in older women. Crit Rev Oncol Hematol, 62(1): 84-91. | 13. Wei, L. J. (1992). The accellarated failure time model: a useful alternative to the Cox regression model in survival analysis. Statistics in Medicine, 11, 1871-79. |