An Analysis of the Survival times of the Patients with AML (Acute Myeloid Leukemia) using Cox Proportional Hazard Model with changing form of the Hazard Function

Bahrawar Jan
Department of Statistics
University of Peshawar
Peshawar, Pakistan

Iftikharuddin
Government Post Graduate College
Mansehra, Pakistan

Abstracts

Survival pattern can be conveniently inspected with the help of hazard function. Thus, it is often informative to examine the hazard function when time to response data is being analyzed. In this study Cox proportional hazard model is fitted to patient data of Acute Myeloid Leukemia. The natural log of the hazard function is the response and two of the explanatory variables age and cellularity level of the patient. Three different forms of the hazard rates are incorporated in the hazard function and are tried out as response.

The fitted model based on three different hazard function revealed, that estimates of the regression coefficients and their relevant standard error are quit close when $h_2(t)$ and $h_3(t)$ are used as hazard rate. Both reflect positive but insignificant effect for the two explanatory variables, while $h_1(t)$ is giving misleading results.

AML patient of age 50 years are below have marginal risk of 12.57% and that for the cellularity of 100% or less it.

1. Introduction

It is customary to fit the model with the log hazard function as the response variable in survival time data. The same information may not be so easily extracted from the survival or death density functions.

The cox proportional hazards models (Cox, 1972) is a big achievement for the analysis of survival time date with covariates (Lawless, 1982). In this study an extensive use of the cox proportional hazards models has been made and different proposed approaches for the hazard function estimates are compared. The cox model is fitted to acute myeloid leukemia (AML) data. The log hazard function is fitted with two explanatory variables namely age of the patient at death and cellularity level. Three different ways of hazard rates are tried to estimate the regression coefficients.

Adult acute myeloid leukemia (AML) is a cancer of the blood and bone marrow. This type of cancer usually gets worse quickly if it is not treated. It is the most common type of acute leukemia in adults. AML is also called acute myelogenous leukemia, acute myeloblastic leukemia, acute granulocytic leukemia, and acute nonlymphocytic leukemia (Bennett JM, et al, 1985).

Normally, the bone marrow produces stem cells (immature cells) that develop into mature blood cells. In AML, the stem cells usually develop into a type of immature white blood cell called myeloblasts (or myeloid blasts). The myeloblasts in AML are abnormal and do not mature into healthy white blood cells. Sometimes in AML, too many stem cells develop into abnormal red blood cells or platelets. These abnormal white blood cells, red blood cells, or platelets are also called leukemia cells or blasts. Leukemia cells are unable to do their usual work and can build up in the bone marrow and blood so there is less room for healthy white blood cells, red blood cells, and platelets. When this happens, infection, anemia, or easy bleeding may occur.

Acute myeloid leukemia (AML) is one of the most aggressive forms of leukemia (cancer of the blood) in adults.

According to Linet MS, Devesa SS (1991), AML comprises about 40% of leukemia in the Western world. Approximately 6500 cases are diagnosed in adults in the US annually. The incidence of AML increases steadily with increasing age. The median age is 55 to 60. AML is more common in males than in females, and more common in whites than in blacks.

Since there is no proper data recording system in Pakistan to look in to the prognostic factors of AML, the survival times of 30 patients of AML are taken from literature along with two prognostic covariates as patient's age and the cellularity level for analysis. It is suspected that these two factors have a significant influence on the survival times.

In this study interest lying on two-fold statistical objectives; i.e. one to compare three different hazard functions, when used as response simultaneously in the cox proportional hazard model, and other to analyze the survival time and its dependency on the prognostic factors.

2. Methodology

For analyzing survival data it is sensible to model the hazard function on explanatory or prognostic covariates because it will depict more realistic and easily interpretable features of the survival times. The Cox Proportional Hazard Model is one of the most widely used model in survival data analysis. The model is specified by the hazard relationship;

$$h_i (t) = h_o (t) \exp(X \beta) \text{-----------------} 1$$

Where $X=[x1, x2,……..xp]$ is the matrix of covariates, β is the vector of regression coefficients and $h_o (t)$ is the base line hazard function when no covariates are in the model.

The above model is non parametric in the sense that is involves an unspecified function in the form of base line hazard function $h_o(t)$. The above model is in multiplicative form, which makes its interpretation very complex so we convert it to loglinear form, to make it additive, by taking natural log on both sides, it becomes

$$\ln[h_i (t)/h_o(t)]=X \beta \text{---------------------} 2$$

This format is not only easy to fit but it is also interpretable. To estimate vector of parameters β for the function with response $h_i(t)$ and set of covariates X, we tried three different forms, in order to obtain values of hazard rate. They are

1. The most simple form is $h_1(t) = 1/k_i(t)$ where k_i are survival times ranked inversely in descending order the hazard for censored observation is taken to be zero.
2. The second form is
$$h_2(t) = [S(t_i)q_i/b_i]/\{0.5[S(t_i)+S(t_{i+1})]\}$$
where $S(t_i)$ is the survival rate at time t_i, q_i is the conditional proportion of dying for ith observation (patient), b_i is the length of time between two successive uncensored survival times.
3. The third form for the hazard rate is,
$$h_3(t) = -\ln(p_i)/b_i$$
where pi is the probability of survival for the ith patient.

3. Analysis

Table-I given below contain the parameter estimates of the Cox Proportional Hazard Model with related tests and risks using the three different mathematical forms of hazard rates $h_1(t) = 1/k$, $h_2(t) = [S(t_i)q_i/b_i]/\{.5[S(t_i)+S(t_{i+1})]\}$ and $h_3(t) = -\ln(p_i)/b_i$. The covariates used are X_1 (age) and X_2 (cellularity level).

Table I: Parameter Estimates of Cox Proportional Hazard Model using different

Hazard Estimate	Model Fitted with	Base-line Hazard	Reg. Co-efficient		T-value	P-value	Risk 1 (Dying)	Risk 2 (Cell ;)	R. Risk
			β1 (S.E)	β2 (S.E)					
$h_1(t)$	X1	0.0919	-0.3878 (0.1995)	---------	-1.94	0.1305	0.1542	---------	---------
	X2	0.0537	--------	0.1395 (0.2093)	0.6664	>0.50	---------	0.1154	---------
	X1,X2	0.0718	-0.3743 (0.2086)	0.0622 (0.2036)	-1.794(β1) 0.3056(β2)	0.1096(β1) 0.5000(β2)			.4845(x1=0) .5924(x2=0)
$h_2(t)$	X1	0.049	0.4466 (0.2875)	---------	1.5535	0.2571	0.1256	---------	--------------
	X2	0.0617	---------	0.1256 (0.2869)	0.4379	>0.5000	-------	0.1397	-------------
	X1,X2	0.0424	0.4835 (0.2953)	0.2042 (0.2797)	1.6375(β1) 0.7300(β2)	0.2772(β1) 0.5150(β2)	----------	------	0.4492(x1=0) 0.4240(x2=0)
$h_3(t)$	X1	0.0442	0.4751 (0.2957)	---------	1.6065	0.0655	0.1257	---------	------------
	X2	0.620	---------	0.1224 (0.2964)	0.4129	>0.5000	-------	0.1320	------------
	X1,X2	0.0416	0.5123 (0.3037)	0.2052 (0.2877)	1.6868(β1) 0.7132(β2)	0.1106(β1) 0.4922(β2)	--------	-------	0.4487(x1=0) 0.3748(x2=0)

Hazard Functions

In all the three cases it is clear that no covariate is significantly related with hazard function due to larger P-values. All the regression coefficients are positive except model specified by the hazard risks defined as $h_1(t)=1/k$ which indicates that X_1(age) is negatively related to the response variable in both cases i.e. in separate fitting of X_1 or in simultaneous fitting of X_1 and X_2. The later two estimators' $h_2(t)$ and $h_3(t)$ give almost the same estimates and the sign of all regression coefficients are positive as well.

Using $h_1(t)$ the estimated risk of dying for patients of age 50 years or below is 0.1542, which means that 15.42% of the patients are expected to die at age 50 years or below, ignoring the cellularity level completely. The estimated risk of patient, having cellularity level of 100% or less is 0.1154. Which means that 11.54% of the patients having the cellularity level of 100% or less are expected to die irrespective of their age factor. The relative risk when both the covariates are considered simultaneously in the model is 0.4845, for $X_1=0$,with $X_2=0,1$and 0.5924 for $X_2=0$ and $X_1=0$, 1, which means that on the average 48.45% of the total patients will die before reaching to 50th year of their age for both levels of cellularity, and on the average 59.24% of the patients will have less than 100% cellularity level for both levels of their age.

Taking $h_2(t)$ as the estimate for hazard, we found slight change in the relative risk. The estimated risk of dying for patients of age 50 years or below declines to 0.1256, which indicates that 12.56% of the patients are expected to die at age 50 or below, when cellularity is not considered in the model.

Similarly, the patients at risk with cellularity level of 100% or lesser is 0.1397 i.e. 13.97% of the patients are expected to die when their cellularity level is 100% or lesser and age consideration is not counted. The relative risk when both the covariates are considered in the model is 0.4492 (when $X_{1=0,}$ $X_2 =0$, 1) and 0.4240 ($X_2=0$, $X_1=0$, 1) Which mean that 44.92% of the total patients are expected to die before celebrating their 50th birthday when considering both levels of cellularity. The second measure shows that 42.40% are expected to die of those having cellularity level of 100% or lesser for both levels of age.

Incorporating $h_3(t)$ in the Cox Proportional Hazard Model, the Base line Hazard along with the estimate of the regression co-efficient changes, but still much closer to those obtained in case of $h_2(t)$. The Patients of age 50 years or below now possess the risk of 0.1257 i.e. 12.57% Patients are expected to die at age 50 years or below, when cellularity is ignored altogether.

The Patients at risk with cellularity level of 100% or lower is 0.1320, which means 13.20% of the Patients with cellularity level 100% or lower are expected to die, ignoring the age factor. The relative risk when both age and cellularity level are simultaneously included in the model is 0.4487 (for $X_1=0$, $X_2=0$, 1) and 0.3748 (for $X_2=0$, $X_1=0$, 1). This leads to the conclusion of that 44.87% Patients are expected to die at the age of 50years or below when their cellularity is at any

level. Further, 37.48% are expected to die when their cellularity level is 100% or lower, for both categories of age.

Overall no discrimination can be made between $h_2(t)$ and $h_3(t)$ on the basis of their performance while $h_1(t)$ is misleading due to negative sign of the age coefficients. The third way of hazard rates is slightly better than the second one, due to lesser P-values for the significance of the estimates.

4. Conclusion

Adult acute myeloid leukemia (AML) is a type of cancer in which the bone marrow makes abnormal myeloblasts (a type of white blood cell), red blood cells, or platelets. In this study AML data has been analyzed due to its importance in life threatening.

Proportional Hazard regression Model is a nonparametric approach usually used for the survival time analysis depends on covariates. It is customary to fit the model with the log hazard function as the response variable. In this paper analysis has been made of the survival times using the Cox proportional hazards models with two explanatory variables namely age of the patient at death and cellularity level. Three different estimates of hazard functions are used as response variables and the regression coefficients are estimated. The main objectives are to know whether the said prognostic variables have a significant relationship with survival times on one side and to compare the different hazard function on the other. We obtain then the Risk of Dying before or after 50 years age and the Risk of having 100% or less than 100% cellularity level. We also estimate the Relative Risk considering both covariates simultaneously in the model.

Using first way of hazard rate $h_1(t)$ the marginal estimated risk of dying for patients of age 50 years or below is very low, that is 15.42% of the patients are expected to die at age 50 years or below. The marginal estimated risk of patient, having cellularity level of 100% or less is also very low, that is 11.54% of the patients having the cellularity level of 100% or less expected to die. The relative risk when both the covariates are considered simultaneously implies that the interaction of the variable is more important. Having partial estimated risks, on the average 48.45% of the total patients will die before reaching to 50th year of their age for both levels of cellularity, and on the average 59.24% of the patients will die having less than 100% cellularity level for both levels of their age.

Taking the second way of hazard rate $h_2(t)$, we found slight change in risks. The marginal estimated risk of dying for patients of age 50 years or below 12.56% and the patients at risk with cellularity level of 100% or lesser is 13.97%. The partial relative risk for interaction in the model is 0.4492 (when $X_1=0$, $X_2=0$, 1) and 0.4240 ($X_2=0$, $X_1=0$, 1). Which mean that 44.92% of the total patients are expected to die before celebrating their 50th birthday when considering both levels of cellularity. The second measure shows that 42.40% are expected to die of those having cellularity level of 100% or lesser for both levels of age.

Estimates of the regression coefficients based on $h_3(t)$ are much closer to those obtained from $h_2(t)$. The Patients of age 50years or below now possess the marginal risk of 12.57% and with cellularity level of 100% or lower, it is 13.20%. The partial relative risk when both age and cellularity level are simultaneously included in the model is 0.4487 (for $X_1=0$, $X_2=0$, 1) and 0.3748 (for $X_2=0$, $X_1=0$, 1). The P-values for the third form of hazard rates are smaller shows an edge over the other.

References

1. Bennett JM, Catovsky D, Daniel MT (1985): *Proposed revised criteria for the classification of acute myeloid leukemia*: A report of the French-American-British Group. *Ann Intem Med*; 103:460-462.

2. Cox, D. R. (1972). Regression Models with Life Tables (with discussion). *Journal of the Royal Statistical Society*, Series B, 34: 187-220.

3. Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

4. Linet MS, Devesa SS (1991): *Descriptive epidemiology of the leukemias*, in Henderson ES, Lister TA ledS: Leukemia, ed 5. *Cancer Facts and Figures- 1991.* Atlanta, American Cancer Society Inc, pp 1-31.