Estimation of Relative Hazard between Groups of Cancer patients in North West Nigeria

[1*]Boyi Abubakar Dalatu, [2]Shehu Ladan, [3]Nwoji O. Jude, [4]Shamsuddeen A. Sabo

[1,2]Department of Statistics, Waziri Umar Federal Polytechnic, Birnin Kebbi, Nigeria.
[3]Department of Computer Science, Waziri Umar Federal Polytechnic, Birnin Kebbi, Nigeria.
Department of Mathematics, Mewar University, Chittorgarh, Rajasthan, India.
*abubakarboyidalatu@gmail.com and shela909@gmail.com

ABSTRACT

A problem in population-based cancer survival analysis is the estimation of relative hazard between patients with different characteristics handled in the same registry center. In this work, we use the Cox proportional hazard model to find an estimated hazard ratio between two patients having different characteristics. The procedure was then fitted has been applied on the data collected from Ahmadu Bello University Teaching Hospital (ABUTH) Zaria Cancer Registry Center. The result indicated that female and frail patients are more prone to failure than male and non frail patients respectively.

Keywords : Population Based, Cancer Patients, Relative Hazard, Cox Proportional Hazards

I. INTRODUCTION

Survival analysis refers to collection of statistical procedures used for the study of random variable T, which represents the time between entry to a study of patients and occurrence of events such as death or recurrence of disease. The analysis is often called time-to-event analysis and the outcome of interest is the elapsed time between a well-defined starting point and a well-defined end point. The study focuses on predicting the probabilities of response, survival, or mean lifetime, comparing survival distributions of patients and identification of risk and prognostic factors related to response, survival and development of a disease. In such studies, it is usual to make comparison between groups with different characteristics. A measure of the increased risk of failure or risk of contracting a particular disease or attaining a particular response is required. When it is desired to study simultaneously, the effect of explanatory variables upon survival, the hazard function is modeled to include a vector of covariates such as age, sex, cancer type and frailty.

II. EXISTING WORK

The survival experience of a cohort can be measured using either the survivor function, or the hazard function, which are mathematically related. It is usual to work with survivor function for descriptive

analysis; and the hazard function for assessing the association between explanatory variables and survival of patients which usually requires statistical modelling.

Survivor Function

The survival time for an individual can theoretically take on any non-negative value and is represented by the non-negative random variable *T*. The actual survival time for an individual is denoted by *t* and is assumed to be a realization of the random variable *T*. This random variable *T* is generally expected to have a probability distribution function $f(t)$ and corresponding cumulative distribution function $F(t) = P_r(T \le f) = \int_0^1 f(x)dx$. The probability of an individual surviving until at least time *t* is given by the survivor function, $s(t)$, (sometimes called the survival function)

$$s(t) = P_r \sim (T > t) = 1 - F(t) \qquad (0.1)$$

The Hazard Function

The hazard function denoted by $\lambda(t)$, describes the instantaneous death rate at time *t*, conditional on survival up to time *t*. In contrast to the survivor function which describes the probability of not failing before the time *t*, the hazard function focused on the failure rate at the time *t* given by

$$\lambda(t) = \frac{f(t)}{s(t)} \qquad (0.2)$$

That is a lower value for $\lambda(t)$ implies a higher value for $s(t)$ and vice-versa. Hazard is a rate not probability and therefore can take on any value between 0 and infinity as opposed to $s(t)$ which is restricted to the interval $[0,1]$.

Cox proportional hazard model

The model proposed by Cox (1972) has been used primarily in medical research to model the effect of explanatory variables on survival. Its strength lies in its ability to model and test many inferences about survival without making any specific assumptions about the form of the life distribution model. Thus, suppose that the data collected on n subjects are denoted by $(t_1, \delta, \mathbf{X})$ where t_1 is the time to failure of the *ith* subject, δ is the censoring indicator such that for the *ith* subject $\delta_i = 1$ if a subject is observed to fail and $\delta_i = 0$ if the time is right censored (i.e. we observe some value c with the knowledge that $t_i < c$) and $\mathbf{X}$ is a p-dimensional vector of covariates. Cox (1972) model assumes that the hazard function for the *ith* subject with covariate value $\mathbf{X}$ has the form

$$\lambda(t_i, \ X) = \lambda_0(t)\exp\beta'\mathbf{X} \qquad 2.0$$

Where $\lambda_0(t)$ is arbitrary baseline hazard function and β' is a p-vector of unknown regression coefficients. Under the proportional hazards model (2.0), it is possible to analyze survival data and compute maximum likelihood estimate and use likelihood ratio test to determine which explanatory variables are significant to failure or survival. The log-rank test is derived as the score (Rao) test from Cox proportional hazard model (2.0). Hanagal (2011)

Survival Analysis of Cancer Patients Data collected from ABUTH Zaria Cancer Registry Centre with Covariates

The Data, table 3.1, on the survival times of about 400 cancer patients was collected from the Ahmadu Bello University Teaching Hospital (ABUTH) Zaria, Cancer Registry Centre. The center is one of the eleven (11) cancer centers charged with the responsibilities of

collecting and classifying information about cancer patients and prevalence in north western Nigeria. More than 50% of the patients are female. Cancer registry centers do not admit patients. This phenomenon makes heavy censoring evident which is a common feature of population based survival data. The data is prepared in such a way that the most common cancer cases i.e. breast and cervical are considered as group one (1) type, while all other cases are grouped under group two (2). To consider cases of random effects, all patients suffering from other diseases than cancer were considered frail. Thus, frailty is treated as an indicator variable with unity (1) for frail and zero (0) for non frail patients. Other variables considered are the age of patient at the time of diagnosis and the sex of the patient. For convenience, the value of the variable sex is reduced to unity (1) for females and zero (0) for males respectively. Thus the variables are: - age, sex, frail and group. Hence, sex, frail, and group are all factor variables with two levels each and are fitted using indicator variable sex, frailty and group respectively. We use the hazard model since our interest is to ascertain the influence of the covariates on survival.

Table 3.1 a sample of cancer data collected from ABUTH Zaria Cancer Registry Centre (in appendix va)

id	sex	age	time	frailty	status	group
1	1	39	30	1	0	1
2	0	50	480	1	0	2
3	0	34	30	0	0	2
4	1	40	270	0	0	1

Therefore, the proportional hazard model for the ith individual with the said variables is

$$\lambda_i(t) = \exp(\beta_1 Age_i + \beta_2 sex_i + \beta_3 frailty_i + \beta_4 group_i)\lambda_0(t)$$
.... (3.1)

Where the subscript i on an explanatory variable denotes the value of that variable for the ith individual

Relative Risk or Relative Hazard

From (2.0), the hazard ratio can be expressed as

$$\frac{\lambda(t, \mathbf{X}_i)}{\lambda(t, \mathbf{X}_j)} = \frac{\lambda_o(t)\exp(\beta'\mathbf{X}_i)}{\lambda_o(t)\exp(\beta'\mathbf{X}_j)} = \exp\left[(\beta'\mathbf{X}_i) - (\beta'\mathbf{X}_j)\right]$$

(3.2)

Which is constant for all time points; equivalently

$$\log\left[\frac{\lambda(t, \mathbf{X}_i)}{\lambda(t, \mathbf{X}_j)}\right] = (\mathbf{X}_i - \mathbf{X}_j)^T \beta', \quad for\ all\ t \geq 0$$

(3.3)

So that with one unit increases in X_k while other covariates values being held fixed,

$$\log\left[\frac{\lambda(t, X_k + 1)}{\lambda(t, X_k)}\right] = \log(\lambda(t, X_k + 1) - \lambda(t, X_k)) = \beta_k$$

(3.4)

Therefore, β_k is the increase in the log hazard at any time with unit increase in the kth covariate X_k. Equivalently,

$$\frac{\lambda(t, X_k + 1)}{\lambda(t, X_k)} = e^{\beta_k}, \quad for\ all\ t \geq 0$$

(3.5)

So, $\exp(\beta_k)$ is the hazard ratio associated with one unit increase in the X_k. This quantity is referred to as relative risk.

If gender is considered; then the relative risk of female with indicator variable unity (1) with respect to male with indicator variable 0 is

$$\frac{\lambda(t/X = 1)}{\lambda(t/X = 0)} = \frac{\lambda_o(t)\exp(\beta)}{\lambda_o(t)} = \exp(\beta)$$

(3.6)

Similarly, equation (3.6) above can suffice if we intend to estimate the relative risk of frail patients

with respect to non frail since their indicator variables are equally unity (1) and 0. Estimate of coefficient β greater than 0 is an indication of increased risk of failure thereby decreasing survival time, while estimate of β less than 0 is evidence of decreased risk and increased survival

The Working Hypothesis

The working hypotheses is that; there is no relationship between risk of failure of cancer patients and the sex or cancer type. Our interpretation would focus on the estimated values of the regression coefficients i.e. β and the probability values $\left(p-value \right)$, and the value of the likelyhood ratio test as well. Where positive value of β indicates greater

risk of failure and negative value indicates increased survival time. If $p-value$ is small, the test is viewed as significant and therefore the null hypothesis is unlikely to produce more extreme values than the observed one. Hence if the $p-value$ is less than the (0.05, which is the common practice in survival analysis) significant level, the null hypothesis is rejected

Result of Proportional Hazard Model Fitting on ABUTH Data

Below is the result of fitting the proportional hazard model to the ABUTH Cancer data.

Relative Risk of Female with Respect to Male Patients, In Respect Of (ABUTH) Cancer Data.

Table 4.2 : Result of Relative Risk of Female with Respect to Male Patients

	Coef	exp(coef)	se(coef)	z	Pr(>|z|)		exp(coef)	exp(-coef)	lower .95	upper .95
sex	-0.5339	0.5863	0.7646	-0.698	0.485	sex	0.5863	1.706	0.131	2.624

Rsquare = 0.001 (max possible= 0.179)
Likelihood ratio test = 0.47 on 1 df, p = 0.4928
Score (log rank) test = 0.5 on 1 df, p = 0.4798

Discussions on Relative Risk of Female with Respect to Male Patients in Respect of BUTH Cancer Data
Table 4.2.0 is a result of proportional hazard fitting when only sex is considered as an explanatory variable. Referring to section 3.3, the relative risk of female with respect to male is given as which in this result is 0.586. Therefore, female is 0.586 more likely to fail than male. The values of the likelihood ratio and that of the log-rank, with their respective p-values are all evidence for rejection of the null hypothesis of no relationship between the covariates and hazard.

III. CONCLUSION

On fitting the proportional hazard model on the cancer data collected form ABUTH cancer centre, the relative risk analysis of failure of female patients with respect to male indicates that, female are 0.586 more likely to fail than male. This could be due to the fact that, the two most common cancer types are more common with female patients.

IV. REFERENCES

[1]. Collet D (2003). Modeling Survival Data in Medical Research. Chapman & Hall/CRC

[2]. Cox DR (1972). "Regression Models and Lifetables" Journal of the Royal Statistical Society Series B (Methodological) 34(2) 187 – 220.

[3]. Dickman, P.W., Hakilinen, I. (2004). Population Based Cancer Survival Analysis. Chapman & Hall.

[4]. Hanagal D (2011). Modelling Survival Data using frailty models. Chapman & Hall/CRC Press. Taylor and Francis Group, LCC.

[5]. Pohar, M. and Stare, J. (2006). Relative Survival in R: Computer Methods and Programs in Biomedicine. 81(2006) 272 – 278. Journal Home Page : www.intt.elsevierhealth.com/journals/cmpb

Cite this article as :