Evaluating the Influence of Frailty in Survival Analysis of Cancer Patients in North-Western Nigeria

Boyi Abubakar Dalatu^{1*}, Shehu Ladan², Faruk Abdullahi Zagga³ and Ishaq .O. Olawoyin⁴

^{1,2,3}Department of Statistics, Waziri Umar Federal Polytechnic, Birnin Kebbi, Nigeria.

⁴Department of Statistics, Kano University of Science and Technology, Wudil, Nigeria.

*abubakarboyidalatu@gmail.com, shela909@gmail.com, zagga07@yahoo.com,
babington4u@gmail.com

Tel: 08065523733

Abstract

Not much is known about the influence of frailty along side other explanatory variables on cancer patients' survival in the north-western Nigeria. In this work, the impact of variables; age sex cancer type, and frailty on patient survival has been estimated using Cox proportional hazards model. Fitting the model to the cancer data obtained from Ahmadu Bello University Teaching Hospital (ABUTH) Zaria, Cancer Registry Center, the results indicate that the covariates studied are relevant prognostic factors to patients' survival. Also, presence of frailty has impact on the estimates. Female patients were found to be slightly more prone to failure than the male patients.

Key words: Cancer Survival, Covariates, Frailty, Cox Proportional Hazards

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. In time to event analysis, the outcome of interest is the elapsed time between a well-defined starting point and a well-defined end point. Sometimes, we would like to describe the survival experience of the population in terms of prognostic factors such as age, sex, year of diagnosis, stage and region of residence. Sometimes there are some characteristics that are shared by more than one observed event times, leading to randomness which leads to the concept of random effect known in biostatistics as frailty. In this approach, the hazard function is modeled when it is desired to study simultaneously, the effect of prognostic factors (variables) upon survival. Patients' survival is the most important single measure for monitoring and evaluating cancer control activities. Evaluation of patients' survival at population level is performed using population-based data collected by cancer registries. A registry is a center whose role is to collect and classify information on all cancer cases in order to produce statistics on the occurrence of cancer in a defined population and to provide a frame work for assessing and controlling the impact of cancer in the community [1].

Existing Work

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 modeling. Our focus therefore should be on the hazard function.

Cox proportional hazard model

The model proposed by [2] 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 $(t_i, \delta, \mathbf{X})$ where t_i is the time to failure of the i th subject, δ is the censoring indicator such that for the i th 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. [2] model assumes that the hazard function for the i th subject with covariate value $\mathbf{X}$ has the form

$$\lambda(t_i, \mathbf{X}) = \lambda_0(t) \exp \beta' \mathbf{X} \quad (1)$$

Where $\lambda_0(t)$ is arbitrary baseline hazard function and β is a p -vector of unknown regression

In frailty model, it is assumed that the hazard function can be separated into components of frailty, baseline hazard and linear predictor. Thus for individual j in group i , the right hand side of (1) would be

$$z_i \lambda_0(t) \exp(\mathbf{X}_{ij} \beta') \quad (2)$$

Under the proportional hazards model (1), 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 (1). [3] observed that the log-rank test is more suitable when the alternative to the null hypothesis of no difference between two groups of survival times is that the hazard of death at any given time for an individual in one group is proportional to the hazard at that time for a similar individual in the other group. This is the assumption of proportional hazard, which underlies a number of methods for analyzing survival data. Log-rank test is the optimal non parametric test for testing the null hypothesis $\beta' = 0$ in the proportional hazards model when studying the influence of the explanatory variable $\mathbf{X}_i$ on the survival time T_i of individual $i = (i = 1, 2, \dots, n)$, when the baseline hazard function is totally unknown

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

The Data, table 1.0, 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 considered as a factor variable with unity (1) for frail and zero (0) for non-frail. 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.

Table 1.0 a sample of cancer data collected from ABUTH Zaria Cancer 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, proportional hazard model for the i^{th} individual with the said variables is

$$\lambda_i(t) = \exp(\beta_1 \text{Age}_i + \beta_2 \text{sex}_i + \beta_3 \text{frailty}_i + \beta_4 \text{group}_i) \lambda_0(t) \quad (3)$$

Where the subscript i on an explanatory variable denotes the value of that variable for the i^{th} individual

The working hypothesis

The working hypothesis is that; there is no relationship between survival time of cancer patients and the explanatory variables age, sex and frailty. Our interpretation would focus on the estimated values of the regression coefficients i.e. β and the probability values (p -value), were 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.

Rejection of the working hypothesis will imply that the explanatory variables in our study have some influence or impact on the survival or hazard of patients. It also implies that belonging to a group may be a factor to patients' failure or survival. On the other hand, accepting the null hypothesis will imply that hazard and survival depends solely on the disease and no any contribution from the explanatory variables. In practice, if one of the explanatory variables indicates a rejection of the null hypothesis, the test is viewed as significant [4].

Result

Below is the result of resting the proportional hazard model to the ABUTH Cancer data. We first considered analysis without frailty, and subsequently considered frailty so as to examine how much the presence of frailty will affect the estimates and overall result

Result of (ABUTH) Cancer Data Analysis; Estimation when Frailty is not considered

Table 2.0 Result of the estimated values of the explanatory variable on fitting the proportional hazard model

Coe	exp(coef)	se(coef)	z	Pr(> z)	exp(coef)	exp(-coef)	lower .95	upper .95
sex	0.416996	1.517397	0.810775	0.514	0.6070	sex	1.5174	0.6590
age	-0.001737	0.998264	0.025955	-0.067	0.9466	age	0.9983	1.0017
group	2.229487	9.295097	1.146044	1.945	0.0517	group	9.2951	0.1076

Rsquare= 0.014 (max possible= 0.179)

Likelihood ratio test = 5.57 on 3 df, p=0.1346

Score (logrank) test = 5.37 on 3 df, p=0.1467

Discussions on Result of (ABUTH) Cancer Data Analysis; Estimation when Frailty is not considered

On fitting the proportional hazard model ignoring the presence of frailty, table 4.1.0 shows that the p -value in the case of group indicates that belonging to a group could have an influence on patient's failure. While p -value for sex and age do not indicate evidence to

reject the null hypothesis (of no relationship between the variables and hazard). The negative estimated coefficient in the case of age suggests increased survival and decreased risk of failure. The likelihood ratio test value is 5.57 on 3 degree of freedom (and big enough for rejection of the null hypotheses) and the p-value in respect the variable group render the test significant and therefore we reject the null hypothesis.

Results of (ABUTH) Cancer Data Analysis; Estimation when Frailty is included

Table 3.0 Result of Estimation when Frailty is included

coef	exp(coef)	se(coef)	z	Pr(> z)	exp(coef)	exp(-coef)	lower .95	upper .95		
sex	0.399212	1.490649	0.811408	0.492	0.6227	sex	1.4906	0.6708	0.30388	7.312
age	-0.002412	0.997591	0.025571	-0.094	0.9248	age	0.9976	1.0024	0.94883	1.049
frailty	-0.580329	0.559714	1.082211	-0.536	0.5918	frailty	0.5597	1.7866	0.06711	4.668
group	2.243404	9.425357	1.146875	1.956	0.0505	group	9.4254	0.1061	0.99558	89.232
Rsquare = 0.015 (max possible= 0.179)										
Likelihood ratio test = 5.9 on 4 df, p=0.2069										
Score (logrank) test = 5.69 on 4 df, p=0.2233										

Discussion on Results of (ABUTH) Cancer Data Analysis; Estimation when Frailty is included

On fitting the proportional hazard to the cancer data considering the presence of frailty, table 4.2.0 the *p* – value in respect of the variable, group, perfectly shows evidence to reject working hypothesis (of no influence the variables on hazard or survival). Inclusion of frailty slightly reduces the values of the estimated coefficient. The values of the likelihood ratio and that of the log-rank, with their respective p-values and the p-value in respect of the variable group, are all evidence for rejection of the null hypothesis of no relationship between the covariates and hazard. Inclusion of frailty does not significantly change the result.

Conclusion

On fitting the proportional hazard model on the cancer data collected from ABUTH cancer centre, the results indicate that the analysis in the presence of frailty slightly reduces the values of the estimated coefficients, though, it does not significantly change the final decision on the results. In the final analysis, the results show that, the explanatory variables are relevant prognostic factors to survival or hazard of cancer patients.

References

- I. P. W. Dickman, I. Hakilinen. Population Based Cancer Survival Analysis. Chapman & Hall, 2004.
- II. D. R. Cox DR. “Regression Models and Lifetables” Journal of the Royal Statistical Society Series B (Methodological) Vol. **34**(2) pp.187 – 220, 1972.
- III. D. Hanagal. Modelling Survival Data using frailty models. Chapman & Hall/CRC Press. Taylor and Francis Group, LCC, 2011.
- IV. D. Collet. Modeling Survival Data in Medical Research. Chapman & Hall/CRC, 2003