

Prognostic Models with Competing Risks: Methods and Application to Prostate Cancer Data

N.P.K.S. JAYAWARDANA AND M.R. SOORIYARACHCHI

Department of Statistics, University of Colombo, Sri Lanka

ABSTRACT

Fundamental statistical methods for analyzing competing risks data have been in discussion for decades. However there is still an uncertainty on how to approach this type of data due to its complexities and thus there exist several gaps in the available methodology particularly in the area of modelling and model validation. Hence, this has become a topic of interest for many researchers. We review competing-risk regression models with a view toward: testing for prognostic factors, testing for treatment effects after adjusting for prognostic factors and model validation. An example of prostate cancer data from a French study is used to illustrate the methods examined. This includes the application of the Lunn and McNeil regression model for testing prognostic factors and treatments and the adaptation and modification of a goodness-of-fit test, suggested in the literature, to test the hypothesis whether to include the covariates in a multiplicative Cox proportional hazards model, against the hypothesis whether to include the covariates in a more general class of additive-multiplicative model. Serum prostatic acid phosphatase, Combined index of stage and histological grade, Size of primary tumor, Serum hemoglobin level, Performance rating and age were identified as the more vital factors for the survival of patients from death by prostate cancer. Furthermore, the active treatments (different levels of estrogen) significantly affects time to death by prostate cancer, where the survival experience of patients showed improvement for higher doses of The goodness of fit test indicated that the model fit was adequate and that all prognostic factors in the model had a multiplicative effect on hazard.

Keywords: Additive-Multiplicative Hazards model, Competing risks, Cumulative Incidence Function, Proportional Hazards, Stratified Cox model

INTRODUCTION

Survival analysis focuses on one possibility for the occurrence of the event. Often times the subjects under investigation are exposed to more than one possibility of experiencing an event of interest and this is known as competing risks. In the real world competing risks situation is often predominant, where the end point consists of several distinct events of interest. For example in a treatment episode, treatment

maybe discontinued, switched, a supplement treatment may be given or death may occur. To investigate the effect of one specific cause to the end result, the other risks affecting the outcome should be controlled. Basic statistical methodology for analyzing competing risks data have been discussed for decades. The fundamental tools for the analysis of survival data in the presence of competing risks were developed in the 1950s, however there is still an uncertainty on how to analyze this type of data and how to use the numerous tools and methods introduced by researchers interested in this area of statistics, over a period of time. The Main problem in a competing risks situation includes the unknown and inestimable interrelations among competing risks and the interpretation of the results in the presence of such situations. Many research papers are on quite a formal mathematical level where the concentration has been mainly drawn upon mathematical functions and definitions. Particularly the area of modelling competing risks has several gaps where many papers discuss the mathematical details of the modelling procedure and no proper application to real world clinical trials is available. An area that is really lacking in development is model validation where only a single paper (Sun *et al.*, 2006) could be found that discussed goodness of fit techniques for this type of model.

Based on the gaps in the literature the following objectives were formulated. The primary objective of the study is to illustrate methodology for the modelling of competing risks data and to adapt and extend the only model validation procedure in the literature, by way of an example. A secondary objective is to determine the factors affecting the survival of prostate cancer patients in the presence of competing risks.

The data set was taken from a randomized clinical trial conducted during 1967-1969, comparing treatments for prostate cancer and reported in Byar and Green, (1980). The full data set is given in Andrews and Herzberg (1985). There are 18 explanatory variables. The response variable of interest is the months of follow-up of the patients. The data set has been collected on 502 prostate cancer patients, where after removing patients with unknown cause of death, there are 495 patient records. Here, the interest is focused on the death by Prostate cancer when the other competing risks are controlled. The potential prognostic factors together with the outcome variable, their levels and the meanings of the levels are given in table 1.

Table 1: Description of data used in the study

Variable	Notation	Levels	Code
The patient number	patno		
Stage	stage	3 - local extension beyond the prostate gland	1
		4 - distant metastases, elevated acid phosphatase or both	2
Treatments of estrogen	Rx	Placebo	1
		0.2mg estrogen	2
		1.0mg estrogen	3
		5.0mg estrogen	4
Age (in years)	Age	≤60	1
		61 – 70	2
		>70	3
Weight index = wt(kg) – ht(cm) + 200	Wt	<85	1
		85-104	2
		105-124	3
		>125	4
Performance rating	Pf	Normal	1
		confined to bed ('limitation of activity')	2
History of cardiovascular disease	Hx	Yes	0
		No	1
Systolic blood pressure/10	Sbp	<16	1
		≥16	2
Diastolic blood pressure/10	Dbp	<9	1
		≥9	2
Ekg	Ekg	Normal	1
		Benign	2
		rhythmic disturb	3
		Heart block or strain	4
		MI	5
Serum hemoglobin(g/100ml)	Hg	≤10	1
		10<hg≤15	2
		15<hg≤25	3
Size of primary tumor (cm ²)	Sz	0-19	1
		20-39	2
		40-69	3
Combined index of stage and histological grade	Sg	5-9	1
		10-15	2
Serum Prostatic Acid Phosphatase	Ap	≤10	1
		>10	2
Bone Metastases	Bm	0	0
		1	1
Date on study	sdate		

Table 1 Contd...: Description of data used in the study

Variable	Notation	Levels	Code
The status given by survival from different causes (Outcome)	status	Alive	0
		dead-Prostate cancer	1
		dead-cerebrovascular	2
		dead-heart or vascular problems	3
		dead-other cancer	4
		dead- non-cancer	5
		dead-lung and respiratory diseases	6
Months of follow-up	dtime (tym)		

Section 2 includes a review of the available literature on modelling competing risks and model validation procedures. Section 3 gives the methods used/extended in this study. Section 4 presents the analysis and results obtained by applying the methodology in section 3 to the data. Section 5 gives conclusions and recommendations based on the study.

REVIEW OF LITERATURE ON COMPETING RISK MODELS

Since 1950's there has been vast amount of discussions in literature on the models that could be used in the presence of competing risks. Gail (1975) reviewed and gave criticisms on some competing risks models proposed by various researchers. He introduced a notation, allowing the models to be defined easily. Kay (1986) used an approach described by Kalbfleisch and Prentice (1980), which involved fitting separate models for each type of failure, treating other failure types/competing risks as censored data. A main drawback identified in this model is that the different types of failures are not treated jointly, thus the parameter estimates of different failure types could not be compared directly. To avoid this drawback alternative approaches have been used where more complex models are fitted considering different failure types. Larson and Dinse (1985) introduced a mixture model with a multinomial distribution for types of failures and a piece-wise exponential distribution for conditional distribution of failure times given type of failure. Since this provides a regression framework, associated variables could be adjusted for and their effects on the joint distribution of time and type of failure could be assessed. This model does not assume independence between the types of events or the competing risks. Kuk (1992) also followed this approach of fitting more complex models incorporating the different failure types. But in this approach, difficulties arise due to the inadequacy of the standard software tools.

Lunn and McNeil (1995) demonstrated that standard software could be used in the analysis of survival data in the presence of competing risks. The authors presented two methods (Method A and Method B) for joint estimation of parameters in models for competing risks, where Cox's proportional hazards regression model was used in both cases. The method, initially assuming two types of failures in addition to censoring ($\delta = 0$ or 1), adopts the different types of failures by augmenting the data using a data duplication method. Each entry is duplicated, one for each failure type but the second entry always censored. Here the hazards of the two types of risks are assumed to be additive, so the hazard of failure is the sum of those two and the time to failure is the minimum of two failure times. Method A uses δ as a covariate in an unstratified Cox regression, assuming that the hazard functions associated with the two types of failures has a constant ratio. In general proportionality of the hazards cannot be assumed, but a good fit can be obtained by taking separate time zones and separating the model based on that. Method B uses δ as a stratifying variable in a Cox regression, stratified by the type of failure. Here assumptions on proportional baseline hazards are not required. The advantages of the augmented data approach are that it can run immediately on the existing software packages and that it does not over-parameterize the model. Ali and Babiker (2002) demonstrated the 2 approaches proposed by Lunn and McNeil (1995), by fitting those models on augmented data using 'stcox' in stata. Their programme produced cumulative incidences with point-wise confidence interval apart from the parameter estimates and their standard errors. They performed this demonstration duplicating the data k times, for the k failure types. .

A semi-parametric proportional hazard model was proposed by Fine and Gray (1999), for the sub distribution also known as the cumulative incidence function (CIF) (Pintilie, 2006). The form of the partial likelihood used here is similar to that used in Cox proportional hazards model (Cox, 1972). This versatile method of modelling the hazard of the sub distribution was considered as a natural extension of the Cox proportional hazards model. Just as the Cox model requires the proportionality of the cause-specific hazard, this approach requires the hazards of the CIF to be proportional. The estimator obtained for a subject under this method is uniformly consistent. The availability of the software for the above analysis and the ability for incorporating time-dependent covariates make it a definite advantage.

Gelfand *et al.* (2000) proposed a semi-parametric version of the conventional proportional hazards survival model, where a flexible class of parametric specifications providing better modelling and an interpretation of the hazard as a latent competing risk model were suggested. Here also as in Lunn and McNeil, a data augmentation scheme is followed.

Many authors have proposed models that combine the Cox model and additive model (Lin and Ying, 1995). Martinussen and Scheike (2002) proposed a flexible additive multiplicative hazard model that allows both time-varying and fixed covariate effects. Here the two components of the model included additive covariate effects through an additive Aalen model (Aalen, 1980), and multiplicative covariate effects through a Cox regression model (Cox, 1972). Model was also extended to allow for time-varying covariate effects in the multiplicative part of the model. Sun *et al.* (2006) proposed a generalization of the above model to account for a competing risks situation. The authors produced inference procedures in estimating non-parametric and parametric components of the flexible additive multiplicative model. Robust variance estimation and a goodness-of-fit procedure are also proposed by these authors.

MATERIALS AND METHODS

Description

Initially the data on prostate cancer patients will be analyzed descriptively to visualize the features prevalent in the data. For this purpose Cumulative Incidence Function (CIF) (Pintilie, 2006) was used. Using CIF graphs the data set would be visualized graphically, with the hope of identifying whether the CIFs change according to the levels of each variable. Before fitting a model, the variables should be tested for inclusion in the model with the hope of preserving simplicity of the model. Thus, various tests are carried out to test the covariates. For testing the covariates with only two levels, Pepe and Mori's test (Pepe and Mori, 1993) was used. As this does not provide for testing of the covariates with more than two levels, an extension of this test by Lunn (1998) was used in this case. Having identified the variables for inclusion in the model, a Cox proportional regression model would be fitted to the data using a data augmentation procedure, following Lunn and McNeil (1995). When it comes to model fitting, two approaches of Lunn and McNeil (1995) are explored. Model validation of a proportional hazard (PH) model mainly focuses on checking the validity of the assumptions of proportionality of hazards. A goodness-of-fit test procedure proposed by Sun *et al.* (2006), based on an additive-multiplicative model presented by Martinussen and Scheike (2002), will be followed. This test will be adjusted to suit the needs of this study. There are two mathematical approaches described in Pintilie (2006) for dealing with competing risks, where both methods have been discussed in the literature in the usage of different competing risks tools. These two methods are namely, Competing risks as bivariate random variable and competing risks as latent failure times. As the latter method cannot make use of the dependence structure between times to different types of events in this study more concentration is towards bivariate approach.

Lunn and McNeil's Model for Competing Risks

Of the models reviewed in section 2, the Lunn and McNeil's model was selected for modelling due to it is based on the familiar Cox proportional hazards model (Cox, 1972) and also due to the fact that there was available software (Stata) for fitting this model. Under the augmented data approach introduced by Lunn and McNeil (1995), there are two approaches to joint estimation of the parameters in models for competing risks. These two approaches, both using a Cox proportional hazards model would be discussed here with regard to the study. Initially considering that there are only two types of failure, each data entry would be entered twice, one for each failure type, ($\delta = 0$ or 1). The covariates z_i 's are augmented to allow for possible interactions with type of failure.

The Cox proportional hazard models used in this study, with regard to the two methods, can be given as follows:

$$h_{i\delta}(t) = h_{00}(t) \exp(b_0\delta + b'z_i + \theta'\delta z_i). \quad (1)$$

$$h_{i\delta}(t) = h_{0\delta}(t) \exp(b'z_i + \theta'\delta z_i) \quad (2)$$

Here, z_i is a p -dimensional vector of measured covariates for the δ^{th} event, $h_{i\delta}(t)$ is the hazard function of the i^{th} subject on the δ^{th} failure type at time t . b 's are vectors of regression parameters to be estimated.

In model (1), $h_{00}(t)$ is the baseline hazard function that is common to all the events. (i.e. $h_{i\delta}(t) = h_{00}(t)$ for $\delta = 1, \dots, k$)

Model (2), known as a stratified Cox proportional hazard model, stratified by the failure type, allows the baseline hazard to vary over each failure type; $h_{0\delta}(t)$. Models (1) and (2) are fitted by determining in each case the values of the parameters which maximize the partial log-likelihood. Lunn and McNeil (1995) provide more details on this.

Model Validation Techniques

Among the very few methods available in the literature on model validation techniques in the presence of competing risks, the test procedure proposed by Sun *et al.* (2006) is considered here and this test is modified to suit our study. The goodness-of-fit test proposed by Sun *et al.* (2006) is based on a flexible additive multiplicative hazards model. This model generalizes the flexible additive multiplicative model proposed by Martinussen and Scheike (2002) for survival analysis, to accommodate for competing risks. This model allows both time-varying and fixed covariate effects, where the two components of the model includes additive covariate effects through an additive Aalen model (Aalen, 1980), and multiplicative covariate effects through a Cox regression model (Cox, 1972). The model used by Sun *et al.* (2006), is as follows:

$$h_1(t; x, z) = \alpha'(t)x + h_{10}(t) \exp(\beta'_0 z) \quad (3)$$

Where x and z are vectors of covariates of dimensions q and p , $\alpha(t)$ is an unknown q -vector of time varying components representing the effects of covariates x on h_1 , β_0 is a p -vector of unknown regression parameters denoting effects of z on h_1 , and $h_{10}(t)$ is an unspecified baseline hazard function.

According to Sun *et al.* (2006), the following test statistic could be used to test if the contribution from each covariate towards the multiplicative part of the above model (3) is appropriate.

$$F_1 = \text{Sup}_{0 \leq t \leq \tau} \left| n^{-\frac{1}{2}} u_j(\hat{\beta}; t) \right| \quad (4)$$

Where $u_j(\beta; t)$ is the j^{th} component of $u(\beta; t)$, the score function and τ is a pre-specified constant such that $P(T > \tau) > 0$. The percentiles of this test statistic can be estimated empirically under a number of simulation processes discussed in Sun *et al.* (2006) and Lin, Fleming and Wei (1994).

Modification of the Goodness-of-fit Test of Sun *et al.* (2006)

Using the test proposed by Sun *et al.* (2006) it is possible to test the null hypothesis H_0 , whether a Cox proportional hazards model is appropriate against the alternative hypothesis H_1 , of a more general class of additive-multiplicative hazards model.

Here $H_0: \alpha(t) = 0$ (in equation 1) and $H_1: \alpha(t) \neq 0$

The same test statistic (F_1) used by Sun *et al.* (2006) can be used for testing the Cox proportional hazards model in this study where

$$u_i(t) = \int_0^t [z_i(s) - \bar{z}(s)] d\hat{M}_i(s) \quad (5)$$

is the score process for covariate z for the i^{th} subject at time t and $\hat{M}_i(t)$ refers to the martingale process for subject i at time t . Similarly as Sun *et al.* (2006), in order to approximate the limiting distribution of the test statistic $\{G_i; i = 1, \dots, n\}$ are simulated from the standard normal distribution independently of the data and $\hat{M}_i(s)$ is replaced by $G_i \hat{M}_i(s)$ in the equation (5). This gives the following result:

$$\int_0^t [z_i(s) - \bar{z}(s)] d[G_i \hat{M}_i(s)] = G_i \int_0^t [z(s) - \bar{z}(s)] d\hat{M}_i(s) = G_i u_i(t) \quad (6)$$

Replacing this result (5) in the test statistic (4) it is obtained,

$$F_1 = \text{Sup}_{0 \leq t \leq T} \left| n^{-\frac{1}{2}} G_j u_j(\hat{\beta}; t) \right| \quad ; j = 1, \dots, p \quad (7)$$

For repeated sets of normal random samples $\{G_i; i = 1, 2, \dots, n\}$ given the observed data, this gives different values of the test statistic and thus, the limiting distribution of the test statistic. By sorting these values in ascending order, percentiles can be obtained for the test statistic for each covariate. Then the p value can be obtained as,

$p = \Pr[\tilde{F}_1 > F_1]$. Suppose F_1 test for covariate z_1 gives p value $>$ the significance level. This indicates that z_1 should be included in the multiplicative part of the model, rather than in an additive part of the model in a general class of additive-multiplicative hazards model (Martinussen and Scheike, 2002 and Sun *et al.*, 2006 and). Similarly all the variables can be tested for the goodness-of-fit.

ANALYSIS AND RESULTS

Descriptive Analysis

(a) CIF plots

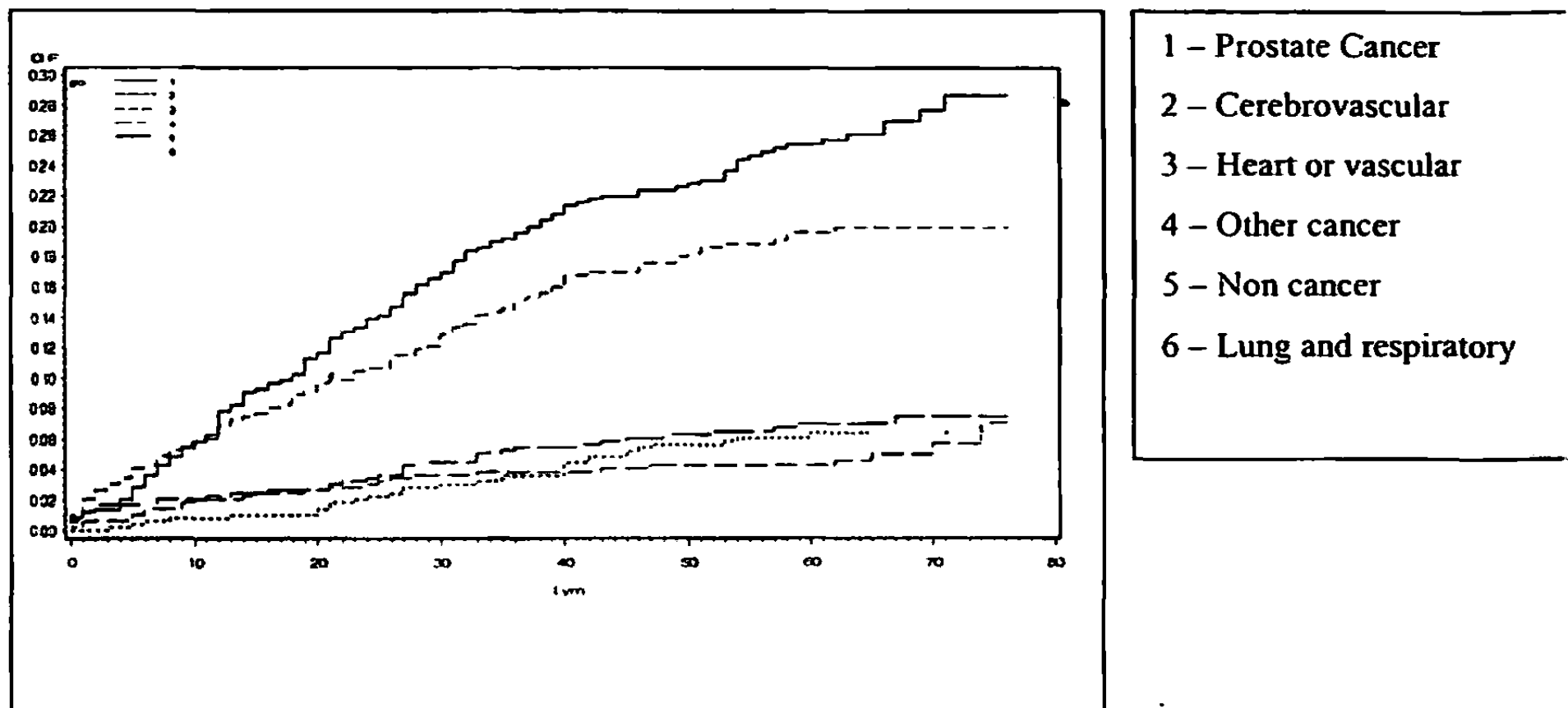


Figure 1: CIF for all diseases

From the graph it can be clearly seen that the patients have the highest risk of experiencing death by prostate cancer. This result prevails, except for the early months of follow-up ($t_{ym} < 10$), where the patients have the highest risk for death by heart or vascular diseases. Ketchandji *et al.* (2009) found that men with early stage prostate cancer and with low to moderate grade of tumor (55% in our data) experienced a survival not substantially worse than men without prostate cancer. In these men, cardiovascular disease and other cancers were the leading cause of death. CIF graphs were plotted for each prognostic factor for the outcome death by prostate

cancer. To give a flavour of this mode of presentation and interpretation, one such graph is given for the variable treatment. Figure 2 shows the CIF for death by Prostate cancer for four treatment groups indicated by "rx". Group 2 (0.2mg estrogen) shows the highest level of risk for death by Prostate cancer on or before time t, except before the period of approximately 10 months of follow-up. The placebo group (group 1) shows the next highest risk in experiencing death by Prostate cancer, and this shows the highest risk before the period of 10 months of follow-up.

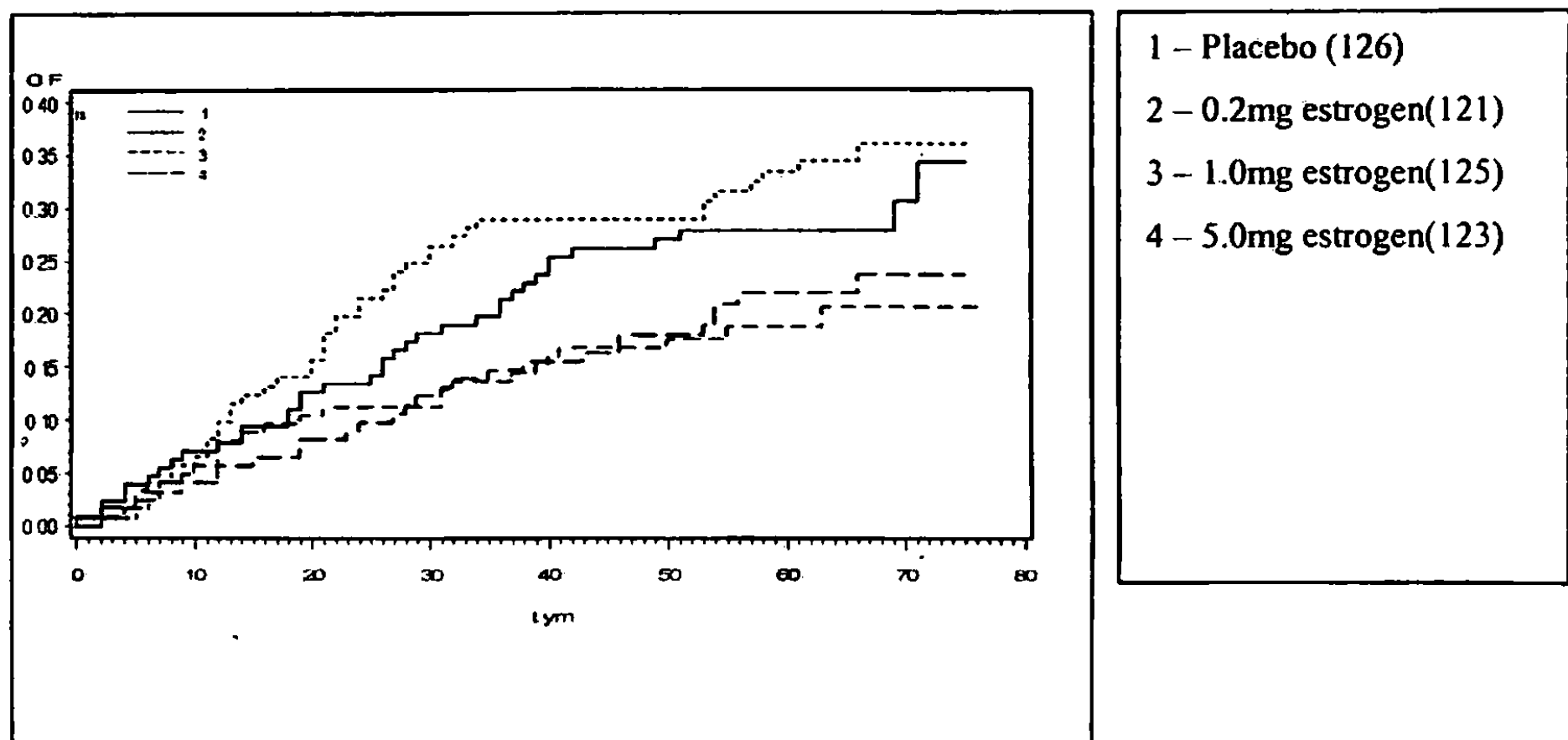


Figure 2: CIF for event 1 by rx

(b) Log Cumulative Hazard (LCH) Plots (Pintille, 2006)

The figure 3 illustrates the LCH plots for all the six diseases/causes. For prostate cancer and Cerebrovascular disease, the plots are approximately parallel. Also, LCH plot for heart and vascular diseases and lung and respiratory diseases are approximately parallel. The other lines violate the proportionality assumption as they appear to be crossing each other.

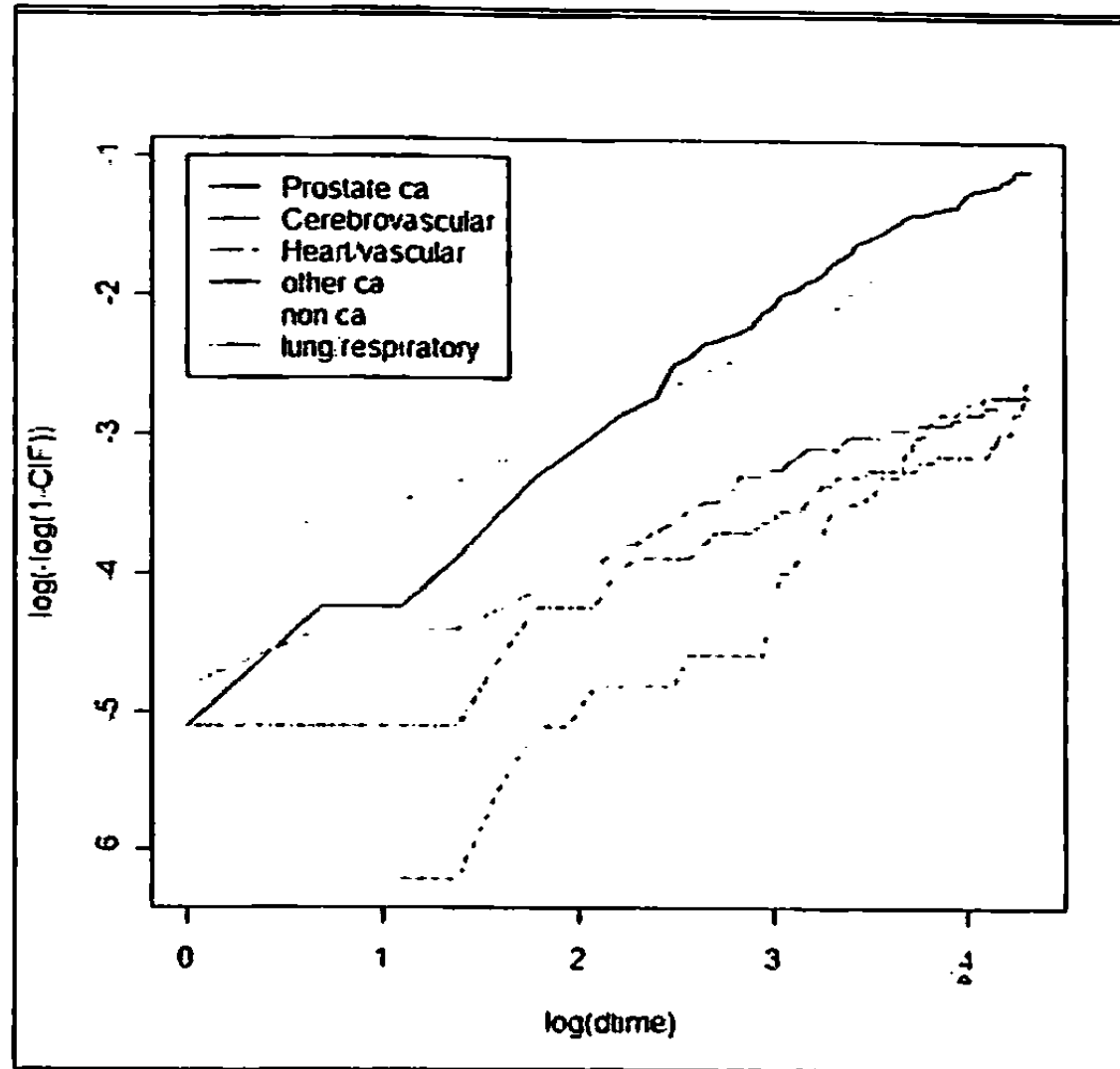


Figure 3: LCH plot for all diseases

Univariate Analysis

Having visualized the features of the data set and having investigated each and every covariate and their levels separately, these covariates must be tested for their effects before a model is fitted.

Testing the effect of covariates

Since these tests test the difference between the CIFs for the different levels of the covariates, significance of the test imply significant difference and thus the effects of the covariates are significant. Therefore those covariates could be tested for inclusion in the model.

Pepe and Mori's method (Pepe and Mori, 1993) is chosen in this study to test for covariates with two levels. As for the covariates with more than two levels the method of Lunn (1998) is used.

(a) Pepe and Mori's method

The test is performed on the eight variables which have two levels each. Results are given in table 2.

Table 2: Results of Pepe and Mori's test

Covariate	Test Statistic	p-value	Significance at 5% level
Stage	61.5996	0.00000	Significant at 1% level
Pf	4.0197	0.04497	Significant at 5% level
Hx	1.9096	0.16701	Significant at 20% level
Sbp	0.2633	0.60786	Not significant at 20% level
Dbp	0.7763	0.37826	Not significant at 20% level
Sg	80.4763	0.00000	Significant at 1% level
Ap	28.8007	0.00000	Significant at 1% level
Bm	28.8799	0.00000	Significant at 1% level

Using the outcome in the table 2, the results of the Pepe and Mori's test can be summarized as follows. In summarizing the results, covariates which are significant at a liberal 20% level, are selected. This is because covariates that are fairly significant could become significant at 5% level once the other variables are adjusted for. Results of table 2 indicate that out of the eight variables five variables are significant under 5% level of significance and one variable is significant at 20% level. The covariates "sbp" and "dbp" are not significant at the 20% level of significance.

(b) Lunn's method

A programme to perform this test is currently unavailable. Thus, the SAS macro used for the Pepe and Mori's test was modified by the authors. The macro should be run separately for each level of the covariate to obtain initial results. Having obtained the initial values intensive calculations are performed using a SAS IML programme to obtain the final results.

Table 3 presents a summary of Lunn's test applied to prognostic factors with more than 2 levels.

Table 3: Results of Lunn's test

Covariate	Test statistic	Degrees of freedom	$\chi^2_{df, 5\%}$	p-value	Significance
Rx	5.8185	3	7.8147	0.1208	Significant at 20% level
age	4.7641	2	5.9915	0.0677	Significant at 20% level
Wt	2.2428	3	7.8147	0.5236	Not significant at 20% level
Ekg	0.6091	4	9.4877	0.9621	Not significant at 20% level
Hg	9.7263	2	5.9915	0.0077	Significant at 1% level
Sz	21.7773	2	5.9915	0.0000	Significant at 1% level

Results in table 3 indicate that four out of the six covariates are significant at 20% level of significance. It is seen that the covariates *wt*: weight index and *Ekg* are not significant at the 20% level of significance.

Modelling prostate cancer in the presence of competing risks

Figure 1 in section 4.1 shows that the LCH plots for outcomes (Status) were not parallel. Thus the assumption of proportional hazards between outcomes is in doubt. In this type of situation a stratified Cox regression model should be used stratifying by failure types or the causes of death (model B of Lunn and McNeil, 1995). For modelling the variable status was regrouped in to two categories (death by prostate cancer and otherwise) as our interest was mainly to determine factors effecting survival to prostate cancer death. For modelling, the data set was adjusted to suit requirements of method B proposed by Lunn and McNeil (1995), following the instructions given by Ali and Babiker (2002). This is done using 'stcox' in stata, as Ali and Babiker (2002) proposed. To find the best model using the significant covariates identified in section 4.2, a forward selection method is used. Here the significance of the parameter estimates based on the Wald test will be the criterion considered in selecting the best model. Dummy variables were created for each covariate with more than two levels, to represent the levels of the covariates. Interactions of the covariates with the type of event (δ) and interactions of the covariates with $(1-\delta)$ are created so as to obtain the terms for the reparameterized form of Lunn and McNeil's (1995) model, and they are represented by "covariate_name* δ " and "covariate_name*(1- δ)" respectively. For death by prostate cancer $\delta = 1$ and $\delta = 0$ otherwise. For instance, the effect of 'stage' for prostate cancer death is represented by;

stage δ =stage*type

And the effect of 'stage' for death by other causes is represented by;

Stage(1- δ)=stage*(1-type)

Selecting the best model using forward selection

In the selection of the most suitable model, a forward selection procedure is followed. Variables are selected considering the significance of the parameter estimates based on the Wald z statistic. Here,

$$\text{Wald's } Z \text{ value} = \frac{\text{parameter estimate}}{\text{Standard Error}}$$

The strategy used in selecting variables is to select the most significant interaction of covariate with type and include this interaction with the interaction between covariate and (1-type). Then proceed with the forward selection procedure, selecting the remaining most important covariates with type interaction and including it with the corresponding covariate by (1-type) interaction. In this process, the priority is given to selecting the covariate-type interactions over the covariate by (1-type) interactions because the primary interest is to determine risk factors for prostate cancer death and secondarily for death by other causes. This procedure resulted in the model,

$$\begin{aligned} h_{i\delta}(t) = h_{0\delta}(t) \exp(&(-0.28761)ap(1-\delta)_i + (-0.37450)sg(1-\delta) \\ &+ (0.54108)sz1(1-\delta)_i + (0.37746)sz2(1-\delta)_i \\ &+ (-0.33504)rx1(1-\delta)_i + (-0.31386)rx2(1-\delta)_i \\ &+ (-0.47695)rx3(1-\delta)_i + (-1.38196)age1(1-\delta)_i \\ &+ (-0.39933)age2(1-\delta)_i + (0.12891)bm(1-\delta)_i \\ &+ (0.05701)hg1(1-\delta)_i + (0.07356)hg2(1-\delta)_i \\ &+ (0.53188)pf(1-\delta)_i + (0.96426)ap\delta_i + (1.4598)sg\delta_i \\ &+ (-1.41514)sz1\delta_i + (-0.79709)sz2\delta_i + (0.83317)rx1\delta_i \\ &+ (0.97965)rx2\delta_i + (-0.05754)rx3\delta_i + (0.67109)age1\delta_i \\ &+ (0.03837)age2\delta_i + (0.44264)bm\delta_i + (0.74674)hg1\delta_i \\ &+ (-0.06019)hg2\delta_i + (0.48878)pf\delta_i) \end{aligned}$$

where $\delta = 1$ for prostate cancer death
and, $\delta = 0$ for death by other causes

Goodness-of-fit test

The goodness-of-fit test proposed by Sun *et al.* (2006), modified for this study, is used here, for the purpose of further assessing the validity of the model. Here each and every covariate is checked for inclusion in the multiplicative model that is used in the present study, rather than in a more general class of additive multiplicative models.

H_0 : covariate (Z_1) should be included in the multiplicative part of the model

H_1 : covariate (Z_1) should be included in the additive part of an additive-multiplicative model

Using the modified goodness-of-fit test proposed in section 3 the test statistic, 5% critical value, p-value and significance was determined. These results are given in table 4. Table 4 indicates that for each covariate, corresponding test statistic (F_1) is less than the critical value and p value > 0.05 , (the significance level of the test). This indicates that each and every covariate given above should be included in the multiplicative part of the model rather than in an additive part of the model in a general class of additive multiplicative hazard model. This signifies the goodness-of-fit of the Cox model used in this study, over a more generalized class of additive multiplicative hazard model.

Interpretation of parameter estimates

Table 5 presents the robust parameter estimates, p values, hazard ratios and the 95% confidence intervals of the hazard ratios. Here, the significant variables under 5% level of significance in the final model are interpreted. The interpretations are based on two areas; for death by prostate cancer, which is the event of interest and for death by other causes. For covariates with two levels, the hazard ratio represents the ratio of the hazard function of the higher level with respect to the lower level, which is the baseline hazard. For covariates with more than two levels, the hazard ratio represents the ratio of the hazard function of each level versus the highest/last level, which is the baseline hazard in this case.

Table 4: Results of the Goodness-of-fit test

Covariate	F_1 (Test Stat)	5% Critical value	p-value	Significance at 5%
ap(1- δ)	0.0189729	0.060016	1.000	Not Significant
sg(1- δ)	0.0113187	0.036047	1.000	Not Significant
sz1(1- δ)	0.0152675	0.047266	1.000	Not Significant
sz2(1- δ)	0.0156286	0.047835	1.000	Not Significant
rx1(1- δ)	0.0142989	0.043791	1.000	Not Significant
rx2(1- δ)	0.0145014	0.044779	1.000	Not Significant
rx3(1- δ)	0.0189729	0.049644	1.000	Not Significant
age1(1- δ)	0.0178686	0.047938	0.919	Not Significant
age2(1- δ)	0.0156982	0.046057	1.000	Not Significant
bm(1- δ)	0.0163174	0.047328	0.999	Not Significant
hgl(1- δ)	0.0179354	0.044959	0.906	Not Significant
hg2(1- δ)	0.0155778	0.044359	1.000	Not Significant
Pf(1- δ)	0.0166199	0.046938	0.999	Not Significant
ap δ	0.0271055	0.055880	0.616	Not Significant
sg δ	0.0175702	0.044750	0.965	Not Significant
sz1 δ	0.0242253	0.050851	0.789	Not Significant
sz2 δ	0.0249778	0.053112	0.808	Not Significant
rx1 δ	0.0289328	0.058274	0.700	Not Significant
rx2 δ	0.0309362	0.060840	0.496	Not Significant
rx3 δ	0.0169832	0.043541	0.993	Not Significant
age1 δ	0.0176765	0.041383	0.942	Not Significant
age2 δ	0.0229219	0.048791	0.894	Not Significant
bm δ	0.0265907	0.054490	0.611	Not Significant
hgl δ	0.0267856	0.052283	0.552	Not Significant
hg2 δ	0.0229460	0.044577	0.837	Not Significant
pf δ	0.0281105	0.054631	0.658	Not Significant

Table 5: Details of the final model

Covariates	Coefficients (b)	p value	Hazard ratio [h]	95% CI of h	
$ap(1-\delta)$	-0.28761	0.382	0.750054	0.393545	1.429536
$sg(1-\delta)$	-0.37450	0.022	0.687633	0.499534	0.946561
$sz1(1-\delta)$	0.54108	0.310	1.717861	0.604375	4.882855
$sz2(1-\delta)$	0.37746	0.499	1.458575	0.487863	4.360732
$rx1(1-\delta)$	-0.33504	0.098	0.715309	0.480864	1.064058
$rx2(1-\delta)$	-0.31386	0.119	0.730621	0.492653	1.083525
$rx3(1-\delta)$	-0.47695	0.021	0.620674	0.413507	0.931639
$age1(1-\delta)$	-1.38196	0.000	0.251086	0.124838	0.505009
$age2(1-\delta)$	-0.39933	0.030	0.670769	0.467456	0.962511
$bm(1-\delta)$	0.12891	0.629	1.137588	0.673889	1.920355
$hg1(1-\delta)$	0.05701	0.886	1.058666	0.485911	2.306566
$hg2(1-\delta)$	0.07356	0.675	1.076333	0.763097	1.518146
$pf(1-\delta)$	0.53188	0.043	1.702129	1.016525	2.850174
$ap\delta$	0.96426	0.000	2.622846	1.691896	4.066082
$sg\delta$	1.45980	0.000	4.305098	2.402116	7.715567
$sz1\delta$	-1.41514	0.000	0.242892	0.127871	0.461368
$sz2\delta$	-0.79709	0.013	0.450638	0.240093	0.845819
$rx1\delta$	0.83317	0.004	2.300600	1.307321	4.048555
$rx2\delta$	0.97965	0.001	2.663524	1.528918	4.640073
$rx3\delta$	-0.05754	0.849	0.944084	0.521341	1.709635
$age1\delta$	0.67109	0.001	1.956369	1.309218	2.923437
$age2\delta$	0.03837	0.885	1.039116	0.619105	1.744067
$bm\delta$	0.44264	0.050	1.556812	0.999370	2.425166
$hg1\delta$	0.74674	0.049	2.110110	1.003596	4.436652
$hg2\delta$	-0.06019	0.823	0.941586	0.554887	1.597772
$pf\delta$	0.48878	0.049	1.630326	1.001982	2.652705

(a) Effect of treatments

For death by prostate cancer, $rx1\delta$ is highly significant (p value; 0.004). The hazard ratio 2.3006 indicates that the hazard of death by prostate cancer for $rx1\delta$ (placebo), is approximately 2.3 times higher than the hazard of death by using 5.0 mg of estrogen (4th category). The lowest level of estrogen, $rx2\delta$, is also highly significant and has a hazard, 2.66 times the hazard of 4th treatment category, indicating that the usage of 0.2mg of estrogen increases the risk of death by prostate cancer compared to 5.0mg of estrogen. $rx3\delta$ is not significantly different from the highest level. When considering all other causes of death, the covariate $rx1(1-\delta)$ is not significant. (p value: 0.098). This indicates that the placebo group is not significantly different from 5.0mg of estrogen. Only $rx3(1-\delta)$ is significant when the other causes of death are considered. Indication of this is that the treatment of $rx3(1-\delta)$; 1.0mg of estrogen, is significantly different from the highest level of the treatment, which is 5.0mg of estrogen. The hazard ratio 0.620674, indicate that the hazard is approximately 38% less for $rx3(1-\delta)$ when compared to 4th level of the treatment.

(b) Effect of other prognostic factors

For death by prostate cancer, the hazard ratio for $ap\delta$ is 2.6228. This hazard ratio indicates that, compared to lower level of ap ; Serum prostatic acid phosphatase, the higher level has a 2.62 times hazard of death by prostate cancer. Interpreting the other prognostic factors similarly it is seen that compared to 1st level of sg : lower combined index of stage and histological grade, 2nd level of sg : higher index, has a 4.3051 times hazard of death by prostate cancer. Hazard of death by prostate cancer for $sz1\delta$, is lower by approximately 76% than 3rd level of sz . Hazard is 55% less for $sz2\delta$, compared to the highest category of size of primary tumor. The $age1$ category, that is the youngest age group, has nearly 2 times more hazard of death by prostate cancer than the oldest age group which consists of patients over 70 years. When considering the Serum hemoglobin level ($hg\delta$), the hazard of the patients of $hg\delta = 1$ (lowest level) is approximately twice the hazard of patients having the highest level hemoglobin ($15 < hg \leq 25$). $pf\delta$ with a hazard ratio of 1.6303 indicates a higher hazard for patients with limited activity, when compared to normal patients. However, the significance of $hg\delta$ and $pf\delta$ are only marginal at a 5% level of significance.

For death by other causes $sg(1-\delta)$ is significant indicating that compared to patients with lower combined index of stage and histological grade (5-9), the patients with higher index have a lower hazard by an amount of 31%. Another significant

prognostic variable is $age1(1-\delta)$. For the other causes of death, $age1(1-\delta)$: the youngest age group, has approximately 75% less hazard than the oldest age group. Also $age2(1-\delta)$ renders a significant coefficient in the final model, indicating that the hazard of age group 2 is 33% lower than the oldest patients, depicted by a hazard ratio of 0.6798. $pf(1-\delta)$ is marginally significant, with a hazard ratio of 1.7021. Compared to patients of level 1: normal patients, $pf(1-\delta)=2$: patients with limitation of activity has approximately 1.7 times hazard of death by other causes. The confidence interval does not include 1, suggesting that the hazards are significantly different.

CONCLUSIONS AND RECOMMENDATIONS

In this study, the descriptive and univariate analysis were done using SAS and R, and Stata was used for the modelling. New programmes were developed in SAS for conducting the Lunn's test (Lunn, 1998). The goodness of fit test proposed by Sun *et al.* (2006) was modified to suit our example. This too required the development of new SAS programmes. The study was successful in illustrating the use of available methodology for modelling and wherever there were gaps in the methodology suitable developments were made so as to give the reader a complete picture of analyzing competing risk data from a clinical trial. The conclusions obtained from the prostate cancer data are as follows.

The patients have the highest risk of experiencing death by prostate cancer over all other causes of death (competing risks). This result was as anticipated because the data consisted of prostate cancer patients. The next highest risk was for patients with heart or vascular problems.

It can be concluded that the instantaneous rate of death for prostate cancer, has been lowered by treatment of estrogen. Thus, it can be concluded that the survival experience of patients is improved by the estrogen treatment, however there is no significant difference between levels 1.0mg and 5.0mg of estrogen treatment. When the other causes of death are considered, the hazard was approximately 34% less for 1.0mg of estrogen when compared to highest level of the treatment. Thus an estrogen level of 1.0 mg can be recommended for these patients.

When death by prostate cancer is considered; the hazard is higher for higher levels of Serum prostatic acid phosphatase (ap), size of primary tumor (sz), Combined

index of stage and histological grade (*sg*) and Performance rating (*pf*) while the hazard is lower for higher values of age and Serum hemoglobin level (*hg*).

The effect of the prognostic factors for other causes of death suggested that the effects of Combined index of stage and histological grade (*sg*) and Performance rating (*pf*) and age are significant. The implications of the results are that the patients with higher stage and histological grade and younger ages have lower hazard while the patients with limited activity has higher hazard compared to their respective baseline hazards. The result for younger age groups is sensible, as generally older patients are expected to have higher risk of death from any disease. For death by prostate cancer, the hazard of the youngest group of patients was almost twice as that of oldest patients. This is quite contrary to the conventional expectation that younger patients have much higher chance of escaping death from cancers. Recent research (Carter and Coffey, 1990) however has given some clues to support our findings on age with several studies showing worse prognosis of prostate cancer death for younger patients.

Model validation of Cox PH models mainly involves validating the assumptions. A goodness-of-fit test was performed to assess the inclusion of each and every variable in the multiplicative part of the Cox proportional hazard model, than in the additive part of a more general class of additive multiplicative models. The approach used suggested the goodness-of-fit of the model. That is the model is adequate.

REFERENCES

- Aalen, O.O. (1980) A model for nonparametric regression analysis of counting process. Springer Lect. Notes Statist. 2: 1-25.
- Ali, M. and A. Babiker, A. (2002) Applying the Cox proportional hazard regression model to competing risks, Stata Users Group series United Kingdom Stata Users' Group Meetings 2002, Number 11.
- Andrews, D.F. and A.M. Herzberg (1985). *Data: A Collection of Problems from Many Fields for the Student Research Worker*, Springer-Verlag, New York.
- Byar, D. and S. Green (1980). The choice of treatment for prostate cancer patients based on covariate information. *Bulletin Cancer*, 67: 477-488.
- Carter H. and D. Coffey (1990). The prostate: an increasing medical problem. *Prostate*, 16(1): 39-48.
- Cox, D. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 24: 187-220.

- Fine, J. and R. Gray (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94: 496-509.
- Gail, M. (1975, March). A review and critique of some models used in competing risk analysis. *Biometrics*, 31: 209-222.
- Gelfand, A., S. Ghosh, C. Christiansen, S. Soumerai and T. McLaughlin (2000). Proportional Hazards Models: A Latent Risks Approach. *Applied Statistics*, 49: 385-397.
- Kalbfleisch, J. and R. Prentice (1980). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons Inc : New York.
- Kay, R. (1986). Treatment effects in competing risks analysis of prostate cancer data. *Biometrics*, 42: 203-211.
- Ketchandji, M., Y. Kuo, V. Shahinian, J. Goodwin (2009). Cause of death in older men after the diagnosis of prostate cancer. *The Journal of the American Geriatrics Society*, 57(1): 24-30.
- Kuk, A. (1992). A mixture model for the regression analysis of competing risks data. *Australian Journal of Statistics*, 34: 169-180.
- Larson, M. and G. Dinse (1985). A Mixture Model for the Regression Analysis of Competing Risks Data. *Applied Statistics*, 34: 201-211.
- Lin, D. and Z. Ying (1995). Semiparametric Analysis of General Additive-Multiplicative Hazard Models for Counting Processes. *The Annals of Statistics*, 23: 1712-1734.
- Lunn, M. (1998, December). Applying k-sample tests to Conditional Probabilities for Competing risks in a Clinical Trial. *Biometrics*, 1662-1672.
- Lunn, M. and D. McNeil (1995). Applying Cox Regression to Competing Risks. *Biometrics*, 51: 524-532.
- Martinussen, T. and T. Scheike (2002). A flexible additive multiplicative hazard model. *Biometrika*, 89: 283-298.
- Pepe, M. and M. Mori (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, 12: 737-751.
- Pintilie, M. (2006). *Competing Risks: A Practical Perspective*. John Wiley & Sons Ltd.
- Sun, L., J. Liu, J. Sun, and M. Zhang (2006). Modelling the subdistribution of a competing risk. *Statistica Sinica*, 16: 1367-1385.