ISSN: 2577 - 8005
Review Article
Medical & Clinical Research

Survival Analysis of Multiple Myeloma Cancer (MMC) Using the Cox-Proportional Hazard Model

Lohuwa Mamudu[1*], Chris P Tsokos[1,2] and Otunuga Oluwaseun E[3]

[1]Doctoral Candidate in Statistics, Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA

[1,2]Distinguish University Professor, Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA

[3]Doctoral Candidate in Statistics, Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA

[*]Corresponding author
Lohuwa Mmamudu, Doctoral Candidate in Statistics, Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA
Email: lohuwa@usf.edu, Tel: 423-617-2135

Submitted:21 July 2020; **Accepted**: 27 July 2020; **Published**: 05 Aug 2020

Abstract

Though multiple myeloma cancer (MMC) remains incurable, research into improving the therapeutic strategy has increased dramatically in recent years. But it is unclear if sustained improvements have been achieved. We studied the survival times of 48 patients diagnosed and treated with alkylating agents. The semi-parametric Cox proportional hazard model was employed to examine the survival probability taking into account the sixteen risk factors presumed to be contributing to the survival times. A careful and rigorous assessment of the risk factors based on the AIC of the stepwise selection technique revealed seven risk factors, and one interaction term are statistically significantly contributing to the survival times. They are blood urea nitrogen (BUN)/serum creatinine, white blood cells (WBC), Bence Jone protein in the urine (BJPU), fractures, proteinuria, gender, platelets, and the interaction of infections and serum calcium. The final Cox-PH model was well-validated and satisfied the key assumptions. The identified risk factors are rank according to the prognostic effect on the survival time based on the hazard ratio. Blood urea nitrogen (BUN)/serum creatinine was the greatest prognostic factor (most contributing factor, and highly negatively related to the MMC deaths or survival times), followed by white blood cells (WBC), and normal platelet was found to be the minimum prognostic factor (least contributing factor to MMC death or survival times). This study offers prognostic and therapeutic significance for further enhancement in the treatment strategy of the multiple myeloma cancer disease.

Introduction

Multiple myeloma cancer (MMC) is a type of cancerous disease that remains incurable. The development of MMC (plasma cell myeloma) starts from the malignant plasma cell, specifically the white blood cell [1]. Myeloma occurs when the plasma cell becomes an anomalous caused by the inability of the plasma cells to fight against hazardous substances like germs. The name multiple myeloma is the consequence of the accumulation of the abnormal plasma cells in the bone marrow crowding the active blood cells and destroying the solid part of the bone [2-8]. The growth of the myeloma cell is given by Figure 1 below [4, 5].

The abnormal plasma cells produce abnormal antibodies that cause kidney problems and overly thick blood [9-11]. The initially identified causes of MMC are obesity, radiation exposure, family history, and certain chemicals [12-15]. Generally, no specific risk factors or causes have been identified for patients diagnosed with MMC. The recent treatment of MMC is based on high-dose chemotherapy, commonly with bortezomib-based regimens, and lenalidomide-dexamethasone followed by autologous hematopoietic stem-cell transplantation (ASCT). For MMC patients under 65 years old, there are suggestions of transplantation of a persons' stem cells [16]. There has been an improved

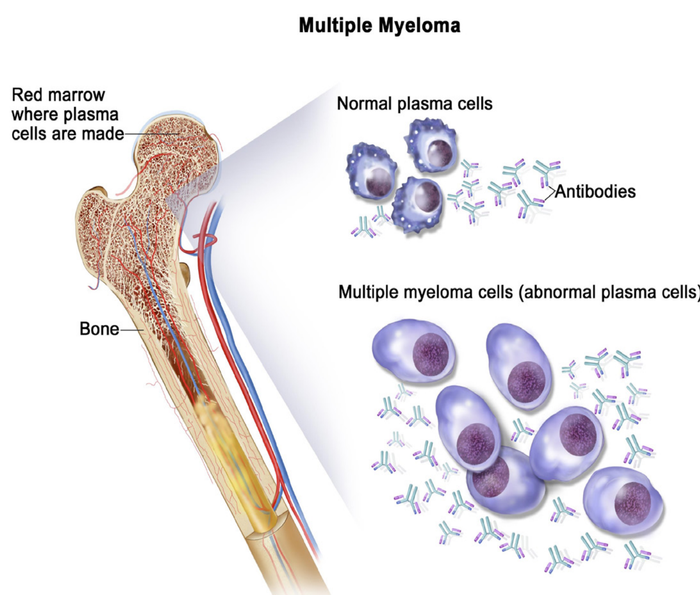

Figure 1: Growth of the Myeloma Cell

progression-free survival and overall survival for treatment based on post-ASCT maintenance therapy with lenalidomide for a patient at normal risk [17]. For intermediate and high-risk patients, the bortezomib-based maintenance regimen has been reported suitable according to a clinical trial conducted in 2012 [18].

There have been some intriguing statistics about multiple myeloma cancer in the recent times. 30,000 new patients are reported to be diagnosed with MMC in the United States (U.S.) every year, the second most common hematologic malignancy in the U.S. [19-21]. MMC is ranked among the 14 top cancer diseases in the U.S., a report by the Surveillance, Epidemiology, and End Results (SEER) Cancer Institute in 2019 [2]. SEER reported an estimated new cases of 32,110 MMC patients, an increase compared with the 24,050 reported in 2014 [3, 6]. Most risk factors of MMC are reported to be common among the age, males, black race, and families with MMC history [2, 7].

There are no major findings of what specifically causes MMC, given that the disease remains incurable. However, research has discovered several risk factors presumed to have some relation with the duration of the survival of patients with MMC [9, 10]. Most of these factors were identified at the time a patient was diagnosed with the MMC disease. Some common risk factors identified through clinical trials and research studies included hemoglobin, immunoglobulin type, extent and type of lesions, serum calcium, age, sex, white blood cells, blood urea nitrogen, serum calcium, serum albumin, infections, platelets, hemoglobin, presence of Bence Jones protein, and performance status, all at the time patients were diagnostic with MMC, are believed to have contributed to the survival of patients with MMC [10, 19, 20].

A struggle to find a lasting solution or treatment to the incurable MMC has resulted in research into some statistical analysis on the survival of patients with MMC given the event that a patient died or survived. Kaplan-Meier technique has been commonly used for analyzing cancer survivorship data in recent times due to the simplicity of its usage. It is often used to compare the survival difference of observations/groups base on the log-rank test. KM is mostly used for longitudinal studies like a cohort study; an example to the present study (i.e. the survival time of patients diagnosed with multiple myeloma) [22-24]. The disadvantage of using KM is that it does not take into consideration the risk factors (covariate) contributing to the length of patients' survival duration of the MMC disease, hence, nullifying the relevance of KM if risk factors are contributing in the given survival data.

Brain et al [9] used Kaplan-Meier to test whether there was a significant difference in the survival duration between the categories of risk factors based on the generalized Wilcoxon test and the log-rank test. They further used a non-linear Cox regression to ascertain the combination of patients' characteristics relative to survival duration. They identified a significant difference in the survival duration among patients based on performance status, cell mass and percentage labelling index, Nephrotic status, Hemoglobin, age, and k/λ subtype. John M. Krall et al developed a set-up procedure for selecting variables associated with the survival times of patients with MMC utilizing the data used in the present study [22]. They found blood urea nitrogen, hemoglobin, percent plasma cell in bone marrow, and Serum calcium to be associated with the survival of patients with multiple myeloma. Shaji K. Kumar et al found continued improvement in survival in MMC with changes in early mortality and outcomes in older patients. Giampaolo Merlini, Jan G. Waldenstrom, and Suresh D. Jayakar [26] proposed a new improved clinical staging system for the survival of MMC based on analysis of 123 treated patients. In their findings, serum calcium, % bone myeloma plasma cell (% BMPC) and serum creatinine/BUN were significantly related to the survival of IgG myeloma stage; hemoglobin, serum calcium, and M-component related significantly with the survival of IgA myeloma stage; and creatinine/BUN, % BMPC and serum calcium to be related significantly with the survival of BJ myeloma stage; but no significant relation to survival with age or sex. Our study found five new significant attributable out of the sixteen risk factors presumed to be contributing to the survival times of MMC. They are platelets, gender, white blood cells, fractures, and an interaction term between infections and serum calcium. In most of the research studies, either one or two of the five newly identified risk factors were analyzed, but not found significant or not part of the data analyses.

We studied the semi-parametric Cox-PH survival analysis of the survival times to estimate the survival rate of patients diagnosed with multiple myeloma. We utilized the Cox-PH model to analyze the proportion of survival time, taking into account the 16 risk factors, Table 1, considered to be contributing to the survival time of the patients diagnosed with MMC. Thus, we assessed the relationship between the proportion of survival time as a function of 16 attributable risk factors and two-way interactions based on the Cox proportional hazard (PH) model. The significant attributable risk factors identified were carefully investigated and selected based on the stepwise model selection method, with the final model representing the model with the least AIC. The final Cox-PH model was validated to satisfy all the key assumptions, and no presence of multicollinearity measured based on the variance ination factor (VIF).

Data Description

The data was provided by Harley from West Virginia University Medical Center [22, 23]. The original data consist of the survival times of 72 multiple myeloma cancer (MMC) patients diagnosed and treated with alkylating agents [22]. Out of the 72 patients, we have 65 complete data of patients on 16 concomitant variables (risk factor). The remaining 7 patients' information were discarded due to missing data in at least one of the 16 risk factors. The data collection process involved recording the 16 risk factors of the patient diagnosed with MMC and monitoring how long the patient survived the MMC disease (called the survival time from diagnosis to the nearest month). Given the 65 patients with complete data, 48 were dead, and the remaining 17 were alive. Therefore, our analysis involves 48 patients whose survival times are known. The response variable is the survival times of the patient, which is continuous. Given the 16 risk factors, 11 are continuous, and 5 are categorical. Table 1 below gives the complete description of the response variable and the 16 risk factors (contributing variables).

Table 1: Variables Recorded for Multiple Myeloma Patients

Symbol	Variable Name
t	Survival time from diagnosis to nearest month +1
X_1	Log blood urea nitrogen (BUN)/serum creatinine at diagnosis
X_2	Hemoglobin at diagnosis
X_3	Platelets at diagnosis 0 abnormal, 1 normal
X_4	Infections at diagnosis 0 none, 1 present
X_5	Age at diagnosis (complete years)
X_6	Gender 1 male, 2 female
X_7	Log white blood cell (WBC) at diagnosis
X_8	Fractures at diagnosis 0 none, 1 present
X_9	Log %BM at diagnosis (log % plasma cells in bone marrow)
X_{10}	% Lymphocytes in peripheral blood at diagnosis
X_{11}	% Myeloid cells in peripheral blood at diagnosis
X_{12}	Proteinuria at diagnosis
X_{13}	Bence Jone protein in urine at diagnosis 1 present, 2 none
X_{14}	Total serum protein at diagnosis
X_{15}	Serum globin (gm%) at diagnosis
X_{16}	Serum calcium (mgm%) at diagnosis

Our analysis of the survival times of the 48 MMC patients started by first comparing whether there is a difference in the survival times between the males and the females using the Kruskal-Wallis rank-sum test given in Table 2. The test resulted in a large p - value = 0.5224, thus, we do not reject the null hypothesis (i.e. H0:$\mu_M = \mu_F$). Hence, there is strong statistical evidence that the survival times of the males and the females are not different. Therefore, we utilized the combined data of both males and females for Cox-PH analysis of the survival times of the MMC patients.

Table 2: Kruskal-Wallis rank sum test of the Difference in Survival probability Between 3p-lognormal and Kaplan Meier.

Type of Test	Survival Probability	Data: list[Male, Female]
Kruskal-Wallis	chi - squared($\tilde{\chi}^2$) = 0.40914	p - value = 0.5224

Review of the Cox Proportional Hazard Model

In survival analysis, two things are of utmost importance; time and event. Thus, survival analysis models the time an event occurred called the survival time. For example, the time a patient died of MMC. The survival time can be associated/influenced by one or several attributable factors/risks, often termed as covariates by most survival analysis literature. Cox proportional hazard model is also known as the Cox model, introduced by Cox (1972) has been widely recommended for semi-parametric modelling of the relationship of the survival time as a function of the covariates in survival analysis. A good basic review of the introduction and methodology is given by Kleinbaum, and more extensive discussions have been provided by Kalbeisch and Prentice [25-31]. We are given a brief review of the Cox proportional hazards model in this section. An important aspect of the Cox model is the hazard function. The hazard function measures the rate of death at time t. We define the hazard function as follows; Let random variable T denote the survival time with cumulative density function $F_T(t)$, given by

$$F_T(t)=P(T\leq t)=\int_0^t f_T(t)dt$$

Thus, $F_T(t)$ is the probability of failure by time t and $f_T(t) = dF_T(t)/dt$ is the probability density function. The survival function is defined as

$$S_T(t) = P(T > t) = 1 - P(T \leq t) = 1 - F_T(t).$$

Therefore, the hazard function which examines the risk of instantaneous death at time t, is conditional on the survival function defined by

$$\begin{aligned} h(t) &= \lim_{\partial t \to 0} \frac{F_T(T+\partial t) + F_T(t)}{\partial t . S_T(t)} \\ &= \lim_{\partial t \to \infty} \frac{P(t < T \leq t + \partial t)}{\partial t . S_T(t)} \\ &= \lim_{\partial t \to \infty} \frac{P(t < T \leq t + \partial t | T > t)}{\partial t} \\ &= \frac{f_T(t)}{S_T(t)} \end{aligned} \quad (1)$$

From the hazard function given by equation (1), we can obtain the cumulative hazard function, expressed as

$$H(t) = \int_0^t h(s)ds.$$

The integral can be expressed in close form as $H(t) = - \ln S(t) = - \ln R(t)$.

The Cox model which includes interacting covariates is expressed by the hazard function, estimated as follow:

$$h_i(t) = \alpha_i(t) \exp\left(\sum_{i=1}^{k} \beta_i X_i + \sum_{i \neq j=1}^{k} \rho_{ij} X_i X_j\right)$$

and **(2)**

$$In\left(\frac{h_i(t)}{\alpha_i(t)}\right) = \sum_{i=1}^{k} \beta_i X_i + \sum_{i\neq j=1}^{k} \rho_{ij} X_i X_j$$

where t is the survival time, $h_i(t)$ is hazard function obtained by the set of k covariates, β_i is the coefficients measuring the impact of the covariates X_i on $h_i(t)$,ρ_{ij} is the coefficient measuring the impact of interacting covariates $X_i X_j$ on $h_i(t)$, $\alpha(t)$ is the baseline value of $h_i(t)$ if all X_i and $X_i X_j$ equals zero. The Cox model is a multiple linear regression of the logarithmic form of the hazard on X_i's and $X_i X_j$'s, with $\alpha(t)$ as an intercept that varies with time t. A major assumption of the Cox model is the proportional hazard assumption, which explains that the hazard function of observations (or patients) should be proportional and independent of time t [29]. Consider the case of two patients i and i' with varying values of covariates; the corresponding hazard functions for i^{th} patient is

$$\eta_i(t) = \alpha(t)\exp\left(\sum_{i=1}^{k}\beta_i X_i + \sum_{i\neq j=1}^{k}\rho_{ij}X_iX_j\right)$$

and the corresponding hazard functions for i'^{th} is

$$\eta_i'(t) = \alpha(t)\exp\left(\sum_{i'=1}^{k}\beta_{i'} X_{i'} + \sum_{i'\neq j'=1}^{k}\rho_{i'j'}X_{i'}X_{j'}\right)$$

The hazard ratio of the two patients is

$$\frac{\eta_i(t)}{\eta'_i(t)} = \frac{\alpha(t)\exp\left(\sum_{i=1}^{k}\beta_i X_i + \sum_{i\neq j=1}^{k}\rho_{ij}X_iX_j\right)}{\alpha(t)\exp\left(\sum_{i'=1}^{k}\beta_{i'} X_{i'} + \sum_{i'\neq j'=1}^{k}\rho_{i'j'}X_{i'}X_{j'}\right)} \quad (3)$$

$$= \frac{\exp\left(\sum_{i=1}^{k}\beta_i X_i + \sum_{i\neq j=1}^{k}\rho_{ij}X_iX_j\right)}{\exp\left(\sum_{i'=1}^{k}\beta_{i'} X_{i'} + \sum_{i'\neq j'=1}^{k}\rho_{i'j'}X_{i'}X_{j'}\right)} = \exp(coef),$$

Which is independent of time t. The consequence of the above hazard ratio implies that the Cox model is a proportional-hazards model. The hazard ratio is a relative measure of the hazards between observations/groups [28]. We interpret the hazard ratio (HR) in the following three ways: (1) $HR = 1$; implies that there is no hazard effect. Thus, the covariates have no relationship with the event probability, hence, no influence on the length of survival. (2) $HR > 1$ (i.e. equivalently $\hat{\beta}_i > 0$), implies an increase in hazard. That is, the covariates have a positive association with the event probability, hence, a negative association with the length of survival (bad prognostic factor). (3) HR < 1 (i.e. equivalently $\hat{\beta}_i < 0$), implies a decrease in hazard. That is, the covariates are negatively associated with the probability of the event, hence, positively associated with the length of survival (good prognostic factor). A comprehensive review of the hazard ratio have been provided by L. Douglas Case et al [32].

To compute the baseline hazard function, we performed the following computation:

$$\hat{\alpha}(t) = \sum_{t_i \leq t}\hat{h}(t_i)$$

with

$$\hat{h}(t_i) = \frac{d_i}{\sum_{i\in R(t_i)}\exp(X_i'\hat{\beta})}$$

Where $t_1 < t_2 < \ldots < t_n$ denote the distinct event times, d_i is the number of events at t_i, and $R(t_i)$ is the risk set at t_i containing all individuals still susceptible to the event at t_i. The base line hazard function can assume any functional form of the covariates. In section 1.3.1, we discussed in detail the major assumptions of the Cox-PH model. We will show that the assumptions are satisfied once we have developed the Cox-PH model for the given data.

Cox-Proportional Hazards (PH) Model Assumptions

A good Cox proportional hazard model should satisfy the following three key assumptions, prior to its implementation. Failure to satisfy the assumptions will lead to wrong decision about the subject matter.

Proportional hazard (PH) assumption

The PH assumption of the Cox model can be assessed based on formal statistical tests. A non-statistical significance of the covariates and the global test is an indication that the PH assumption is valid. Another method to check for the PH assumption is by investigating the plot of scaled Schoenfeld residuals against the transformed time. The Schoenfeld residuals are independent of time; a non-random pattern against time is evidence of a violation of the PH assumption. We calculate the Schoenfeld residuals with one per observation per covariate. This can be expressed as

$$r_{ik} = X_{ik} - \hat{\bar{X}}_{wik}(\beta, t_i)$$

Where X_{ik} denotes the value of the k^{th} covariate for i^{th} observation. $\hat{X}_{wik}(\beta, t_i)$ represents the weighted mean values of covariates at risk at the given event time, t_i, denoted by $R(t_i)$, and given by

$$\hat{\bar{X}}_{wik}(\beta, ti) = \sum_{j\in R(t_i)} X_{ik}w_i(\beta, t_i)$$

The weight function, $w_i(\beta; t_i)$ for i^{th} observation at risk, $R(t_i)$ is the probability that observation i fails at time t_i, defined by

$$w_i(\beta, t_i) = \frac{\exp(\beta^T X_i)}{\sum_{I\in R(t_j)}\exp(\beta^T X_i)}$$

A positive value of r_{ik} depicts an X value higher than expected at that death time. For a binary (0, 1) variable, Schoenfeld residuals will be between -1 and

In that situation,

$$r_{ik} = \begin{cases} 0 - \hat{\bar{X}}_{w_i k} & \text{for X} = 0 \\ 1 - \hat{\bar{X}}_{w_i k} & \text{for X} = 1 \end{cases}$$

Linear functional form of continuous covariates

We assume that the functional form of the covariates are linear. T. Therneau and P. Grambsch suggested this assumption could be checked by visualizing the plot of Martingale residuals against the continuous covariates with fitted lowess (locally weighted smoothing) line function. A trend or pattern in the plot is evidence of a violation of the linear functional form of the covariates. Martingale residual is defined by

$$\hat{M}_i = \delta_i - \hat{\Gamma}_0(t_i)\exp(\hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}),$$

Where δ_i denotes the event indicator for i^{th} observation, $\hat{\Gamma}_0(t_i)$ is the estimated cumulative hazard at the final follow-up time for i^{th} observation, and $\exp(\hat{\beta}_1 X_{1i} + \ldots)$ is the estimated coefficients applied to the observed covariate for the i^{th} observation. Martingale residuals, $\hat{M}_i$, have a skewed distribution. The $\hat{M}_i$ values are

$$\hat{M}_i = \begin{cases} 1, & \text{for maximum possible values} \\ -\infty, & \text{for minimum possible values} \end{cases}$$

A positive Martingale residual value implies individuals demised too soon, negative value implies individuals lived too long. A transformation of $\hat{M}_i$ to obtain approximate symmetric distribution can be essential. Such a transformation is motivated by deviance residuals defined below.

Examining influential observations (or outliers)

In examining influential observations, we visualized the dfbeta values. The dfbeta values estimates the influence of i^{th} - case (or observation) on the regression coefficients. A very high value of dfbeta should be closely investigated. Another technique for checking influential observations is by assessing the deviance residuals (symmetric/normalized transformation of the Martingale residuals) plot. The deviance residual is defined by

$$d_i = \sin(\hat{M}_i)\sqrt{2}\sqrt{-\hat{M}_i - \delta_i \log(\delta_i - \hat{M}_i)}.$$

Note that $d_i = 0$ is only when $\hat{M}_i = 0$. The square root shrinks the large negative martingale residuals, while the logarithm transformation expands those residuals that are close to zero (i.e as $n \to \infty$). The distribution of the residuals should be roughly symmetrical about zero mean and standard deviation of one. A very large/small/distant deviance residual values indicate influential observations or outliers. The values of the deviance residual values can be compared with the expected value of survival time. A positive value implies individuals demised too soon, negative value implies individuals lived too long.

Now, we proceed to develop the Cox-PH model for the survival times of multiple myeloma patients. After we develop the model, we will verify that the above three assumptions are met to validate the applicability and quality of the proposed model.

Proposed Cox-PH Model for the Survival Times of Patients with MMC

We started by fitting the Cox-PH model to the survival times t as a function of all 16 covariates X_i given in Table 1 together with their two-way interactions. A stepwise model selection method was adopted to select the final model with the least Akaike information criterion ($AIC = 2ln(L) + 2k$, where L is the value of the maximum likelihood function of the model and k represents the estimated model parameters) [25]. AIC gives an estimation of the relative amount of information missing in the model; hence, the smaller the AIC value the better the quality of the model. It deals with the danger posed by overfitting or under-fitting the model.

The stepwise model variable selection procedure is one of the best ways used for determining significant covariate for Cox-PH models. It is based on iterations between forwarding and backward steps. All covariates and their interactions are included to be part of the "variable list" for selection. The significance levels for entry (α_{entry}) and stay (α_{stay}) are suggested to be set at 0.15 or larger for being conservative. Then, the best Cox-PH model is obtained by manually removing the covariates with *p-value* > 0.05 one at a time until all model coefficients are statistically significant at the chosen level of significant, $\alpha = 0.05$. The final model with all significant covariates and possible interactions is the model with the least AIC value. Hence, based on the stepwise model selection procedure criteria, our final proposed model that significantly contributes to the probabilistic survival time of patients diagnosed with multiple myeloma includes seven significant contributable covariates (risk factors) and one interaction; given by

$$\begin{aligned} \text{In}\left(\frac{h_i(t)}{\alpha_i(t)}\right) &= 2.008X_1 - 1.608X_3 normal - 0.815X_6 female \\ &+ 1.878X_7 + 0.854X_8 present + 0.108X_{12} \\ &+ 1.576X_{13} none + 0.114X_4 present.X_{16} \end{aligned} \quad \textbf{(4)}$$

The Table 3, below displays the estimates of the model coefficients/parameters, their hazard ratios (HR) (exp ($\hat{\beta}$)), standard error of coefficients, statistical significance, and 95% confidence interval. The significant contributing coefficients or risk factors have been ranked based on the prognostic effect to the survival times of patients diagnosed with MMC using the hazard ratio (HR). Thus, we ranked from the most contributing factor to the least contributing factor to the death or survival times of MMC patients. The covariates, X'_i s are defined in Table 1.

Table 3: Ranking of the Significant Contributing Covariates (Risk Factors) Base on Prognostic Effect to the Survival Time Using the Hazard Ratios

Rank	Covariates	coeff ($\hat{\beta}$)	HR (exp($\hat{\beta}$))	S.E.($\hat{\beta}$)	Pr(>|z|)	lower .95	upper .95
1	X_1	2.008	7.454	0.619	$1.165e^{-3**}$	2.217	25.056
2	X_7	1.878	6.543	0.773	$1.505e^{-2**}$	1.439	29.745
3	X_{13}	1.576	4.835	0.418	$1.63e^{-4***}$	2.131	10.972
4	X_8	0.854	2.349	0.409	$3.693e^{-2*}$	1.053	5.243
5	$X_4 : X_{16}$	0.113	1.121	0.040	$4.873e^{-3**}$	1.035	1.213
6	X_{12}	0.108	1.114	0.030	$3.84e^{-4***}$	1.049	1.183
7	X_6	-0.815	0.443	0.391	$3.711e^{-2*}$	0.206	0.952
8	X_3	-1.608	0.200	0.502	$1.355e^{-3**}$	0.075	0.536

Table 4: Global statistical significance of the model

Type of test	Test statistic	df	P - value
Likelihood ratio test	32.6	8	$7e^{-05}$
Wald test	30.38	8	$2e^{-04}$
Score (log-rank) test	32.49	8	$8e^{-05}$

Table 5: First Ten Baseline Hazard Estimates

Obs.	Baseline Hazard	Time
1	$2.498e^{-06}$	1.25
2	$1.463e^{-05}$	2.00
3	$1.951e^{-05}$	3.00
4	$3.162e^{-05}$	5.00
5	$6.104e^{-05}$	6.00
6	$8.818e^{-05}$	7.00
7	$9.806e^{-05}$	9.00
8	$1.549e^{-04}$	11.00
9	$1.676e^{-04}$	13.00
10	$1.818e^{-04}$	14.00

In Table 3, we tested the statistical significance of each of the chosen risk factors and interaction (coefficients) in the equation (4) based on the p - value from Wald statistic value. All the selected risk factors are tested significant, with “three stars ***” indicating a very highly statistically significant risk factor. A positive coefficient ($\hat{\beta} > 0$) means a higher hazard rate, and thus a bad prognostic factor. By contrast, a negative coefficient ($\hat{\beta} < 0$) means a lower hazard rate, and thus a good prognostic factor. For instance, $\hat{\beta}_6$ = -0.815 representing gender implies that females are good prognostic of the survival time of MMC; thus, females have a lower risk of death (higher survival rates) of MMC than males. The exp($\hat{\beta}$) is the hazard ratio measures the size of the effect of the risk factor. Thus, exp(-0.815) = 0.443 < 1 for gender means being a female has a reduced risk of dying with MMC than being a male. The ranking of the significant attributable risk factors based on the HR shows that blood urea nitrogen (BUN)/serum creatinine (X_1) is the greatest prognostic factor to the survival of MMC, followed by white blood cells (WBC) (X_7), and platelets (X_3) is the least prognostic factor. Table 4 displays three different tests for the overall significance of the proposed Cox model; the likelihood-ratio test, the Wald test, and score log-rank statistics. The three tests are asymptotically equivalent and give similar results for large samples. However, for small samples like in our case, the likelihood ratio test is robust and generally preferred. The global statistical significance test demonstrates that the proposed Cox-PH model in equation (4) is highly statistically significant. In Table 5, we displayed the baseline hazard function $\hat{\alpha}(t)$ for the first ten observations. Figure 2, below is a graphical display of the results given in Table 3 according to the order of prognostic effect based on the hazard rate. Clearly, we can see that blood urea nitrogen (X_1) is the greatest prognostic factor to the survival time of MMC patient, followed by white blood cells (X_7), and platelets (X_3) is the least prognostic factor.

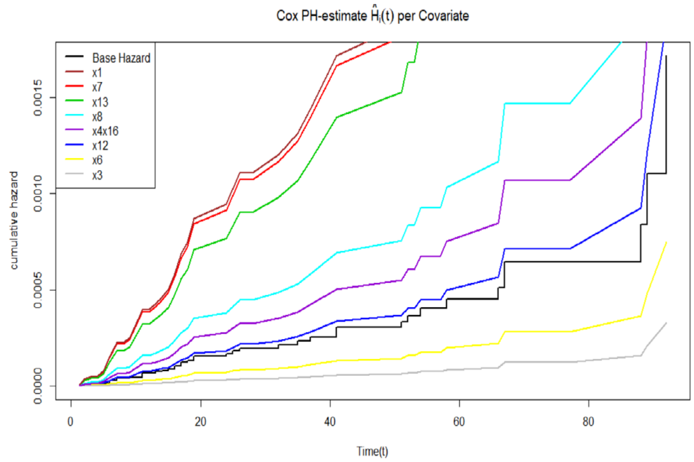

Figure 2: Ranking of Prognostic Eect of Risk Factors

Validation of the Proposed Cox-PH Model

We validated the goodness-of-fit of the proposed Cox-PH model by satisfying the three major Cox-PH model assumptions outlined in section 1.3.1. Firstly, we verified that the proportional hazard assumption is satisfied. Figure 3, shows the plot of the scaled Shoenfeld residual against time. It shows that there is no pattern as a function of time. Thus, the residuals are randomly scattered with no systematic departures from the horizontal fitted smoothing spline deep line (i.e. the residuals are independent of time). A formal test for the PH assumption is given in Table 6. The covariates and the global test are non-statistically significant given by the large p-values. This is a further justification of the validity of the PH assumption for our proposed model.

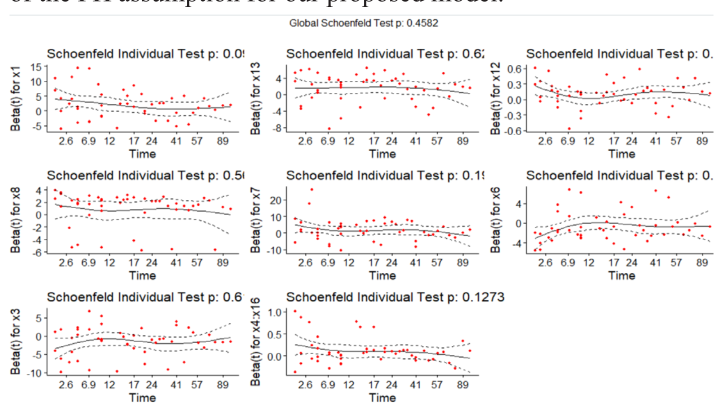

Figure 3: Testing Proportional Hazard Assumption

Table 6: Formal Test of Proportional Hazard Assumption

Covariate	ρ	X^2	*P - value*
X_1	-0.2136	2.8454	0.0916
X_3	0.0636	0.2491	0.6177
X_6	0.1391	1.1587	0.2817
X_7	-0.1569	1.6895	0.1937
X_8	-0.0861	0.3291	0.5662
X_{12}	-0.0323	0.0651	0.7986
X_{13}	-0.0607	0.2348	0.6280
$X_4 : X_{16}$	-0.2067	2.3255	0.1273
GLOBAL	**NA**	**7.7505**	**0.4582**

Secondly, we assessed the functional form of continuous covariates. The continuous covariates are expected to have a linear form. However, categorical covariates do not have any issue of nonlinearity. Figure 4 is a plot of Martingale residuals against continuous covariates with fitted lowess (locally weighted smoothing) function. The plot demonstrates no major trend or pattern. Thus, the linear functional form of continuous covariates is reasonable. Therefore, continuous covariates have a linear functional form. We further investigated the presence of influential observations (or outliers). In Figure 5, we plot the magnitude of *dfbeta* against the model coefficients. We can see that there are no major Influential observations, given that all the residuals are within one standard deviation of the residuals. Multicollinearity can negatively impact the precision of the estimated model coefficients and prediction. In Table 7, we employed the variance inflation factor (VIF) to assess multicollinearity. A V IF > 2.5 and V IF > 5 for categorical and continuous covariates, respectively, are evidence of the presence of multicollinearity. Given the VIFs in Table 7, implies that there is no multicollinearity in our proposed Cox PH model.

Table 7: Variance Inflation Factor of Model Coecient

Covariates	X_1	X_{13}	X_3	X_{12}	X_8	X_7	X_6	$X_4\ X_{16}$
VIF	1.564	1.602	1.283	1.612	1.656	1.574	1.385	1.243

The Proposed Cox-PH Model Survival Function

The survival function of the Cox-PH model is a reverse process of the hazard function in equation (1). In equation (1), we are given *f (t)* and $\hat{S}(t)$, then we proceed to find *h(t)*. In Cox-PH, we find $\hat{S}(t)$ given *f(t)* and *h(t)*. In addition to the relationship between *h(t)* and $\hat{S}(t)$ in equation (1), another alternate relation is given by

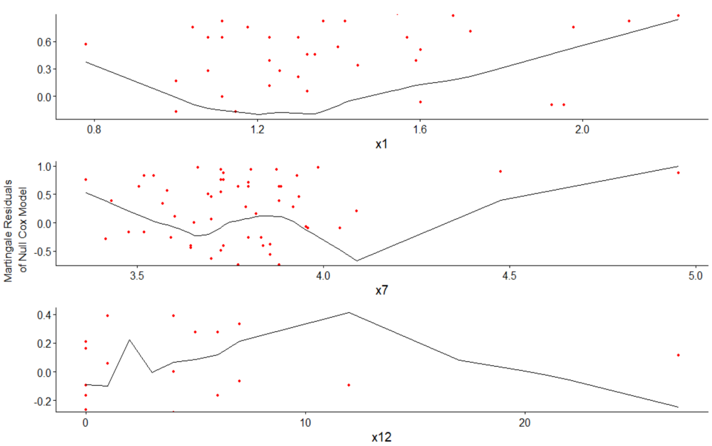

Figure 4: Assessing the Functional Form of the Continuous Covariates

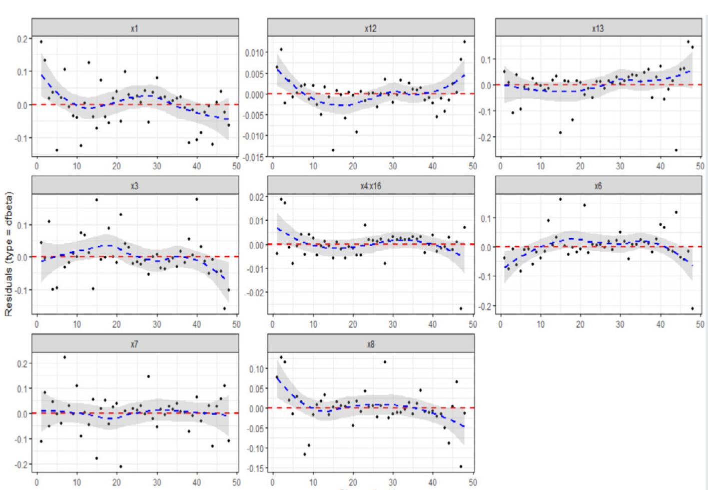

Figure 5: Assessing Influential Observations (or Outliers)

$$\hat{S}(t) = \exp\left(-\int_0^t h(t)dt\right) = \exp\big(-H(t)\big) \tag{5}$$

The Cox-PH model given by equation (2) can be re-written in the form

$$h(t_i; X_i, X_iX_j) = h_0(t)\exp\left(\sum_{i=1}^{k}\beta_i X_i + \sum_{i\neq j=1}^{k}\rho_{ij}X_iX_j\right) \tag{6}$$

We can modify the Cox-PH model for the survival function by employing equation (5) above. Therefore, the survival function of the Cox-PH model can be expressed as

$$\begin{aligned}
\hat{S}(t_i; X_i, X_iX_j) &= \exp\left(-\int_0^t h(t_i; X_i, X_iX_j)dt\right)\\
&= \exp\left(-\int_0^t h_0(t)\exp\left(\sum_{i=1}^{k}\beta_i X_i + \sum_{i\neq j=1}^{k}\rho_{ij}X_iX_j\right)dt\right)\\
&= \exp\left(-\exp\left(\sum_{i=1}^{k}\beta_i X_i + \sum_{i\neq j=1}^{k}\rho_{ij}X_iX_j\right)\int_0^t h_0(t)\,dt\right) \qquad (7)\\
&= \exp\left(-\int_0^t h_0(t)\,dt\right)^{\exp\left(\sum_{i=1}^{k}\beta_i X_i + \sum_{i\neq j=1}^{k}\rho_{ij}X_iX_j\right)}\\
&= S_0(t)^{\exp\left(\sum_{i=1}^{k}\beta_i X_i + \sum_{i\neq j=1}^{k}\rho_{ij}X_iX_j\right)}
\end{aligned}$$

where $h_0(t)$ is baseline hazard function, which assumes any functional form of the covariates. The coefficient parameters of the covariates has been estimated, given by Table 3.

In section 1.3.3, we validated the quality of the proposed Cox-PH model in equation (4) by showing that it satisfied all the three model assumptions outlined in section 1.3.1. Given that the model is of high quality, we estimated the Cox-PH survival function (i.e. the proportions of survival beyond time t) of patients diagnosed with multiple myeloma cancer (MMC) as a function of the seven covariates (risk factors) and the interaction term, given by equation (7). Figure 6 shows the proposed Cox proportional hazard model survival function of the survival time. This plot demonstrates the predicted survival proportion at any given point in time for the mean values of the risk factors. The cox model is very useful in predicting the probability of the survival time for an individual patient based on the significant attributable risk factors that we have identified. Thus, given that a patient is diagnosed with MMC, we put into the model the seven contributing risk factors and the interacting factor to estimate the probability of survival beyond a given survival time (death time).

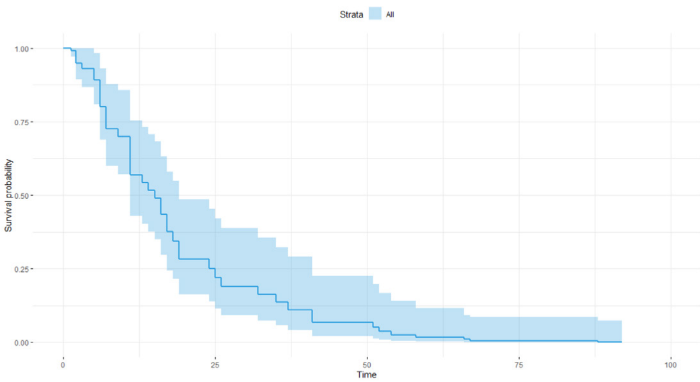

Figure 6: Survival Estimate $\hat{S}(t)$ from the Proposed Cox-PH Model

Discussion

The multiple myeloma cancer (MMC) cancer diseases may be incurable. However, the introduction of therapeutic agents such as thalidomide, lenalidomide, bortezomib, and high-dose chemotherapy and stem-cell rescue (ASCT) has improved the treatment progress, hence the survival time of the patient. Also, many research techniques and approaches have been adopted to enhance the patients' survival time after been diagnosed with MMC.

In the present study, we first assessed whether there is a difference between the survival times of the MMC males and females given by the Kruskal-Wallis rank-sum test in Table 2. It was revealed that there is no difference in the survival times of the males and the females. Therefore, we performed the Cox-PH model analysis of the survival times without considering stratification of the data. We then estimated the proportion of the survival time as a function of covariates utilizing the Cox proportional hazard regression model.

Our data on the survival times of patients diagnosed with MMC included 16 risk factors presumed to be contributing to survival times. We used the Cox proportional hazard model to estimate the proportion of survival time because it takes into consideration the risk factors (covariates) contributing to the patients' survival time. Therefore, we developed the Cox-PH model for the survival time of patients diagnosed with MMC base on the sixteen risk factors. Our final proposed Cox-PH model given by equation (4) identified seven significant risk factors and one interaction term as contributing to the survival probability. They are blood urea nitrogen (BUN)/serum creatinine, white blood cells (WBC), Bence Jone protein in the urine (BJPU), fractures, proteinuria, gender, platelets, and the interaction infections and serum calcium. It is interesting and highly important to point out that the two interacting risk factors do not individually significantly contribute to the survival probability, a highly useful finding. Seldom do we see interaction terms in Cox-PH models because they are difficult to find. However, not including interaction(s) in the Cox-PH model given that they exist and affect the survival time of the patient can result to the wrong estimation of the proportion of the survival time, hence diminishing the true relevance and quality of the Cox-PH model, and consequently endangering the treatment process of the MM cancer disease.

The proposed Cox-PH model satisfied all the key assumptions of the Cox-PH model discussed in section (1.3.1). Most research uses the Cox-PH models without discussing or validating the assumptions that allowed for its usage. Therefore, we cannot justify the quality of the Cox-PH model they proposed in their findings. Our proposed Cox-PH model is of high quality because the underlying key assumptions are satisfied and well-validated. The proposed Cox-PH model has the least AIC based on the stepwise model selection procedure with uncorrelated covariates given by the very small VIF values in Table 7. We rank the identified significant attributable risk factors or covariates based on the prognostic effect (i.e. the highest contributing factor to MMC deaths to the least contributing factor to MMC deaths) on the survival time utilizing the hazard ratio. We found all the identified risk factors, except the female gender and normal platelet to be highly prognostic factors (negatively associated with the survival time).

We found five new risk factors contributing to the survival of patients with MMC. They include the individual risk factors platelets, gender, white blood cells and fractures; and the interaction term, infections and serum calcium. These newly identified risk factors are not found in the findings by [9, 22, 19, 26], who developed statistical models to determine the association of some risk factors to the survival time of MMC. The risk factor, serum calcium was individually found to significantly contribute to the survival time by [26]. Whereas we found it to be significant as it interacts with infections, and it negatively relates to the survival time, given by the hazard ratio (HR > 1). However, we believe our model is more genuine and accurate, given the fact that the identified significant contributing risk factors were carefully assessed and selected based on the AIC of stepwise model selection technique, and validated to satisfy all the key model assumptions.

Conclusion

We estimated the survival probability of patients diagnosed with multiple myeloma cancer (MMC) using the semi-parametric Cox proportional hazard model. We believe the proposed Cox-PH model given by equation (4) gives an accurate estimate of the survival probability of patients diagnosed with MMC. The Cox-PH model was used to estimate the probability of the survival time because it incorporates into the model the additional information about risk factors contributing to the survival time. We identified seven risk factors and one interaction term as contributing to the survival probability of patients diagnosed with MMC. They are blood urea nitrogen (BUN)/serum creatinine, white blood cells (WBC), Bence Jone protein in the urine (BJPU), fractures, proteinuria, gender, platelets, and the interaction of infections and serum calcium. The interacting risk factors (infections and serum calcium), but individually did not significantly contribute to the survival probability. However, both together should be considered as significant interaction when identified at the same time a patient is diagnosed with MMC. Our final proposed Cox-PH model is of very high quality, robust, and efficient given by the fact that it satisfies all the key required assumptions in section (1.3.1). The stepwise model selection procedure was utilized to carefully assess and select the risk factors and the interaction term based on their statistical significance to the survival probability.

The final proposed Cox-PH model is the model with the least AIC. The identified significant contributing risk factors and the interaction have been rank according to the prognostic effect on the survival time using the hazard ratio. The interaction term between infections and serum calcium has been ranked 5, and has negative association with the length of survival time of MMC. The relevance of the proposed Cox-PH model is that we can estimate the survival probability of a patient given the values of the seven identified attributable risk factors and the interaction term. Of the seven risk factors, four of them, and the interaction term are new significant contributing risk factors to the survival time of MMC identified by our proposed model, namely; platelets, gender, white blood cells and fractures; and the interaction term, infections and serum calcium. Our findings offer further prognostic and therapeutic importance for decision making for the treatment of multiple myeloma cancer.

Base on the Cox-PH analysis of the survival times of the MMC patients, we recommend the following. (1) If additional information about what is causing the survival time (risk factors) is known, then we recommend the use of the Cox proportional hazard model to estimate the survival probability. Thus, the Cox-PH model takes into consideration the additional information given by the risk factors, hence the resulting survival probability can be more accurate, robust, and efficient. (2) The investigation of the significance of interaction terms in Cox-PH models should not be overlooked or underestimated because they can greatly affect the accurate prediction of the patients' survival rate of the multiple myeloma cancer disease, leading to dangerous medical and therapeutic/treatment issues.

References

1. Raab MS, Podar K, Breitkreutz I, Richardson PG, Anderson KC (2009) "Multiple myeloma". Lancet 374: 32439.
2. SEER Cancer Facts: Myeloma. National Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/mulmy.html
3. Daniel S, Jenna Voutsinas, Leslie Bernstein, Sophia S Wang (2014) Medication use and multiple myeloma risk Los Angelos County. Cancer Causes Control 25: 1233-1237.
4. Plasma Cell Neoplasms (Including Multiple Myeloma) Treatment" (2017) National Cancer Institute. https://www.cancer.gov/types/myeloma/patient/myeloma-treatment-pdq#section/all
5. Ferri Fred F (2013) Ferri's Clinical Advisor 2014 E-Book: 5 Books in 1. Elsevier Health Sciences 726.
6. Cancer Facts Figures (2014) American Cancer Society, url http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc04215.pdf.
7. Becker N (2011) Epidemiology of multiple myeloma. Recent Results Cancer Res 183: 25-35.
8. About Multiple Myeloma. American Cancer Society. https://www.cancer.org/cancer/multiple-myeloma/about/what-is-multiple-myeloma.html
9. BG Durie, SE Salmon, TE Moon (2011) Pretreatment Tumor Mass, Cell Kinetics, and Prognosisin Multiple Myeloma. Blood 55: 364-372.
10. Alexanian R, Balcerzak S, Bonnet JD, Gehan EA, Haut A, et al. (1975) Prognostic factors in multiple myeloma. Cancer 36: 1192-l2Ol.
11. Van de Donk NW, Mutis T, Poddighe PJ, Lokhorst HM, Zweegman S (2016) Diagnosis, risk stratification and management of monoclonal gammopathy of undetermined significance and smoldering multiple myeloma". International Journal of Laboratory Hematology 38: 11022.
12. World Cancer Report (2014) World Health Organization. Chapter 5: 13.
13. Roberts DL, Dive C, Renehan AG (2010) Biological mechanisms linking obesity and cancer risk: new perspectives". Annual Review of Medicine 61: 30116.
14. Dutta AK, Hewett DR, Fink JL, Grady JP, Zannettino AC (2017) Cutting edge genomics reveal new insights into tumour development, disease progression and therapeutic impacts in multiple myeloma". British Journal of Haematology 178: 196-208.
15. Landgren O, Kyle RA, Pfeiffer RM, Katzmann JA, Caporaso NE, et al. (2009) Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study". Blood 113: 54127.
16. Kyle RA, Rajkumar SV (2008) "Multiple myeloma". Blood 111: 296-272.
17. McCarthy PL, Holstein SA (2017) Lenalidomide Maintenance After Autologous Stem Cell Transplant in Newly Diagnosed Multiple Myeloma: a Meta-Analysis". J Clin Oncol 35: 3279-3289.
18. Sonneveld P (2012) "Bortezomib induction and maintenance treatment in patients with newly diagnosed multiple myeloma". J Clin Oncol 30: 294-655.
19. Durie BG, Salmon SE (1975) A clinical staging system for multiple myeloma. Cancer 36: 842-854.
20. Bergsagel DE (1975) Plasma cell myeloma: Prognostic factors and criteria of response to therapy, in Staquet Mi (ed): Cancer Therapy. New York, Raven 1975: 73-87.
21. Siegel RL, Miller KD, Jemal A (2017) Cancer Statistics CA Cancer. J Clin 67: 730.
22. John M Krall, Vinceent A, Uthofi John, B Harley (1975) A set-up procedure for selecting variables associated with survival. Biometrics 31: 49-57.
23. Harley JB (1971) Ten years of experience in multiple myeloma at the West Virginia University Hospital. In preparation. Morgantown, West Virginia.
24. Meyer Bruce D (1990) Unemployment Insurance and Unemployment Spells" (PDF). Econometrica 58: 757-782.
25. H Akaike (1974) A new look at the statistical model identication, IEEE Trans. Autom Control 19: 716-723.
26. Giampaolo Merlini, Jan G Waldenstrom, Suresh D Jayakar (1980) A New Improved Clinical Staging System for Multiple Myeloma Based on Analysis of 123 Treated Patients. Blood 55: 1011-1019.
27. Cox David R (1972) Regression Models and Life-Tables". Journal of the Royal Statistical Society. Series B 34: 187-220.
28. L Douglas Case (2002) Interpreting Measures of Treatment Effect in Cancer Clinical Trials". The Oncologist 7: 181-187.
29. Brody Tom (2011) Clinical Trials: Study Design, Endpoints and Biomarkers, Drug Safety, and FDA and ICH Guidelines. Academic Press 2011: 165-168.
30. Kleinbaum DG (1997) Survival Analysis: A Self-Learning Text. New York: Springer 1997: 1-324.
31. Kalbeisch JD, Prentice RL (1980) The Statistical Analysis of Failure Time Data. New York: John Wiley Sons 1980: 1-321.

32. L Douglas Case, Gretchen Kimmick, Electra D Paskett, Kurt Lohman, Robert Tucker (2002) Interpreting Measures of Treatment Effect in Cancer Clinical Trials. The Oncologist 7: 181-187.

Citation: *Lohuwa Mamudu, Chris P Tsokos and Otunuga Oluwaseun E (2020) Survival Analysis of Multiple Myeloma Cancer (MMC) Using the Cox-Proportional Hazard Model. Journal of Medical & Clinical Research 5(7):147-156.*