

A New Statistical Modeling Approach for Survival Analysis of Cancer Patients—Multiple Myeloma Cancer

Lohuwa Mamudu*, Chris P. Tsokos

Department of Mathematics and Statistics, University of South Florida, Tampa, USA

Email: *lohuwa@usf.edu, ctsokos@usf.edu

How to cite this paper: Mamudu, L. and Tsokos, C.P. (2021) A New Statistical Modeling Approach for Survival Analysis of Cancer Patients—Multiple Myeloma Cancer. *Open Journal of Applied Sciences*, 11, 365-378.

<https://doi.org/10.4236/ojapps.2021.114027>

Received: February 25, 2021

Accepted: April 9, 2021

Published: April 12, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Background: The Cox Proportional Hazard (Cox-PH) model has been a popularly used method for survival analysis of cancer data given the survival times as a function of covariates or risk factors. However, it is very seldom to see the assumptions for the application of the Cox-PH model satisfied in most of the research studies, raising questions about the effectiveness, robustness, and accuracy of the model predicting the proportion of survival times. This is because the necessary assumptions in most cases are difficult to satisfy, as well as the assessment of interaction among covariates. **Methods:** To further improve the therapeutic/treatment strategy for cancer diseases, we proposed a new approach to survival analysis using multiple myeloma (MM) cancer data. We first developed a data-driven nonlinear statistical model that predicts the survival times with 93% accuracy. We then performed a parametric analysis on the predicted survival times to obtain the survival function which is used in estimating the proportion of survival times. **Results:** The new proposed approach for survival analysis has proved to be more robust and gives better estimates of the proportion of survival than the Cox-PH model. Also, satisfying the proposed model assumptions and finding interactions among risk factors is less difficult compared to the Cox-PH model. The proposed model can predict the real values of the survival times and the identified risk factors are ranked according to the percent of contribution to the survival time. **Conclusion:** The new proposed nonlinear statistical model approach for survival analysis of cancer diseases is very efficient and provides an improved and innovative strategy for cancer therapeutic/treatment.

Keywords

Health Science, Multiple Myeloma Cancer, Cancer Therapeutic, Cox-PH Model, Statistical Model, Survival Analysis, Probability Estimation

1. Introduction

In our previous study [1] [2], we obtained the parametric, non-parametric, and semi-parametric analysis of the survival times of 48 patients diagnosed with multiple myeloma. In the parametric analysis, we found the survival times to follow the three-parameter lognormal distribution, and then we proceeded to obtain the survival function. In the case of the non-parametric analysis, we used the commonly used Kaplan-Meier model to obtain the survival function and then estimate the probability of survival times. For the semi-parametric analysis, we adopted the Cox proportional hazard to obtain the survival function of the survival times. In comparing the parametric and non-parametric analysis of the survival time, we justified that the parametric method is more robust and efficient. However, none of the two methods take into consideration the risk factors contributing to the survival time. Therefore, we believe the Cox-PH model is more relevant for estimating the proportion of patients' survival beyond a given time than the other two because it takes into account the additional useful information given by the risk factors (covariates) contributing to the survival times. The Cox-PH on the other hand has some flaws. The necessary assumptions for applying the Cox-PH model are often difficult to satisfy, so as the finding of interaction among the risk factors. As a result, most research studies use the Cox-PH model without satisfying the underlying assumptions and also finding the interaction among covariates. This makes it difficult to justify the genuineness of conclusions made from using the Cox-PH model and the accuracy of predicting the proportion of survival.

Also in another research study [3], we developed a data-driven nonlinear statistical model of the 48 patients diagnosed with MM cancer and obtained a very accurate and high coefficient of determination, $R^2 = 0.8741$ along with $R_{adj}^2 = 0.8401$. We further utilized the bootstrapping resampling technique to increase the sample size of the survival times by drawing 300 samples from the 48 original observations with replacement, resulting in asymptotic significance of the coefficients or parameters of the risk factors and an increase $R^2 = 0.9116$ along with $R_{adj}^2 = 0.9085$. The nonlinear statistical model resulted in 93% accuracy of predicting the test data. Similar to the Cox-PH model, the nonlinear statistical model takes into consideration the additional information given by the risk factors contributing to the survival times. Most often the Cox-PH model has been employed in analyzing survival data with given covariates or risk factors. Whilst the Cox-PH model is used to estimate the proportion of patients surviving beyond a given time for a given number of risk factors, the nonlinear statistical model is used to estimate or predict the real value of the survival times of patients for a given number of risk factors. In the present study, we developed the survival function from the nonlinear statistical model and use it to estimate the proportion of patients survival of MM cancer beyond a given survival time, and compared it with the survival function of the commonly used Cox-PH model as a means of survival data analysis of the survival time as a function of covariates or risk factors.

2. Methods

2.1. Data

We obtained the data used in the present study from West Virginia University Medical Center provided by Harley [4] [5]. The original data consist of the survival times of 72 multiple myelomas (MM) patients diagnosed and treated with alkylating agents [4]. However, 65 out of 72 patients have complete data on 16 concomitant variables (risk factor) believed to be causing MM, while the remaining 7 were ignored during our analysis due to missing data in at least one of the 16 risk factors. At the time a patient is diagnosed with MM, the 16 risk factors were recorded and the time up to which the patient survived the disease was noted (called the survival time from diagnosis to the nearest month). Of the 65 patients, 48 and 17 were uncensored and censored, respectively. In this study, we utilized only the 48 uncensored patients whose survival times were known (*i.e.* completed the trial). The survival time of patients is the response variable with 16 risk factors. Thus, we have one continuous response variable, 11 continuous risk factors, and 5 categorical risk factors. A detailed description of the response variable and the 16 risk factors are given in **Table 1**.

Before we performed the analysis and modeling of the survival times of the 48 patients with MM, we first investigated whether there is a difference in the survival times between the males and females. Because we have small data of only

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

Table 2. Kruskal-Wallis rank sum test of the difference in survival times between males and females MM Patients.

Type of Test	Survival time	Data: list (males, females)
Kruskal-Wallis	$Chi\text{-squared}(\chi^2) = 0.40914$	$p\text{-value} = 0.5224$

48 patients, we used the non-parametric Kruskal-Wallis test [6] [7] to compare the difference in survival times of males and females. From **Table 2**, the Kruskal-Wallis rank-sum test resulted in a large $p\text{-value} = 0.5224$, hence failing to reject the null hypothesis (*i.e.* $H_0 : \mu_M = \mu_F$) that there is no difference in the survival times of males and females. Given that we have a small sample size of only 48 patients and the fact that survival times of males and females are the same, provides a good justification to use the entire data of both male and female for the analysis and modeling of the survival times of the MM cancer patients.

2.2. The Models

In **Table 3**, we show the significant attributable risk factors identified for the CoxPH model and those of the nonlinear statistical model in ranking order from most significant to the least significant. **Table 4** displays the identified covariates or risk factors by the Cox-PH model along with their coefficient estimates, standard errors, hazard ratios (HR), and the 95% confidence interval of HR. An extensive review of the development of the Cox-PH model is given by Lohuwa Mamudu, Chris P Tsokos, Otunuga Oluwaseun E (2020) [2]. **Table 5** also shows the identified covariates or risk factors by the nonlinear statistical model along with the percent contribution of each of the risk factors to the survival times of MM. A detailed review of the nonlinear statistical model is provided by M. Lohuwa and C. Tsokos (2020) [3]; and more on statistical modeling [8]. Interestingly, we can recognize that almost all the risk factors (covariates) that were identified to be significantly contributing to the survival times of MM cancer by the Cox-PH model were as well found significant in the nonlinear statistical model. The only exception is that the ranking positions are different. This is an extremely important feature to support the high quality and accuracy of our research findings. Thus, the fundamental justification for the comparison of the two models is given below in **Table 3**. We can obtain more and detailed information about the risk factors causing multiple myeloma from [9] [10] [11].

The rankings of the risk factors in the Cox-PH model are based on the hazard ratio, HR, and those of the nonlinear statistical model are based on the coefficient of determination, R^2 . The HR measures the relative risk or the prognostic effect of the covariates to the length of survival time. The higher the HR, the more the impact or contribution of a covariate to the survival time of MM. Generally, $HR > 1$ implies that the covariate has an increased risk of association with the length of survival time, $HR < 1$ implies that the covariate has a decreased risk of association with the length of survival time, and $HR = 1$ means that the covariate has no risk of association with the length of survival time. R^2 , on the other hand, measures the variability in the survival time explained by the covariates or

Table 3. Significant attributable risk factors of the Cox-PH model and the nonlinear statistical model.

Rank	Cox-PH	Nonlinear Statistical Model
1	blood urea nitrogen (BUN)	Bence Jone protein in urine, X_{13}
2	White Blood Cells (WBC)	Blood Urea Nitrogen (BUN), X_1
3	Bence Jone protein in the urine	Infections, X_4
4	Fractures	% Myeloid cells in peripheral blood, X_{11}
5	Infections & serum calcium	serum calcium, X'_{16}
6	Proteinuria	Fractures, X_8
7	Gender	WBC & total Serum protein, $X_7 \cdot X'_{14}$
8	Platelets	Gender, X_6
9		Platelets, X_3
10		Age, X_5

Table 4. Significant attributable risk factors of the Cox-PH model and the nonlinear statistical model.

Rank	Covariates	Coeff (β)	S.E.	HR ($\exp(\beta)$)	95% CI of HR	
1	X_1	2.008	0.619	7.454	2.215	25.056
2	X_7	1.878	0.773	6.543	1.439	29.745
3	X_{13}	1.576	0.418	4.835	2.131	10.972
4	X_8	0.854	0.409	2.349	1.053	5.243
5	X_4, X_{16}	0.113	0.040	1.212	1.035	1.213
6	X_{12}	0.108	0.030	1.114	1.049	1.183
7	X_6	-0.815	0.391	0.443	0.206	0.952
8	X_3	-1.608	0.502	0.200	0.075	0.536

Table 5. Significant attributable risk factors of the nonlinear statistical model.

Rank	Risk Factors	R^2	% Contribution
1	X_{13}	0.2672	30.57
2	X_1	0.2052	23.48
3	X_4	0.0949	10.86
4	X_{11}	0.089	10.18
5	X'_{16}	0.0661	7.56
6	X_8	0.0613	7.01
7	X_7, X'_{14}	0.0379	4.34
8	X_6	0.0329	3.76
9	X_3	0.011	1.26
10	X_5	0.0086	0.98
Total		0.8741	100

risk factors. The higher the percentage of R^2 of a given covariate, the more contribution it makes towards explaining the variation in the survival time. Therefore, both HR and R^2 can be said to play a similar role in determining the prognostic effect of covariates on the survival time. However, R^2 is most recommended and efficient given that it measures the overall contribution of the risk factors in the model. Hence, R^2 gives more accurate information about the impact or prognostic effect of risk factors to the survival times than the HR.

The Cox-PH model identified eight attributable risk factors, including one interaction. The nonlinear statistical model identified ten attributable risk factors, including one interaction. Blood urea nitrogen is ranked first as the highest prognostic factor in the Cox-PH model but ranked second in explaining the variability in the survival times of the nonlinear statistical model. Bence Jone protein in urine was ranked first as the highest contributor in explaining the variability in the survival times but ranked as the third prognostic factor in the Cox-PH model. Interestingly, both models identified only one significant interaction. However, the interacting risk factors are different in the two models. The Cox-PH model identified infections and serum calcium as interaction, and the nonlinear statistical model identified white blood cells (WBC) and total serum protein as interaction factors. Another interesting information we can derive from **Table 3** is that the risk factors making-up the interaction in the Cox-PH model was identified to be individually significantly contributing to the survival times in the nonlinear statistical model.

Furthermore, white blood cells (WBC) individually significantly contributed to the proportion of survival in the Cox-PH model, but it was part of the interaction term identified by the nonlinear statistical model. The risk factor, proteinuria, was identified as significant in contributing to the proportion of Survival by the Cox-PH model, but not identified significant by the nonlinear statistical model. Whereas, age and myeloid cells in peripheral blood were significantly identified to be contributing to the survival time by the nonlinear statistical model, but not identified by the Cox-PH model. For the nonlinear statistical model, a positive coefficient or parameter means that a unit increase in the risk factor increases the survival time by the size of the coefficient, and a negative coefficient means that the survival time decreases by the size of the coefficient whenever there is a unit increase in the risk factor given that the other risk factors remain unchanged.

In the Cox-PH model, a unit increase in a covariate with a positive coefficient leads to a decrease in the proportion of survival beyond a given time by the size of the coefficient. Whereas increasing a covariate with a negative coefficient by a unit will increase the proportion of survival beyond a given time by the size of the coefficient, given that the remaining other covariates remain constant. It is important to recognize that the criterion of model selection of the Cox-PH model was based on choosing the model with the least Akaike information criterion (AIC) [12], whereas the nonlinear statistical model was based on choosing the largest R^2 along with the R^2_{adj} and the least AIC. Though the two models have

some differences, in general, each model identified significantly most of the risk factors found by the other. We recommend that the nonlinear statistical model is more powerful in identifying the risk factors and their percentage contribution to the response, the survival time.

3. Results

3.1. Development of the Survival Function of the Nonlinear Statistical Model

In the present study, we find the survival function of the death times t^* predicted from the final proposed nonlinear statistical model with 300 failure or survival times [3]. The proposed statistical model we developed in [3], is given by

$$t_i^* = \exp\left(-4.377 - 1.097X_1 + 0.332X_{3\text{normal}} + 0.949X_{4\text{present}} + 0.016X_5 + 0.562X_{6\text{femal}} - 0.586X_{8\text{present}} + 0.022X_{11} - 1.268X_{13\text{none}} + 4.151X'_{16} - 0.252X_7X'_{14}\right), \quad (1)$$

where $i = 1, 2, \dots, 300$ and

$$X_j = \begin{cases} 1 - e^{-X_j}, & \text{if } X < 0 \\ -1 + e^{X_j}, & \text{otherwise, for } j = 14, 16 \end{cases} \quad (2)$$

Using the above model, we generated $t_1^*, t_2^*, \dots, t_{300}^*$ survival times that are based on the risk factors that have been identified for each patient of MM. That is, t_1^* is the survival time of patient 1 given the influence of each risk factor, t_n^* is the survival time of the n^{th} patient based on the influence of each of its risk factors. To investigate the distribution of the predicted survival times t^* of the MM patients, we first displayed the descriptive statistics of the survival times t^* . A detailed explanation of the implication of the value of the statistic (especially kurtosis and skewness) is given in our study in [1]. The values of the descriptive statistics given in **Table 6** shows that t^* is skewed, given by the higher value of skewness and kurtosis.

We found the pdf distribution of the t^* data to follow the three parameter-lognormal probability distribution, which is the same distribution we found for the base sample of 48 patient's survival times t . This was expected because the 300 bootstrap samples come from the 48 samples with the 3p-lognormal probability distribution, justifying the high quality of our proposed nonlinear statistical model, [3]. To further support the fact that t and t^* follow the same distribution, we performed a non-parametric Kruskal-Wallis test [6] [7] to compare the difference in two survival times t and t^* . From **Table 7**, the Kruskal-Wallis rank-sum test resulted in a very large p -value = 0.9066 \approx 1, hence failing

Table 6. Descriptive statistics of survival times t^* of multiple myeloma.

Survival time	Mean	Median	Std Err	Std Dev	Kurtosis	Skewness
t^*	22.07	15.68	1.14	19.59	1.46	1.46

Table 7. Kruskal-Wallis rank sum test of the difference between t and t^* .

Type of Test	Survival time	Data: list (t, t^*)
Kruskal-Wallis	$Chi\text{-squared}(\hat{\chi}^2) = 0.013776$	$p\text{-value} = 0.9066$

to reject the null hypothesis (i.e. $H_0 : \eta_t = \eta_{t^*}$), indicating no difference in the survival times t and t^* . In **Table 8** below, we are given the estimates of the parameters of the 3p-lognormal pdf of the t^* . We obtained an approximate estimate of the parameters of the 3p-lognormal parameter distribution utilizing the maximum likelihood estimation (MLE) method as we presented in [1] [13] [14].

Thus, the 3p-lognormal pdf, $f(t^*)$, of the survival times of 300 MM patients is given by

$$f(t^* | \gamma^*, \mu^*, \sigma^{*2}) = \begin{cases} 0, & \text{if } t^* \leq \gamma^* \\ (2\pi\sigma^{*2})^{-\frac{1}{2}}(t^* - \gamma^*)^{-1} \exp\left(-\frac{1}{2}\left(\frac{\ln(t^* - \gamma^*) - \mu^*}{\sigma^*}\right)^2\right), & \text{if } t^* > \gamma^* \end{cases} \quad (3)$$

By substituting the parameter estimates given in **Table 2**, we have

$$f(t^* | \gamma^*, \mu^*, \sigma^{*2}) = \begin{cases} 0, & \text{if } t^* \leq 4.4832 \\ 0.38253(t_i^* - 4.4832)^{-1} \exp\left(-\frac{1}{2}\left(\frac{\ln(t_i^* - 4.4832) - 2.5603}{1.0303}\right)^2\right), & \text{if } t^* > 4.4832 \end{cases}$$

The plot of the pdf, $f(t^*)$, of the survival times of the 300 MM patients is given by **Figure 1**. With the pdf plot, we can compute the probability that the survival time of a patient diagnosed with multiple myeloma will fall between a given time t_k^* and $t_{(k+1)}^*$. For example, we can compute the probability that an MM patient will survive between 20 months and 40 months, given by

$P(20 \leq t^* \leq 40) = 0.025 - 0.008 \approx 0.017$, as shown in **Figure 2**. We interpret this as there is approximately a 1.7% probability that a patient will survive between 20 months and 40 months. On the other hand, for survival times t , $P(20 \leq t \leq 40) = 0.019 - 0.007 \approx 0.012$.

The cumulative distribution function of the 3p-lognormal, $F(t^*)$, of the survival times of t^* is given by

$$F_{T^*}(t^* | \gamma^*, \mu^*, \sigma^{*2}) = \frac{1}{\sqrt{2\pi}} \int_0^{t^*} \exp\left(-\frac{1}{2}z^2\right) dz = \Phi\left(\frac{\ln(t_i^* - \gamma^*) - \mu^*}{\sigma^*}\right) \quad (4)$$

Substituting the parameter estimates obtained we have,

$$F_{T^*}(t^* | \gamma^*, \mu^*, \sigma^{*2}) = P[t^* \leq T^*] = \Phi\left(\frac{\ln(t_i^* - 4.4832) - 2.5603}{1.0303}\right),$$

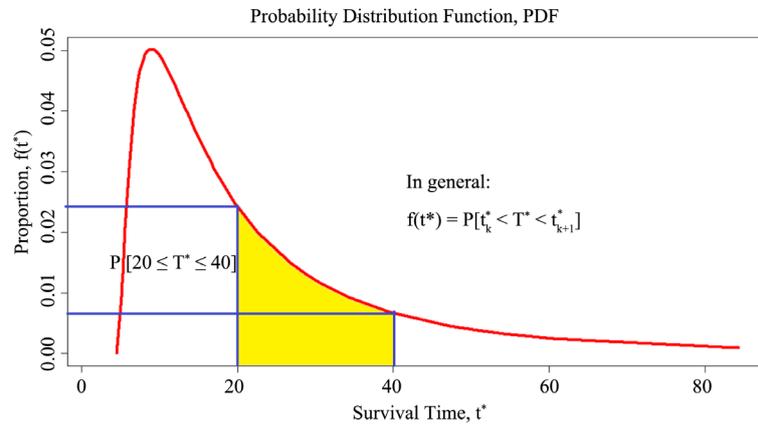


Figure 1. Probability distribution function of survival time, t^* of multiple myeloma.

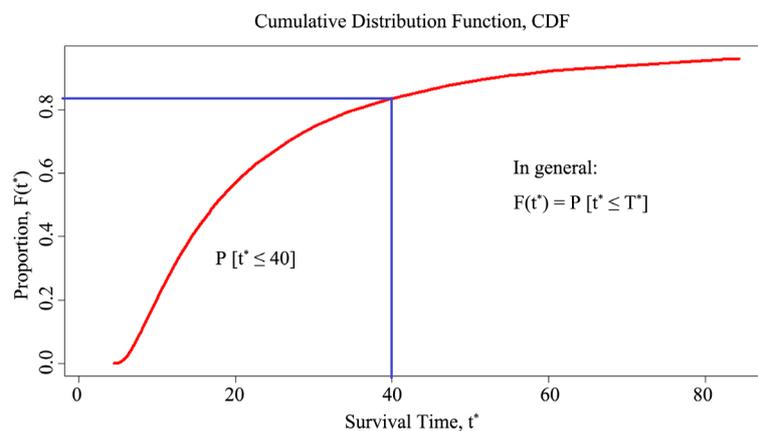


Figure 2. Cumulative distribution function of survival time, t^* of multiple Myeloma.

Table 8. Parameter estimates for the 3p-Lognormal pdf for t^* .

Survival time	Location ($\hat{\gamma}$)	Scale ($\hat{\mu}$)	Shape ($\hat{\sigma}$)
t^*	4.4832	2.5603	1.0303

where $\Phi(\cdot)$ is the standardized normal CDF. Figure 2 is a graph of the CDF of the survival times t^* of multiple myeloma patients. That is, we can estimate the probability that a patient with MM survives up to a given time t^* from Figure 2. For example, the probability that an MM patient will survive up to time $t^* = 40$ months can be computed as; $F(t^* = 40) = P(t^* \leq 40) \approx 0.82$, as shown in Figure 3. Thus, there is about 82% chance that an MM patient will survive up to 40 months. On the other hand, we can find the probability that the patient will survive beyond 40 months to be $P(t^* > 40) = 1 - F(t^* = 40) = 0.18$. For t , $F(t = 40) = P(t \leq 40) \approx 0.83$, and $P(t > 40) = 1 - F(t = 40) = 0.17$.

The survival function $\hat{S}(t^*)$ of the survival times t^* is given by

$$\hat{S}(t_i^* | \gamma^*, \mu^*, \sigma^{*2}) = 1 - F_{r^*}(t^* | \gamma^*, \mu^*, \sigma^{*2}) = 1 - \Phi\left(\frac{\ln(t_i^* - \gamma^*) - \mu^*}{\sigma^*}\right) \quad (5)$$

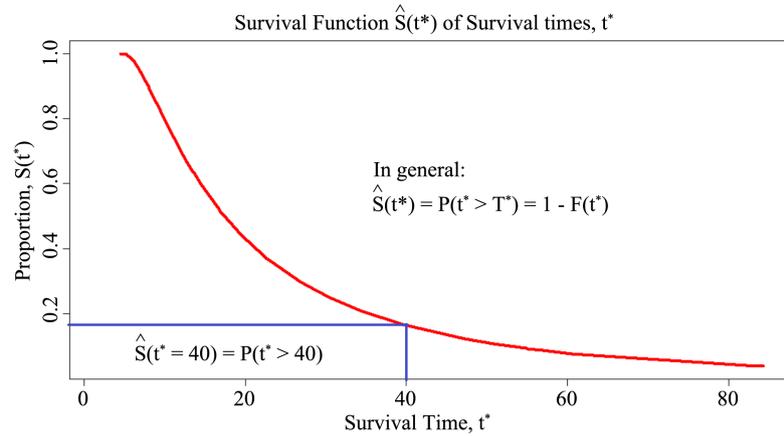


Figure 3. Survival function of the survival times, t^* of multiple myeloma patients.

We substitute the estimates of the parameter given in **Table 2**, we have

$$\hat{S}(t_i^* | \gamma^*, \mu^*, \sigma^{*2}) = 1 - F_{T^*}(t_i^* | \gamma^*, \mu^*, \sigma^{*2}) = 1 - \Phi\left(\frac{\ln(t_i^* - 4.4832) - 2.5603}{1.0303}\right),$$

where $\Phi(\cdot)$ is the standardized normal CDF of the survival time t^* and $\hat{S}(t^*)$ estimates the probability that a patient with multiple myeloma survives beyond a given time t^* . From **Figure 3**, we can compute the probability that an MM patient will survive beyond 40 months; that is,

$$\hat{S}(t^* = 40) = P(t^* > 40) \approx 0.18. \text{ For } t, \hat{S}(t = 40) = P(t > 40) \approx 0.17.$$

Algorithm for the Nonlinear Statistical Modeling to Survival Analysis

The flowchart in **Figure 4** shows the algorithmic process of performing the survival analysis described in this study. The process involves developing a high-quality statistical model that gives high prediction accuracy, followed by making a prediction, and finally performing parametric analysis on the predicted values and finding the survival function for estimating the proportion of survival time.

3.2. Comparing the Survival Function of the Cox-PH Model with That of the Non-Linear Statistical Model of Survival Times of Multiple Myeloma

From [2] [3], we found both the Cox-PH model and the nonlinear statistical model to be of high quality since they satisfy all the respective required model assumptions and pass all the criteria for measuring the robustness and efficiency of a high profile model. Whiles the Cox-PH model predicts the proportion of survival at a given time given the values of the significant attributable risk factors or covariates, the nonlinear statistical model predicts the real value of the survival time given values of the attributable risk factors. In [1], we found the Cox-PH model to be of utmost importance and hence recommended it as more relevant, since it takes into account the additional useful information given by the attributable risk factors of the survival time, at the time a patient is diagnosed with MM cancer. Now, we draw the comparison of the survival estimates of the sur-

vival times t by the Cox-PH model with the survival estimates of the survival times t^* by the nonlinear statistical model of the MM patients, given by **Figure 5**. We can see that the survival function of the nonlinear statistical model lies above that of the Cox-PH model. That is, the nonlinear statistical model consistently gives a higher prediction of the probability of survival of patients diagnosed with MM cancer beyond a given time t^* than the Cox-PH model, making it a better choice. This is because the underlying distribution of the nonlinear statistical model survival function of the survival times t^* is based on a well-defined parametric probability distribution, which is more powerful and sophisticated than the survival function of the survival times t of the semi-parametric Cox-PH model. Therefore, it is not surprising to see that the nonlinear statistical model performs better in estimating the proportion of the survival time than the Cox-PH model.

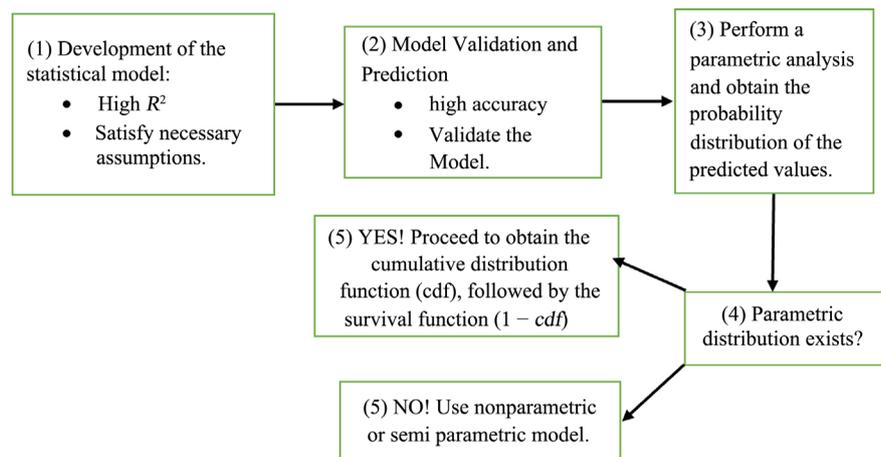


Figure 4. Flow chart of the development of the survival function of the nonlinear statistical model.

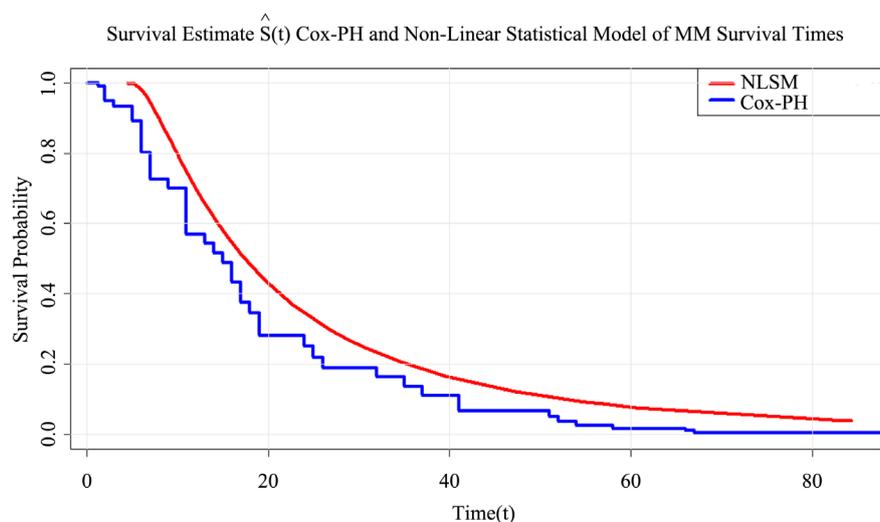


Figure 5. Comparison of the survival function of Cox-PH and the non-linear statistical model of MM.

4. Discussion/Conclusions

4.1. Discussion

In the present study of the survival times t of 48 patients diagnosed with MM, we predicted 300 survival times t^* from the final proposed statistical model in Equation (1) based on the bootstrap resampling method and investigated the probability distribution. We found the pdf probability distribution of the t^* follows the 3p-lognormal (same as the probability distribution of the original sample of 48 MM patients). Using the method of maximum likelihood parameter estimation as described in [1], we obtained the parameter estimates of the 3p-lognormal pdf $f_{T^*}(t^*)$ as given by **Table 8**. We found the CDF $F_{T^*}(t^*)$ by integrating the $f_{T^*}(t^*)$ with respect to t^* , and then find the survival function $\hat{S}(t^*)$ (*i.e.* $1 - F_{T^*}(t^*)$). The CDF estimates the survival proportion up to a given time t^* , and the survival function estimates the proportion of survival beyond a given time t^* . We then compare the $\hat{S}(t^*)$ of the statistical nonlinear statistical given by Equation (1) with the $\hat{S}(t)$ of the Cox-PH model given by Equation (7) in [2], as shown by **Figure 5**. The comparison shows that the $\hat{S}(t^*)$ of the nonlinear statistical model provided a better estimate of the proportion of survival times than the $\hat{S}(t)$ of the Cox-PH model, given that the $\hat{S}(t^*)$ of the nonlinear statistical model is developed from the originally identified well-defined parametric probability distribution of the patients diagnosed with MM.

In **Table 3** of the ranking of the significant attributable risk factors identified by the two models, the Cox-PH model identified infections and serum calcium as an interaction term, ranked fifth, as significantly contributing to the proportion of the survival times of MM patients. However, those two risk factors were individually identified to be significantly contributing to the survival times by the nonlinear statistical model and ranked third and fourth respectively. The ranking process of the Cox-PH model is based on the prognostic effect of the risk factor on the survival time using the hazard ratio, and the nonlinear statistical model ranking of the risk factors is based on the percentage of contribution to the variability in the survival time explained by the significantly identified attributable risk factors (*i.e.* the coefficient of determination, R^2). Both the hazard function and the R^2 can play a similar role in measuring the prognostic effect of a given risk factor on the survival time. However, we recommend the use of the R^2 along with the R^2_{adj} because it measures the entire variability of the survival times of MM explained by the risk factors with a high degree of accuracy.

It is very important to recognize that both models are high profile considering the quality involved in the model building process. However, in reality, medical personnel and patients would be more concerned about the real value of the survival time rather than the probability of surviving. Also, we would be more concerned with the percentage of contribution of a risk factor to the survival time than whether it is a good or bad prognostic factor to the survival time. Thus, making the ranking of the risk factors by the nonlinear statistical model more

relevant. Moreover, developing the Cox-PH model is more difficult satisfying the assumptions and finding the interaction between covariates. Douglas G. Altman and Bianca L. De Stavola (1994) [15] presented the practical problems in fitting a proportional hazards model. Ian Ford, John Norrie, and Susan Ahmadi (1995) [16] also assessed model inconsistency illustrated by the Cox-PH model. These studies allow us to strongly support the robustness of using the statistical model approach to survival analysis, which provides more flexibility than the Cox-PH model. The present finding shows that we can obtain a better and accurate prediction of the proportion of survival time as long as we can find a well-defined parametric probability distribution that characterizes a given cancer survival data. Given that our objective is to maximize the survival times, the nonlinear statistical model is better for improving the therapeutic/treatment strategy of maximizing the survival times of multiple myeloma patients than the Cox-PH model.

4.2. Conclusion

In the present study, we have demonstrated that both the Cox-PH model and the nonlinear statistical model are of high quality and useful. The two models predict the proportion of survival time with a high degree of accuracy. However, we recommend the nonlinear statistical model over the Cox-PH model because it offers a better prediction of the proportion survival of the MM cancer patients, as shown in **Figure 5**. The nonlinear statistical model does not only provide a better prediction estimate of the survival probability, but it also provides us with several useful outcomes. 1) We can predict the real values of the survival time of a patient given the significant attributable risk factors and the interaction term. 2) We can rank the attributable risk factors and the interaction term according to the percentage of contribution to the survival time. 3) We can perform surface response analysis for the maximization of the survival time, given the values of the risk factors and the interaction. 4) We can generate confidence intervals for the survival time. 5) We can also perform parametric analysis of the predicted survival values and obtain better survival estimates (*i.e.* the probability that a patient survival beyond a given time) than the survival estimates from the popularly known traditional survival models. On the other hand, we can only obtain the outcome (5) from using the Cox-PH model. The present study provides therapeutic/treatment significance for further improvement in the survival times of patients diagnosed with multiple myeloma.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Mamudu, L. and Tsokos, C.P. (2020) Parametric and Non-Parametric Analysis of the Survival Times of Patients with Multiple Myeloma Cancer. *Open Journal of Ap-*

- plied Sciences*, **10**, 118-134. <https://doi.org/10.4236/ojapps.2020.104010>
- [2] Mamudu, L., Tsokos, C.P. and Otunuga Oluwaseun, E. (2020) Survival Analysis of Multiple Myeloma Cancer Using the Cox-PH Model. *Medical Clinical Research Journal*, ISSN 2577-8005. <https://doi.org/10.33140/MCR.05.07.05>
- [3] Lohuwa, M. and Tsokos, C. (2020) Data-Driven Statistical Modeling and Analysis of the Survival Times of Multiple Myeloma. *Health Science Journal*, **14**, 1. <https://doi.org/10.36648/1791-809X.14.1.693>
- [4] Krall, J.M., Uthoff, V.A. and Harley, J.B. (1975) A Set-Up Procedure for Selecting Variables Associated with Survival. *Biometrics*, **31**, 49-57. <https://doi.org/10.2307/2529709>
- [5] Harley, J.B. (1971) Ten Years of Experience in Multiple Myeloma at the West Virginia University Hospital. Morgantown.
- [6] Wallis, K. (1952) Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, **47**, 583-621. <https://doi.org/10.1080/01621459.1952.10483441>
- [7] Corder, G.W. and Foreman, D.I. (2009) Nonparametric Statistics for Non-Statisticians. John Wiley Sons, Hoboken, 99-105. <https://doi.org/10.1002/9781118165881>
- [8] Abu Sheha, M. and Tsokos, C. (2019) Statistical Modeling of Emission Factors of Fossil Fuels Contributing to Atmospheric Carbon Dioxide in Africa. *Atmospheric and Climate Sciences*, **9**, 438-455. <https://doi.org/10.4236/acs.2019.93030>
- [9] Bethesda, M.D. (2018) SEER Cancer Facts: Myeloma. National Cancer Institute. <https://seer.cancer.gov/statfacts/html/mulmy.html>
- [10] American Cancer Society (2018) About Multiple Myeloma. The American Cancer Society Medical and Editorial Content Team. <https://www.cancer.org/cancer/multiple-myeloma/about/what-is-multiple-myeloma.html>
- [11] World Health Organization (2014) World Cancer Report 2014. Chapter 5.13. World Health Organization, Geneva.
- [12] Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19**, 716-723. <https://doi.org/10.1109/TAC.1974.1100705>
- [13] Calitz, F. (1973) Maximum Likelihood Estimation of the Parameters of the Three Parameter Lognormal Distribution—A Reconsideration. *Australian Journal of Statistics*, **9**, 221-226.
- [14] Cohen Jr., A.C. (1951) Estimating Parameters of Logarithmic-Normal Distributions by Maximum Likelihood. *Journal of the American Statistical Association*, **46**, 206-212. <https://doi.org/10.1080/01621459.1951.10500781>
- [15] Altman, D.G. and De Stavola, B.L. (1994) Practical Problems in Fitting a Proportional Hazards Model to Data with Udated Measurements of the Covariates. *Statistics in Machine*, **13**, 301-341. <https://doi.org/10.1002/sim.4780130402>
- [16] Ford, I., Norrie, J. and Ahmadi, S. (1995) Model Inconsistency, Illustrated by the Cox Proportional Hazards Model. *Statistics in Machine*, **14**, 735-746. <https://doi.org/10.1002/sim.4780140804>