INDONESIAN COUNCIL OF PREMIER STATISTICAL SCIENCE

Indonesian Council of Premier Statistical Science
ISSN: 3030-9956/https://dx.doi.org/10.24014/icopss.v4i2.37880

Volume 4 Issue 2, 45 – 50, August 2025

Application of the Cox Proportional Hazard Model on Survival Data of Multiple Myeloma Patients Using the R Application

Achmad Riyan[1], Surya Nengsih[2]

[1.2] *Mathematics Study Program, Universitas Islam Negeri Sultan Syarif Kasim Raiu, Indonesia*

Received: 23 July 2025 Revised: 27 August 2025 Accepted: 30 August 2025 Published: 31 August 2025

Abstract - *Multiple Myeloma is a type of blood cancer characterized by the proliferation of malignant plasma cells in the bone marrow and can affect the patient's survival. This study aims to analyze the influence of age, sex, and protein levels on patient survival time. Multiple Myeloma uses the Cox Proportional Hazards model. The data used came from 47 patients with variables of survival time, patient status (dead or alive), age, gender, and protein content. The analysis was carried out using R software. The model match test with the likelihood ratio test also showed insignificant results, but testing of the assumption of proportional hazards through residual Schoenfeld showed that all variables met the model's assumptions. Thus, the Cox PH model in this study is technically valid, but its predictive power is still limited, so further model development is recommended by increasing the amount of data or considering other more relevant variables.*

Keywords: *survival analysis, assumption proportional hazards, Cox proportional hazards, multiple myeloma, likelihood test.*

1. Introduction

Multiple myeloma is a type of cancer that originates from plasma cells, which are part of the immune system that usually helps fight infection. In this disease, plasma cells develop abnormally and excessively in the bone marrow, accompanied by the presence of monoclonal proteins found in blood or urine. Common complications include kidney damage, excessively high blood calcium levels (hypercalcemia), letic bone lesions, and anemia that can lead to significant disability (morbidity) and mortality[1]. Multiple Myeloma accounts for about 10% of hematological malignancies, with an incidence of 5.5 cases per 100,000 population in the 34-44 years of age[2].

This disease often damages bones and causes lytic lesions to form, so that bones become brittle and easily break. Bone abnormalities are found in most patients with multiple myeloma at the time they are first diagnosed. Bone damage can be in the form of bones becoming thinner or more brittle (osteopenia or osteoporosis) to causing fractures for no apparent cause (pathological fractures) this condition can cause extreme pain and decrease the quality of life. Bone damage occurs because cancer cells disrupt the natural balance between the cells that make up bone (osteoblasts) and the cells that destroy bone (osteoclasts) and as a result the bone repair process is disrupted, thus aggravating bone damage[3].

Survival analysis is a statistical method used to study an individual's survival to a specific event, such as death and disease recurrence. This analysis measures the time from the beginning of the observation to the event occurring, and can handle sensor data[4]. One of the models often used is the Cox Proportional Hazard regression model, a semi-parametric approach that models the relationship between survival time and one or more independent variables [5].

In the field of media statistics, survival analysis is a very relevant approach, especially using the Cox Proportional Hazards method which is able to test the influence of many variables on the time of occurrence. Therefore, this study uses the Cox Proportional Hazards method approach, which allows a multivariate analysis of risk factors at survival time, without the need for specific distribution assumptions.

Several previous studies have applied the Cox Proportional Hazards model to analyze survival data. Nainggolan and Manullang (2024) research used this model to analyze the survival of coronary heart patients and found that only the age variable had a significant effect on the length of the patient's stay[6]. Khinanti et al.'s (2021) research applied the Cox model with the Breslow, Efron, and Exact approaches in hemodialysis patients, and the results showed that systolic blood pressure, hemoglobin levels, and dialysis duration had a significant effect on survival time, with the exact approach producing the best model based on AIC values [7]. Research by Puspita et al. (2022) used the Cox Proportional Hazards model to analyze the survival of asthma patients. The results showed that age and comorbidities had a significant effect on recovery time, while gender did not. The model meets the assumption of proportional hazards and can be used to predict factors that affect a patient's recovery time [8]. The research of Elnatan et al. (2018) used the Cox Proportional Hazards model to analyze the survival time of lung cancer patients. The test of the assumption of proportional hazards was carried out through log-log graphs and the Goodness of Fit test using Schoenfeld Residual correlation. The results showed that most of the variables met the PH assumptions. However, age does not meet these assumptions, so the influence of age on risk is not constant all the time [9]. The research of Saragih et al. (2024) applied the Cox Proportional Hazards model with the Breslow approach to analyze the survival of hemodialysis patients in one of the hospitals. Of the eight variables analyzed, only systolic blood pressure was shown to have a significant effect on patient survival time. The initial model was evaluated using the proportional hazards assumption test and the likelihood ratio test, and the test results showed that the model was feasible and valid to use [10].

The selection of the Cox Proportional Hazards model in this study is based on its success in various previous studies in analyzing survival data. This model is semi-parametric, so it does not require a special form of distribution on the hazard function, and is very suitable for data containing censored observations, such as the case of patients who have not experienced death during the observation period. Previous studies, both in patients with heart disease, asthma, lung cancer, and hemodialysis, have shown that Cox's model is able to identify variables that affect survival time effectively. Therefore, in the context of Multiple Myeloma patients who have varying survival rates, the Cox PH model is considered appropriate to test the influence of age, sex, and protein levels on survival time in a flexible and statistically valid manner.

2. Research Methods

2.1. *Data Description*

This study uses case data on the survival of Multiple Myeloma patients. The data used consisted of 47 patients with 5 variables, namely survival time, patient status (death and sensor), age, gender (male and female), and protein content.

Table 1. Multiple Myeloma Patient Survival Data by Survival Time

ID	Survival Time	Patient Status	Age	Gender	Protein
1	13	1	66	1	1
2	52	0	66	1	0
3	6	1	53	2	1
4	40	1	69	1	1
5	10	1	65	1	0
6	7	0	57	2	0
7	66	1	52	1	1
8	10	0	60	1	1
9	10	1	70	1	0
10	14	1	70	1	0
11	16	1	68	1	0
12	4	1	50	2	1
13	65	1	59	1	0
14	5	1	60	1	0
15	11	0	66	2	0
16	10	1	51	2	0
17	15	0	55	1	0
18	5	1	67	2	0
19	76	0	60	1	0
20	56	0	66	1	0
21	88	1	63	1	1
22	24	1	67	1	0
23	51	1	60	2	1
24	4	1	74	1	0
25	40	0	72	1	1
26	8	1	55	1	0
27	18	1	51	1	0

28	16	1	53	1	0
29	50	1	74	1	1
30	40	1	70	2	0
31	1	1	67	1	0
32	36	1	63	1	1
33	5	1	77	1	0
34	10	1	61	1	0
35	91	1	58	2	1
36	18	0	69	2	0
37	1	1	57	1	1
38	18	0	59	2	0
39	6	1	61	2	0
40	1	1	75	1	0
41	23	1	56	2	0
42	15	1	62	2	0
43	18	1	60	2	1
44	12	0	71	2	0
45	12	1	60	2	0
46	17	1	65	2	0
47	3	0	59	1	1

2.2. *Research Methods*

This study used a survival analysis approach to evaluate the influence of age, gender, and protein levels on the timing of events in patients. The data used included survival time, patient status (death/sensor), and independent variables.

2.2.1. Survival Function

The survival function expresses the chance that the individual or subject will still survive after a certain time t. Mathematically, this function is expressed as:

$S(t) = P(T \geq t) = 1 - F(t)$

Where T is the survival time and F(t) is the cumulative distribution function of T. In other words, the survival function is used to determine the probability of survival from the initial time of observation to a certain time t[11].

2.2.2. Metode Cox Proportional Hazards

The Cox Proportional Hazard method is one of the statistical models used to analyze the relationship between survival time and several variables that affect it. This method includes semi-parametric, assuming proportional hazard, i.e., the risk of occurrence between two different individuals has a constant ratio all the time[12]. The Cox Proportional Hazard method can be written as follows[13]:

$h_i(t) = exp\left(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots, \beta_p x_{pi}\right) h_o(t) = exp(\beta' x_i)\, h_0(t)$

With:

$h_i(t)$ = Function of Failure of the Individual I
$h_0(t)$ = basic failure function
β' = coefficient vector of the free variable X1, X2, ..., Xp.
x_i = value vector of independent variables X1, X2, ... Xp

2.2.3. Uji Likelihood

The Likelihood test is used to analyze whether the variables used in the model have a significant effect together[14].

Hipotesis:

$\mathbf{H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0}$

H1 : There is at least one $\beta k \neq 0$, with $k = 1, 2, ..., p$

Test Statistics : G2 = $\boldsymbol{-2 ln\Lambda}$

Critical areas : Subtract H0 if G2 is > or p-value < α.$\boldsymbol{X^2_{p,a}}$

2.2.4. Asumsi Cox Proportional Hazards

The assumption of Proportional Hazards is said to be fulfilled when the hazard ratio value is constant all the time, meaning that the difference in risk of one individual is proportional to that of another individual and is not affected by time. One of the methods used to test this assumption is Goodness of Fit with the Schoenfeld Residual approach[15]. Schoenfeld residual is a residual that each individual and each of its independent variables is based on the first derivative of the likelihood log function[16]. This test generates a test statistical value for each independent variable that is used to assess whether there is a violation of the proportional hazards assumption. If these values show a mismatch, then this model needs to be corrected or modified.

3. Discussion

3.1. *Kurva Survival Cox Proportional Hazards*

The following is a visualization of the survival curve based on the cox proportional Hazards model which illustrates the overall survival probability of Multiple Myeloma patients.

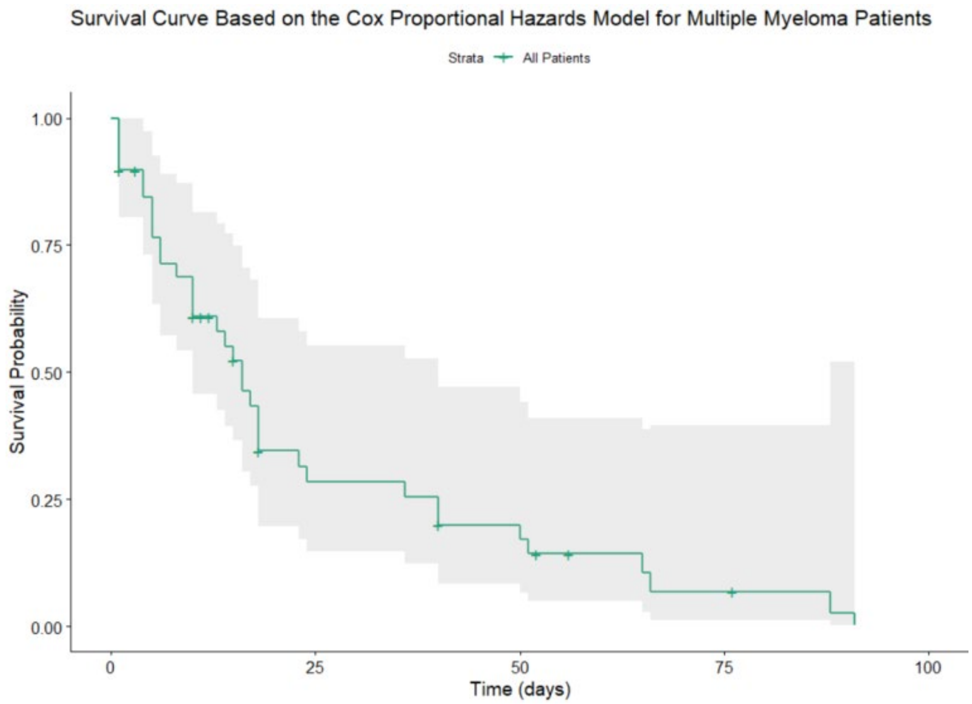

Figure 1. Kurva Survival Cox Proportional Hazards Pasien Multiple Myeloma Overall

Figure 1 shows the estimated chances of survival of Multiple Myeloma patients over time. The green line depicts the estimated chance of survival, while the gray area shows the confidence interval for the estimate. At the beginning of the observation time, the curve decreased sharply, indicating a high mortality rate at the beginning of treatment. Over time, the curve tends to flatten, reflecting a decrease in the number of patients who are still surviving towards the end of the observation period. This curve also shows that overall, there is no significant difference based on age, gender, and protein content variables.

3.2. *Cox Proportional Hazards*

The following are the results of the Cox Proportional Hazards model estimation used to analyze the influence of age, sex, and protein levels on the survival time of Multiple Myeloma patients.

Table 2. Hasil Estimasi Model Cox Proportional Hazards

Variabel	Coef	Exp (coef)	Se (coef)	With	*p*-value
Age	-0,001722	0,998280	0,028917	-0,060	0,953
Gender2	-0,298998	0,741561	0,461878	-0,647	0,517
Protein1	-0,494712	0,609746	0,381051	-1,298	0,194
Protein2	0,483654	1,621990	1,111079	0,435	0,663

Based on Table 2, it is known that the variable age has a hazard ratio of 0.998280, which means that statistically, the increase in age almost does not affect the risk of death, with a p-value of 0.953. For the gender variable, the gender coefficient showed that women had a lower risk of death than men, but this difference was not significant with a p-value of 0.517. Similarly, protein variables encoded as dummy variables also showed no significant effect on survival time with p-values of 0.194 and 0.663.

3.3. *Uji Likelihood*

To find out whether the overall Cox Proportional Hazards model is significant, a likelihood ratio test was performed. The test results are presented in Table 2 below.

Table 3. Likelihood Test Results

Test Statistics	Value	Free Degree (df)	*p*-value	Interpretasi
Likelihood Rasio Test	2,5	4	0,6	Insignificant

Based on Table 3, the results of the Likelihood Ratio test show a test value of 2.5 with 4 degrees of freedom, and a p-value of 0.6. Because the p-value is greater than 0.5, the model is not statistically significant. This means that this model has not been able to explain the variation in survival time of Multiple Myeloma patients well.

3.4. *Asumsi Cox Proportional Hazards*

Testing the assumption of proportional hazards is necessary to ensure that the influence of each variable on hazards is constant over time. This test was performed using the Schoenfeld residual method, and the results are shown in Table 3.

Table 4. Hasil Asumsi Cox Proportional Hazards

Variabel	Chi-square	Free Degree (df)	*p*-value	Interpretasi
Age	0,0649	1	0,80	PH assumption fulfilled
Gender	0,2845	1	0,59	PH assumption fulfilled
Protein	2,3133	1	0,31	PH assumption fulfilled

Based on Table 4, show the test results to see if the influence of each variable in the Cox Proportional Hazards model remains stable on survival time. These results indicate that all p-values are greater than 0.5, which meets the important assumption of the Cox Proportional Hazards model, making the model's results statistically reliable.

4. Discussion

This study aims to determine the influence of age, gender, and protein levels on the survival time of Multiple Myeloma patients using the Cox Proportional Hazards model. Based on the results of the analysis, the model technically meets the assumption of proportional hazards, which means it can be used to analyze survival data. However, no variables showed a significant influence on survival, either individually or overall. These results differ from expectations in the background, where all three variables are thought to affect patient survival. Possible causes may be a relatively small sample count (47 patients), an unbalanced distribution of categories, or the presence of other, more dominant clinical factors that have not yet been analyzed. This finding is also different from some previous studies, such as Puspita et al. (2022), which found that age has a significant effect on the recovery time of asthma patients. Therefore, further research with larger amounts of data and additional variables that are more representative is needed to produce more accurate and statistically robust models.

5. Conclusion

Based on the test results, the Cox Proportional Hazards model used meets the assumption of proportional hazards, so it is technically valid. However, the results of the model match test and the analysis of each variable showed that none of the variables had a significant effect on the patient's survival time, and the overall model was also not statistically significant. This is also reflected in the resulting survival curve, where there are no striking differences that can be attributed to age, gender, or protein content variables. Therefore, even though the model is technically valid, the model's predictive power is still weak and needs further development, either by adding more data and other more relevant predictive variables.

Reference

[1] A. Yusuf and M. Rizki, "Multiple Myeloma: Recognizing the Onset to Prognosis," *J. Med. Hutama*, vol. 4, no. 2, pp. 3265–3270, 2023.

[2] E. Norahmawati *et al.*, “Laporan Kasus : Multiple Myeloma Non Sekretorik Case Report : Non Secretory Multiple Myeloma,” *J. Medicine. Brawijaya*, vol. 28, no. 2, pp. 159–162, 2017.

[3] I. P. A. R. Githa and Anak Agung Davyn Anantha Rheja, "Multiple Myeloma in Patients with Pathological Fractures: Case Reports," *Essence of Medical Science*, vol. 13, no. 3, pp. 651–657, 2022, doi: 10.15562/ism.v13i3.1521.

[4] J. Amalia, "Analysis of Cox Hazard Regression Model in a Case Study of Lung Cancer Patients," *J. Ilm. Simantek*, vol. 4, no. 1, pp. 9–25, 2020.

[5] S. W. Purnami and I. N. Pertiwi, "Cox Proportional Hazard Regression for Survival Analysis of Brain Cancer Patients at C-Tech Labs Edwar Technology Tangerang," *Inference*, vol. 3, no. 2, p. 65, 2020, doi: 10.12962/j27213862.v3i2.7727.

[6] M. Nainggolan and S. Manullang, "Survival Analysis of Coronary Heart Disease Patients with Breslow and Exact Methods on the Cox Proportional Hazard Regression Model," vol. 4, pp. 2350–2360, 2024.

[7] A. S. Khinanti, S. Sudarno, and T. Wuryandari, "Cox Hazard Regression Model in Hemodialysis Patient Survival Data," *J. Gaussian*, vol. 10, no. 2, pp. 303–314, 2021, doi: 10.14710/j.gauss.v10i2.30958.

[8] N. N. R. PUSPITA, M. SUSILAWATI, and N. L. P. SUCIPTAWATI, "Application of Cox Proportional Hazard in Asthma Patients," *E-Journal Mat.*, vol. 11, no. 1, p. 53, 2022, doi: 10.24843/mtk.2022.v11.i01.p360.

[9] E. D. Aditya, S. S. Handajani, and R. Setiyowati, "Test of Proportional Hazard Assumptions on Factors Affecting the Survival Time of Lung Cancer Patients," *Indones. J. Appl. Stat.*, vol. 1, no. 2, p. 126, 2019, doi: 10.13057/ijas.v1i2.26496.

[10] R. T. Saragih, T. Wuryandari, D. Fakhriyana, D. Statistika, and U. Diponegoro, "Analysis of cox proportional hazard regression model on survival data of hemodialysis patients using the breslow method 1,2,3," vol. 13, pp. 443–449, 2024, doi: 10.14710/j.gauss.13.2.443-449.

[11] A. Mutiah, J. Purwadi, and H. A. Mutiah, "Survival Analysis with a Multivariate Adaptive Regression Splines Approach in Cases of Survival Analysis with a Multivariate Adaptive Regression Slines Approach in Cases of Chronic Kidney Disease," pp. D318–D323, 2012.

[12] M. Reza Pahlevi and T. Wuryandari, "Stratified Cox Regression Model on Resilience Data," *J. Gaussian*, vol. 5, no. 3, pp. 455–464, 2016.

[13] E. Setiani, S. Sudarno, and R. Santoso, "COMPARISON OF COX PROPORTIONAL HAZARD REGRESSION MODEL USING BRESLOW AND EFRON METHODS (Case Study: Stroke Sufferers at Tugurejo Hospital, Semarang City)," *J. Gaussian*, vol. 8, no. 1, pp. 93–105, 2019, doi: 10.14710/j.gauss.v8i1.26624.

[14] R. Y. Fa'rifah and Purhadi, "Survival Analysis of Factors Affecting the Recovery Rate of Patients with Dengue Hemorrhagic Fever (DHF) at Haji Hospital Surabaya with Cox Regression," *J. Sains dan Seni ITS*, vol. 1, no. 1, pp. 271–276, 2012.

[15] K. Pasien, T. Lung, Y. Septi, and A. Monica, "Analysis of Factors Affecting the Rate of Weibull Regression and Proportional Cox Regression," vol. 5, no. 2, pp. 450–455, 2016.

[16] N. Zukhaila, "Analysis of the Cox Proportional Hazard Regression Equation Model in the Nutritional Status of Toddlers to Determine the Factors Causing Malnutrition," 2016.

[17] A. N. I. Pradana and Ayunin Sofro, "Application of Survival Analysis in Multiple Myeloma Patients," *J. Ilm. Matt.*, vol. 3, no. 6, pp. 80–86, 2019.