INDONESIAN COUNCIL OF PREMIER STATISTICAL SCIENCE

Indonesian Council of Premier Statistical Science
ISSN: 3030-9956/https://dx.doi.org/10.24014/icopss.v4i2.37879

Volume 4 Issue 2, 68 – 73, August 2025

Analysis of Pneumonia Patients Using Kaplan-Meier and Cox Proportional Hazard Regression

Aulia Azira Putri[1], Dian Desti Ananda[2]

[1,2]*Prodi Matematika, Faculty of Sciences and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia*

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

Abstract - *Pneumonia is one of the leading infectious diseases in the world, especially in the elderly and patients with complications. This study aims to analyze the survival of pneumonia patients treated in the ICU room of Hospital X by 2024 using the Kaplan-Meier method and Cox Proportional Hazard regression. The data used consisted of 37 patients, with independent variables in the form of age and type of pneumonia (mildness, sepsis, and other complications), and dependent variables in the form of patient survival time. The Kaplan-Meier analysis showed that patients with >60 years of age and complicated types of pneumonia tended to have lower survival rates. However, the results of the Log-Rank test showed that there was no statistically significant difference between age groups and types of pneumonia on survival ($p > 0.05$). The results of the Cox regression also showed that no variables had a significant effect on the patient's risk of death, although types of pneumonia with other complications showed a significant tendency at the level of 10% ($p = 0.0699$). This study demonstrates the importance of considering disease severity in the evaluation of the risk of death of pneumonia patients, as well as the need for further research with larger sample sizes and additional clinical variables.*

Keywords: *Pneumonia, Survival, Kaplan-Meier, Cox Regression, ICU*

1. Introduction

Mortality from pneumonia continues to show an alarming trend, particularly in the elderly and patients with comorbidities [1]. Pneumonia ranks fourth as the leading cause of death in the world and is the deadliest infectious disease [2]. Deaths caused by pneunomia in 2000 reached 1.8 million [3]. Pneumonia is an infectious disease that attacks the lower respiratory tract and is characterized by symptoms such as coughing and shortness of breath. Pneumonia affects the lungs, where the lung alveoli are filled with fluid, making oxygen intake limited for breathing. Pneumonia is one of the leading causes of morbidity and mortality worldwide, especially in adult patients with community-acquired pneumonia (CAP) [4]. In addition to having a high mortality rate, pneumonia also requires a relatively long period of hospital treatment, with a minimum duration of four days. Cases of pneumonia that occur in the community tend to increase in line with age.

Increased age is associated with decreased immune function, which leads to a higher risk of infection and complications. Elderly patients are more prone to pneumonia due to decreased immunity, impaired rendiculous response, and decreased mucociliary function [5]. On the other hand, the presence of comorbidities also worsens the patient's condition, affects survival time, and response to therapy. Therefore, it is important to evaluate the factors predicting the survival of pneumonia patients.

Research on predictive factors of pneumonia was conducted in 2015. The method used in the study is the bivariate analysis method using the chi-square test. The factors used in this study were ten variables, namely age group, decreased consciousness, comorbidity (Charlson Comorbidity Index / CCI score >5), sepsis, respiratory failure, severe pneumonia, hemoglobin level 20000/µl, albumin level 200 mg/dl. The results of this study stated that the predictive factors for mortality in adult CAP patients hospitalized at RSCM in the period 2010-2014 were severe pneumonia, sepsis, respiratory failure, comorbidities (CCI score >5), and hypoalbuminemia (albumin levels), while other variables such as age groups were not shown to increase the risk of death [6]. The next study in 2018 aimed to find out the risk factors that affect the survival of pneunomia patients, the variables used were gender, weight, symptoms, and age. The method used was the Cox Regression survival analysis. The Weibull regression model shows that the survival time of pneumonia patients is influenced by male sex and the use of drugs other than cefixime, cefotaxime, ampicillin, Nucef, and Farbivent [7].

In addition, a study in 2019 found the relationship between the elderly and the incidence of community pneumonia at the NTB

Provincial Hospital in 2019. The method used in this study is cross sectional with the consecutive sampling method. This study states that there is an association between the elderly and the incidence of community pneumonia with a significant value of 0.035 [8]. The next research is the relationship between sex and immunization status to the incidence of pneumonia in toddlers at dr. Soedarso Hospital, West Kalimantan Province which was conducted in 2021 using a case control design. The results obtained that there was no significant relationship between sex and immunization status against the incidence of pneumonia in toddlers [9]. A study that discusses the factors that affect hospital mortality in pneumonia was also conducted in 2005. The method used in this study is univariate logistics regression analysis. The results of this study are factors that are independently related to mortality among pneunomia patients, namely increased severity scores, increased age, functional status at the time of hospital admission, and the need for consultation with a respiratory disease specialist [10].

Based on research[6] and research [7] that examined the factors that affect the survival of pneunomia patients as well as the phenomena related to [4], the authors are interested in conducting a similar analysis on the data of pneumonia patients treated in the ICU room of Hospital X to see how survival and what factors affect the survival of pneunomia patients over time. One of the methods often used in nonparametric survival analysis is Kaplan-Meier analysis [11]. However, referring to a study [12] which states that the use of Kaplan-Meir is not sufficient in complex situations. Therefore, the authors combine the Kaplan-Meier approach with Cox Proportional Hazard regression to provide a more comprehensive analysis of the survival of pneumonia patients.

.

2. Research Methods

2.1. *Data Description*

The data used in this study is ICU patient census data obtained from Hospital X during the period January to December 2024. The details of the data can be seen in Table 1 below:

Table 1. Survival Data of Pneumonia Patients

Patient	Age	Category	Time (days)	Statusb	Patient	Age	Category	Time (days)	Statusb
1	33	3	2	1	20	41	1	2	1
2	63	3	2	1	21	59	1	3	1
3	40	3	1	1	22	46	1	4	1
4	55	2	3	1	23	75	3	12	1
5	58	3	2	1	24	51	3	2	1
6	69	2	2	1	25	49	3	1	1
7	56	3	8	1	26	56	3	1	1
8	65	2	1	1	27	61	2	3	1
9	27	2	5	1	28	56	2	3	1
10	69	2	8	1	29	51	2	14	1
11	63	1	8	1	30	51	2	8	1
12	68	1	7	1	31	63	2	10	0
13	78	1	3	1	32	57	2	4	1
14	67	1	4	1	33	50	2	7	1
15	62	1	1	1	34	71	2	1	1
16	44	1	16	1	35	65	1	3	1
17	46	1	1	1	36	59	1	2	1
18	61	1	30	1	37	63	2	9	1
19	58	1	14	1					

Information:

a1, mild pneumonia; 2, pneumonia, sepsis; 3, pneumonia other complications

b1, died; 2, Sensor

2.2. Research Methods

This study is quantitative research with a survival analysis approach to analyze the survival of pneumonia patients. The analysis was carried out using the Kaplan-Meier method to visualize survival functions, as well as *the Log-Rank* test to test the differences in these functions between groups. Furthermore, *a Cox Proportional Hazard* regression analysis was performed to determine the influence of age and severity of pneumonia on the patient's survival time. All analysis was carried out with the help of R software.

2.2.1 Kaplan-Meier Estimation and Log-Rank Test

Kaplan-Meier is a non-parametric method used to estimate the survival function based on data on the time until an event occurs, including censored data. The estimated results are displayed in the form of a survival curve, which shows the probability of survival at certain times. The value of the Kaplan-Meier curve at time *t* is denoted by *S(t)* with the following equation [13]:

$$S(t) = S_{t-1} \times \left(1 - \frac{d_i}{n_i}\right) \quad (1)$$

where:

d_i : The number of individuals who experienced the event at the *time of the*

n_i : The number of individuals who still survived at the *time*

Comparison between two or more Kaplan-Meier curves is performed by the test *Log-Rank*, which is a non-parametric hypothesis test that aims to find out if there are significant differences in the survival function between groups [14]. Here are the hypotheses in the test *Log-Rank*:

H_0 : There was no difference in the Kaplan-Meier survival curve between different groups.

H_1 : There was a difference in the Kaplan-Meier survival curve in at least one group.

This test uses Chi-Square statistics by comparing the observation and expectation values of each group in a time interval. Test formula *Log-Rank* for the two groups are [15]:

$$\chi^2 = \frac{(O_i - E_i)^2}{Var\ (O_i - E_i)} \quad (2)$$

where:

O_i : Individual observation value of group *I*

E_i : the value of individual expectations of the first group

The H0 hypothesis will be rejected if p-value < $\alpha = 0{,}05$.

2.2.2 Regresi Cox Proportional Hazard

Back *Cox Proportional Hazard* is a survival data analysis method used to find models and examine the influence of free variables on hazard functions. Mathematically, this model can be expressed as follows [16]:

$$h_i(t) = h_0(t)\exp\left(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\right) \quad (3)$$

where:

$h_0(t)$: Fungsi hazard baseline

$\beta_1 + \beta_2 + \cdots + \beta_p$: Parameter regresi

$x_{i1} + x_{i2} + \cdots + x_{ip}$: Variable value of the i individual

In survival analysis, ties occur when several individuals experience an event at the same time. This event will cause parameter estimation problems because the determination of the risk set becomes unclear. One alternative method can be used is the Breslow approach, which assumes that the size of the risk set at the time of the joint event is equal [17]. The equation of the Breslow approach is as follows:

$$L(\beta)_{breslow} = \prod_{i=1}^{r} \frac{exp(\sum_{j=1}^{p} \beta_j S_k)}{(\sum_{i \in R(ti)} exp(\sum_{j=1}^{p} \beta_j X_{1j}))^{d_i}} \quad (4)$$

3. Analysis

3.1 *Kurva Survival Kaplan-Meier*

To analyze the influence of each partial free-of-life variable, the Kaplan-Meier curve can be used as a visualization and Log-Rank test to test the statistical significance of **differences.**

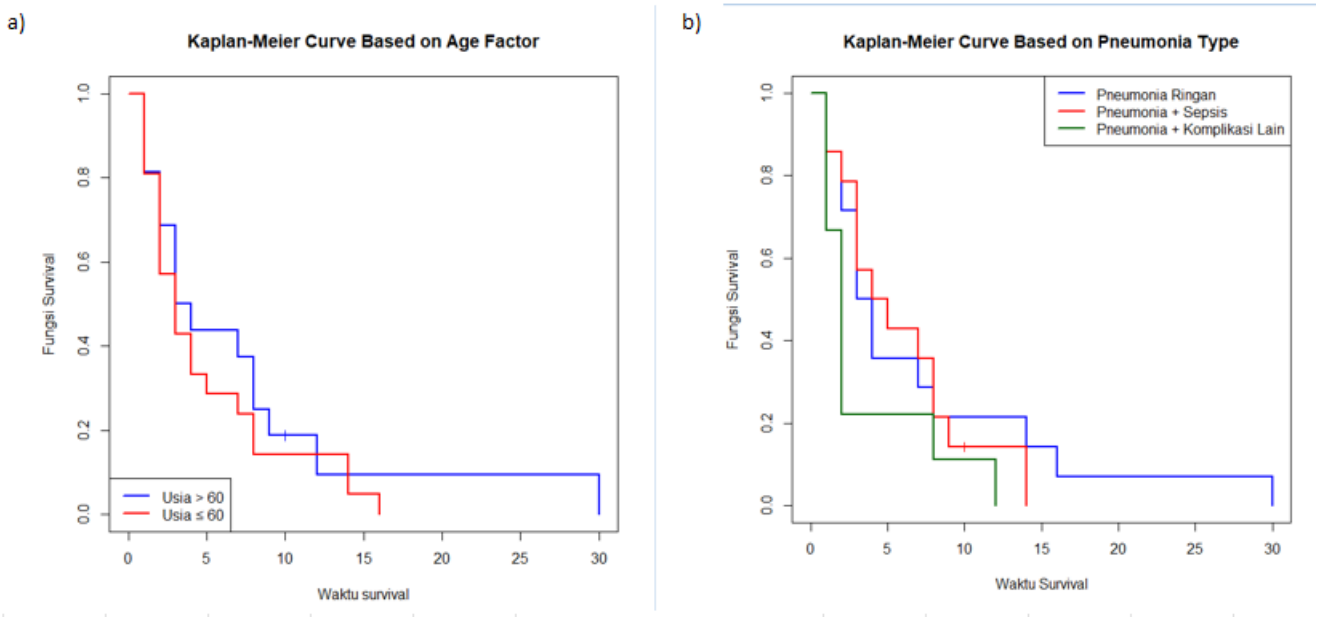

Figure 1. Kaplan-Meier Curve Based on Factors: (a) Age, (b) Type of Pneumonia

Based on Figure 1a, it shows that patients with > age of 60 years have a lower survival rate compared to patients ≤ 60 years of age. A faster decline in survival function occurs in older age groups, indicating that old age is a risk factor contributing to decreased survival in pneumonia patients.

Furthermore, Figure b shows that patients with mild pneumonia have a better survival rate compared to patients with pneumonia with complications of sepsis and other complications. The curve showed the fastest decline in survival function in the group of pneumonia with other complications, followed by pneumonia with sepsis, indicating that the severity and presence of complications had a significant effect on patient survival.

Table 2. Log-Rank Test Results for Each Variable

Variabel	*Log-Rank*	df	p-value	Results
Age Factor	0,6	1	0,4	Receive H0
Types of Pneumonia Factors	3,4	2	0,2	Receive H0

Based on the results of the Log-Rank test in Table 2, it was seen that the p-value was greater than the significance level of 0.05, so it was concluded that there was no significant difference in the chances of survival between groups in each variable.

3.2 *Regresi Cox Proportional Hazard*

In this analysis, it is interesting to know whether the variables of age and type of pneumonia affect the survival time of pneumonia patients.

Table 3. Cox Regression Model Estimation Results

Variabel	Coef	Exp(coef)	Se(coef)	With	p-value	Results
Age	-0,01148	0,98858	0,01568	-0,732	0,4642	Receive H0
Pneumonia + Sepsis	0,08886	1,09293	0,41032	0,217	0,8386	Receive H0
Pneumonia + Other Complications	0,83152	2,29680	0,45878	1,812	0,0699	Receive H0

Based on Table 3, the age variable has a p-value of 0.4642, so it does not have a significant effect on the patient's survival time. The same thing can also be seen in the category of pneumonia with sepsis which has a p value of 0.8386. Meanwhile, the category of pneumonia with other complications had a pvalue of 0.0699, which is close to the 10% significance limit. Although not statistically significant, the Exp(coef) value of 2.29680 indicates that the hazard ratio (HR) for patients with other complications is about 2.3 times greater than for patients with mild pneumonia. This means that there is a tendency to increase the risk of death in this group.

Table 4. Estimated Baseline Hazard

Hazard	Time
0,2948807	1
0,6997683	2
1,1881844	3
1,4886094	4
1,600659	5
1,8555841	7
2,533783	8
2,7688921	9
2,7688921	10
3,0820841	12

Table 4 shows the *baseline hazard estimate* which represents the baseline risk of death without considering the covariate variable. Hazard values tend to increase over time, suggesting that the longer the patient is treated, the risk of death also increases. A significant increase is seen in the 8th to 12th time, which can indicate a critical phase in the treatment period of pneumonia patients.

4. Discussion

The results of the study showed that descriptively, there were differences in the survival rate of pneumonia patients based on age and type of pneumonia. From the Kaplan-Meier visualization, it can be seen that patients aged >60 years tend to have a lower survival rate compared to patients aged ≤60 years. This is in line with previous findings that suggest that old age is a major risk factor in pneumonia cases, due to decreased immune function and other physiological conditions in the elderly [5].

However, based on *the Log-Rank* test, no statistically significant difference in survival by age or type of pneumonia (p-value > 0.05) was found. This indicates that although there is a graphically different tendency for survival, the available data are not strong enough to conclude that there are real differences between groups. The limited number of samples and the possibility of heterogeneity in the group can be the cause of the insignificance of this test result.

Follow-up analysis with Cox regression showed that no variables had a significant effect on the risk of death of pneumonia patients at a significance level of 5%. However, the type of pneumonia variable with other complications showed a p value of 0.0699, close to the significance threshold of 10%, with a *hazard ratio* (HR) value of 2.297. This suggests a tendency that patients with complications other than sepsis have a higher risk of death compared to patients with mild pneumonia.

These results are consistent with previous research [6], which states that disease severity and the presence of complications are strong predictors of mortality in pneumonia patients. However, the insignificance of age in this model differs from some previous studies [10], which stated age as a consistent factor influencing the mortality of pneumonia patients. This is likely due to the limited number and distribution of age data in this study sample. However, these results are in line with research [6], which states that the age variable has not been shown to increase the risk of death.

In general, these results reinforce the importance of a combined approach between Kaplan-Meier analysis and Cox regression to obtain a more comprehensive picture of survival, although the final results still indicate the need for additional data to strengthen the conclusions. Further research is suggested to involve more clinical variables and larger sample sizes to obtain more statistically and clinically valid results.

5. Conclusion

This study aims to analyze the survival of pneumonia patients using the Kaplan-Meier method and *Cox Proportional Hazard* regression based on age factors and type of pneumonia. Based on the Kaplan-Meier analysis, there were visual differences in survival function between age groups and types of pneumonia, where elderly patients and patients with complications showed lower survival rates. However, the results of *the Log-Rank* test and the Cox regression model showed that neither age nor type of pneumonia had a statistically significant effect on the survival time of pneumonia patients at a significance level of 5%. However, the pneumonia variable with other complications showed a tendency to have an influence on the increased risk of death, with a hazard ratio about 2.3 times higher than in patients with mild pneumonia.

From these results, it can be concluded that the severity of the disease tends to have a greater effect on the patient's survival than age, although it is not statistically significant. Further research with larger sample counts and the addition of other clinical variables is needed to obtain more accurate and representative results.

Reference

[1] R. Lahmudin Abdjul, S. Herlina, P. Diploma Study of Triple Nursing, and F. Health Sciences, "Nursing Care in Adult Patients with Pneumonia: Case Study," 2020.

[2] F. P. Rima, H. Santosa, L. D. Arde, and I. K. Lubis, "Analysis of the survival of pneumonia patients with cox regression model at Aceh Tamiang Hospital in 2022," *Tropical Public Health Journal*, vol. 4, no. 2, pp. 59–68, Sep. 2024.

[3] A. Rafsanjani *et al.*, "Analysis of Body Temperature Survival in Pneumonia Patients Given Vitamin D."

[4] K. L. Buising *et al.*, "A prospective comparison of severity scores for identifying patients with severe community acquired pneumonia: Reconsidering what is meant by severe pneumonia," *Thorax*, vol. 61, no. 5, pp. 419–424, May 2006.

[5] B. A. Cunha, "REVIE W Pneumonia in the elderly."

[6] M. A. Firmansyah, Z. Amin, T. Loho, and H. Shatri, "Predictive Factors of Community-Acquired Pneumonia Mortality in Inpatient Care at Cipto Mangunkusumo Hospital, Jakarta."

[7] A. Husain Abdullah *et al.*, "Survival Analysis to Estimate the Survival Level of Toddlers with Pneumonia".

[8] "The Relationship of the Elderly with the Incidence of Community Pneumonia at the NTB Provincial Hospital in 2019."

[9] Z. Q. A'yuni, L. S. S. Mamesah, and I. A. Marhana, "Gender Factors and Immunization Status on the Incidence of Pneumonia in Toddlers at dr. Soedarso," *Journal of Smart Midwives*, vol. 4, no. 4, pp. 224–231, Dec. 2022.

[10] T. J. Marrie and L. Wu, "Factors Influencing In-hospital Mortality in Community-Acquired Pneumonia A Prospective Study of Patients Not Initially Admitted to the ICU Clinical Investigations," 2005.

[11] A. Suhartini, R. Rahmwahidinawati, D. Statistics, and F. Science and Mathematics, "Kaplan-Meier Survival Curve Analysis Using Test *Log-Rank* (Case Study :P Coronary Heart Disease at Undata Palu Hospital)," vol. 7, no. 1, pp. 33–42, 2018.

[12] B. N. A Wiyoto and and H. Sumarno, "Competing Risk Analysis for Pneumonia Patients in a Hospital."

[13] Basiludin, Bimasuci, "Analysis of Factors Affecting the Length of Time to Find a Job Using the Kaplan-Meier Curve, Test *Log-Rank*, and regression *Cox Proportional Hazard* ", *Final Project, Islamic University of Indonesia*, 2023.

[14] E. A. Maulida, "Analysis of Survival of Breast Cancer Patients Using Regression *Cox Proportional Hazard* and the Kaplan-Meier method", *Thesis, Sunan Ampel State Islamic University,* 2019.

[15] F. A. Rahmanika, S. W. Purnami, and N. A. Kusumawardani, "Analysis of the Kaplan-Meier Survival Curve in HIV/AIDS Patients with Antiretroviral Therapy (ART) at Prof. Dr. Soekandar Hospital, Mojokerto Regency Using the Test *Log-Rank*", *ITS Journal of Science and Arts*, vol. 5, no. 1, pp. 127–132, 2016.

[16] Sulantari and W. Hariadi, "Survival Analysis of the Cox Regression Model on the Length of Recovery Time of Patients with Moderate Covid-19 Symptoms," *Unisda Journal of Mathematics and Computer Science*, vol. 8, no. 1, pp. 43–54, 2022.

[17] A. S. Khinanti, S. Sudarno, and T. Wuryandari, "Regression Model" *Cox Proportional Hazard* In the Survival Data of Hemodialysis Patients," *Jurnal. Gaussian*, vol. 10, no. 2, pp. 303–314, 2021.