

Survival Analysis to Analyze Factor that Affect the Rate of Recovery Patients of Pneumonia

Sugiyarto, Umami Aulia Septiani, Suliadi Firdaus Sufahani

Abstract: This study aims to apply survival analysis for the survival of patients diagnosed with Pneumonia and to identify the factor that affects the rate of recovery of Pneumonia patients in the treatment period. The data used in this study is the patients' initial time when they were entering the hospital until they were experiencing an event that is when the patients have been declared cured of the disease. This study is an applied research with a causal-comparative approach by observing data from the factor that was suspected as the cause, as a comparison to investigate the possibility of causation. After that, an analysis was done by using Cox regression model in the case and its application to know the factor that can affect the level of healing Pneumonia patients at PKU Muhammadiyah Bantul Hospital; as the duration of therapy in the hospital is the time of survival. In accordance with the Anderson Darling distribution test, the result of the patient's survival time distribution test is a normal distribution with Box-Cox Transformation. The result shows that the respiration rate affects the recovery rate of Pneumonia patients with the coefficient value is -0.04082 .

Index Terms: About four key words or phrases in alphabetical order, separated by commas.

I. INTRODUCTION

In the domain of mathematics, there are inventions of analysis tools that can be used to analyze a problem, known as the branch of statistics. One of them is a life test which is a survival study of a unit or individual in a particular situation. Life tests are commonly used in engineering, biology, medicine, and others. Such studies normally use the data related to the lifetime of an individual. Life test is a research for the survival of a unit or component under certain operational circumstances. This analysis is called as The Lifetime Test Analysis and generally used in industry and health.

The method used to analyze lifetime data is called survival analysis. Survival analysis is a statistical technique which can be used to determine the results of variables that affect an event from beginning to the end of the event such as time

recorded in days, weeks, months, or years. For the initial event, suppose the initial patient diagnosed with a disease and for the final event, suppose the patient's death and the patient's healing [5]. The survival analysis aims to model the underlying distribution of failure time variables and to estimate the dependence of variable failure time with independent variables [4]. In life test analysis, it usually uses the time variable as the survival time, as it indicates that a person has survived for a certain period and typically shows the event as a failure since the event of concern is usually death, illness, or some other event [4].

The previous research, survival analysis with the Cox regression model was done by Riska, *et al.* in 2012 [9] which examined the factors that affect the healing rate of dengue fever patients. Also, there is a tuberculosis case that was done by Pardeshi in 2009 [8] which investigated the survival of tuberculosis patients in India and Indra Maulana *et al.* in 2012 which studied the rate of cure of pulmonary tuberculosis in Jakarta. The thing that distinguishes this research from the previous is the method used is adjusted with the research data and special events that are patient recovered, different research places and variables used differently. The purpose of this method is to model the distribution and acknowledge the factors that affect the recovery rate of patients with pneumonia.

II. MATERIAL AND METHOD

Survival analysis is an analysis of data related to the time, from the beginning until the occurrence of a special event. The duration from the beginning of observation on an individual (time origin) until the occurrence of a special event (endpoint or failure event) is called as the time of survival. There are two main functions in survival analysis, known as survival function and hazard function. The definition of survival function is the probability of an individual can survive until the t time and hazard function is the probability that individual dies within the t time to $t + \Delta t$ interval. If it is known that the individual can still survive until t time or in other words, the hazard function represents an opportunity an object (patient) experiences an event in a moment, hence the hazard function is the reverse of the survival function. Estimation of this case (Pneumonia patients) data distribution is the survival time data. The test may use the Anderson-Darling test to determine the most appropriate distribution of the data.

Revised Manuscript Received on 30 May 2019.

* Correspondence Author

Sugiyarto*, Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Ahmad Dahlan University, South Ringroad Street, Tamanan, Banguntapan, Bantul, Yogyakarta, Indonesia.

Umami Aulia Septiani, Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Ahmad Dahlan University, South Ringroad Street, Tamanan, Banguntapan, Bantul, Yogyakarta, Indonesia.

Suliadi Firdaus Sufahani, Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Educational Hub, 84600 Pagoh, Johor, Malaysia.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Survival Analysis to Analyze Factor that Affect the Rate of Recovery Patients of Pneumonia

The survival function is as follows:

$$S(t) = 1 - F(t) \quad (1)$$

and the hazard function equation is expressed as follows:

$$h(t) = \frac{f(t)}{S(t)} \quad (2)$$

The variables that will be studied are as follows:

- 1) The dependent variable, denoted by Y is the survival time from treatment to the end of the study (within a few days).
- 2) The independent variable, denoted by X consisting of twelve independent variables.
 - a) X1 (Score) = Condition of patient at baseline admission (Karnofsky score)
 - b) X2 (Gender)
 - c) X3 (Age) = Patient's age at the beginning of the study (within a few months)
 - d) X4 (Weight)
 - e) X5 (Therapy) = Therapeutic method given to the patient
 - f) X6 (Temperature) = The initial temperature that indicated from the patient
 - g) X7 (Fever) = Clinical symptoms suffered
 - h) X8 (Cough) = Clinical symptoms suffered
 - i) X9 (Shortness) = Clinical symptoms suffered
 - j) X10 (retraction) = Clinical symptoms suffered
 - k) X11 (Respiration Rate/RR) = Average respiration of patient.
 - l) X12 (Antibiotics) = An antibiotic drug given to the patient

The method used in this research is Cox regression with the distribution model used obtained from research data that is the normal distribution with Box-Cox transformation. The following steps of this research are:

1) Problem Identification

The problem identification begins with a literature study to review the relevant literature sources used to collect the information that are required in this study.

2) Problem Formulation

The formulation of this problem is the limitation of the problem to be studied in this research to be relevant and in accordance with existing research that can be formulated is the average lifetime of patients with pneumonia, survival function and function of failure or hazard. functions that are formed from the survival of patient data, chance of failure and survival opportunities.

3) Data Collection

a) Observation Method

This method is done by collecting data from the patient's medical records. The required data are quantitative data—i.e. data obtained from the measurement results. The following data are patient age, Karnofsky score (100 - 0), the day of initial diagnosis, clinical signs (fever, cough, retraction, and shortness of breath), treatment methods performed by patients, length of patient survival.

b) Literature Method

This literature method is used to select and analyse various reading sources associated with survival analysis.

4) Analysis

Firstly, we performed a descriptive analysis of the patient's pneumonia history. After that, survival analysis was done to determine the factors that influenced the recovery rate of patients with pneumonia. After established a Cox regression model to see the effect of explanatory variables on the dependent variable, the best model is selected and test the model parameters that influence. The proportional assumptions of the survival functions and the coefficient of Cox model using hazard-ratio, and Cox model was applied in assuming the factors that affect the healing rate of Pneumonia patients.

5) Conclusions

Based on the analysis data, we can conclude the variables that affect the dependent variable conclusion of the best model obtained.

III. RESULTS

A. Estimation of Box-Cox Transformation Parameters

If the survival time data follows the normal distribution with Box-Cox transformation, the function $f(t)$ is an opportunity density function of the Box-Cox transformation distribution. The general type of the Box-Cox transformation density function with parameters (λ, μ, σ) is:

$$\text{for } t \geq 0, \mu > 0, \sigma > 0$$

$$f(t|\lambda, \mu, \sigma) = t^{\lambda-1} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{t^\lambda - \mu}{\sigma}\right)^2\right\} \quad (3)$$

The cumulative function of the Box-Cox transformation is as follows:

$$F(t|\lambda, \mu, \sigma) = \Phi\left(\frac{\frac{t^\lambda - 1}{\lambda} - \mu}{\sigma}\right) \quad (4)$$

The survival function which is the probability of survival in time t is defined as follows:

$$S(t|\lambda, \mu, \sigma) = 1 - \Phi\left(\frac{\frac{t^\lambda - 1}{\lambda} - \mu}{\sigma}\right) \quad (5)$$

Hazard function can be formulated as follows:

$$h(t|\lambda, \mu, \sigma) = \frac{t^{\lambda-1} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{t^\lambda - \mu}{\sigma}\right)^2\right\}}{1 - \Phi\left(\frac{t^\lambda - \mu}{\sigma}\right)} \quad (6)$$

Hence the hazard function for *i* patient can be expressed as follows:

$$h_i(t, X) = \frac{t^{\lambda-1} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{t^\lambda - \mu}{\sigma}\right)^2\right\}}{1 - \Phi\left(\frac{t^\lambda - \mu}{\sigma}\right)} \exp(\beta'X_i) \quad (7)$$

Thus, the survival function for *i* patient is:

$$S_i(t, X) = \exp\left[\ln\left(1 - \Phi\left(\frac{t^\lambda - \mu}{\sigma}\right)\right)\right] \exp(\beta'X_i) \quad (8)$$

B. Survival Time

Table 1 Mean and median survival time.

Min.	Median	Mean	Max.
3.00	5.00	4.98	10.00

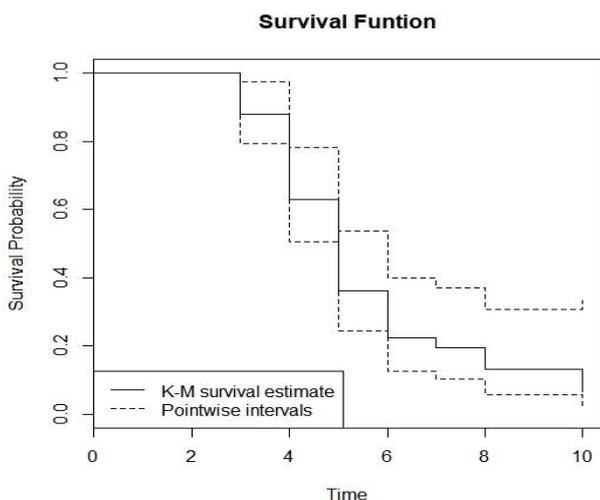


Fig. 1 Survival time.

Based on Table 1 and Fig. 1, it can be seen that on day 0 to day 2, the survival curve still remains with chances 1. This means the chances of the patient not experiencing major

clinical improvement. On day 2 to day 4, the survival curve decreases rapidly with a chance of 0.6 to 1. This means the chances of patients not undergoing clinical improvement decreases. On the 4th day until the 6th, the survival curve decreased rapidly with a chance of 0.2 to 0.6. Therefore, the chances of patients not experiencing clinical improvement are decreasing, with the result that some patients may be discharged from the hospital. On the 6th day until the 8th, the survival curve decreased slowly with a chance of 0.1 to 0.2. These data interpreting the chances of patients not experiencing clinical improvement are descending. On day 8 to 10, the survival curve decreased slowly with a 0.1 chance. These data indicate that many patients can get out of the hospital. On day 10, the survival curve decreased slowly with chances under 0.1. This means the patient has undergone many clinical improvements after the 10th day and many patients have been discharged from the hospital. The survival time has mean 5.55 days and median 5 days.

C. Cox Regression Modeling

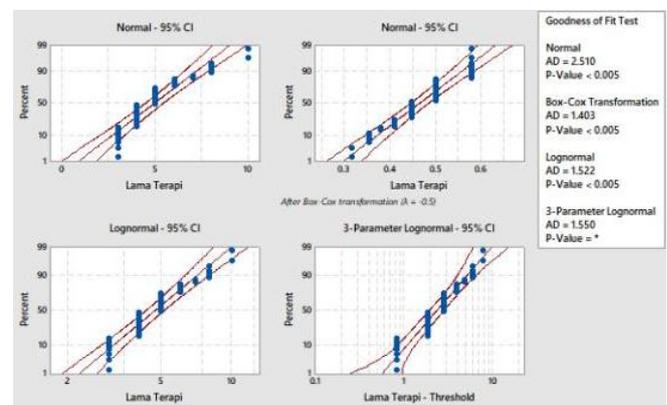


Fig. 2 Distribution plot for the duration of therapy.

Based on the Distribution ID Plot in Fig. 2 shows the Anderson-Darling value and p-value obtained for the duration of therapy. In determining the data following a particular distribution, we can see the Anderson-Darling value comparison of the tested distributions. The data follow an appropriate distribution if the distribution has the smallest Anderson-Darling value. It can be concluded from Table I that the data follows the Normal distribution with the Box-Cox transformation with the Anderson-Darling value 1.365 which is the smallest Anderson-Darling value.

The Cox model that was formed based on patient survival time data from influential variable is as follows:

$$\hat{h}_i(t, X) = h_0(t) \exp(\hat{\beta}X_i) \quad (9)$$

Estimated cox model in (10). After the partial likelihood ratio test then the best model and parametric estimation is shown in Table 2 that the significant estimate is the respiration rate variable. It can be said that the factors that affect the patient's healing rate is the variable and based on the analysis of the decision test the significance of the parameters can be concluded that the variable affects the individual to the time of survival.



Survival Analysis to Analyze Factor that Affect the Rate of Recovery Patients of Pneumonia

$$\hat{h}_i(t, X) = \frac{\frac{1}{2.51} \frac{1}{t^{1.5}} \exp\left\{-\frac{1}{2}\left(\frac{1}{1.73770t^{0.5}} - 2.88\right)^2\right\}}{1 - \int_0^t \frac{1}{2.51} \frac{1}{t^{1.5}} \exp\left\{-\frac{1}{2}\left(\frac{1}{1.73770t^{0.5}} - 2.88\right)^2\right\} dt} \exp(-0.04082RR) \quad (10)$$

Table 2 Best model.

Variable	df	Coefficient	P-Value
RR	1	-0.04082	0.00936

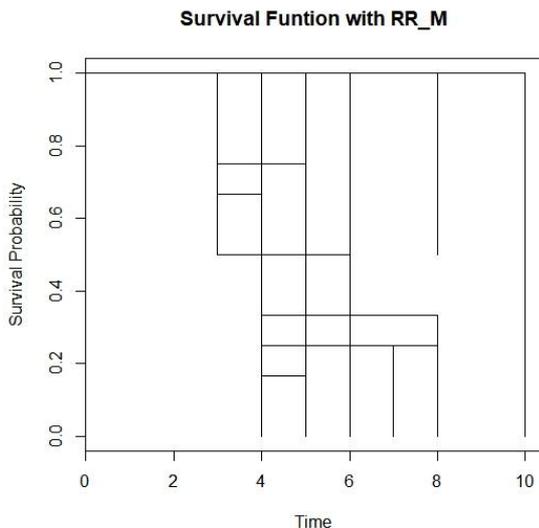


Fig. 3 Survival function plot with respiration rate.

According to Fig. 3, it is shown that the line intersects, it is indicated that the proportional hazard assumption is not met for the RR_M variable. Thus the Cox regression model is called non-proportional hazard regression where there are different hazard functions from one individual to another, the model of the equation involving the RR_M variable is the Cox Nonproportional Hazard model. So the model with Box-Cox Transformation function at its survival time is as follows:

$$\hat{h}_i(t, X) = \frac{\frac{1}{2.51} \frac{1}{t^{1.5}} \exp\left\{-\frac{1}{2}\left(\frac{1}{1.73770t^{0.5}} - 2.88\right)^2\right\}}{1 - \int_0^t \frac{1}{2.51} \frac{1}{t^{1.5}} \exp\left\{-\frac{1}{2}\left(\frac{1}{1.73770t^{0.5}} - 2.88\right)^2\right\} dt} \exp(-0.04082X_{11}) \quad (11)$$

where

t = survival time

X_{11} = respiration rate

Based on the factors that affect the survival of the patient, we can acknowledge the rate of recovery seen from the value of the odd ratio of significant variables by using the formula in the equation. Due to respiration rate is a continuous variable, it can be interpreted that the hazard ratio between individuals with X value is greater 1 unit than other individuals.

Therefore, based on the result of hazard ratio calculation in Table 2 is known that the respiration rate factor has a value of 0.96, which means that every increase of one unit respiration rate, the risk to achieve the patient's recovery is $\frac{1}{0.96}$ times or it can be said that patients who have respiration rate one unit higher have risk to recover is $\frac{1}{0.96}$ or 1,041 times of patients who had one lower unit. the higher the respiration rate of a patient, eating to achieve healing longer.

IV. CONCLUSION

In conclusion, it can be obtained from the analysis results:

- 1) The distribution that was achieved in this study is the normal distribution with Box-Cox transform and can be modeled on survival analysis, as follows:

$$\hat{h}_i(t, X) = \frac{\frac{1}{2.51} \frac{1}{t^{1.5}} \exp\left\{-\frac{1}{2}\left(\frac{1}{1.73770t^{0.5}} - 2.88\right)^2\right\}}{1 - \int_0^t \frac{1}{2.51} \frac{1}{t^{1.5}} \exp\left\{-\frac{1}{2}\left(\frac{1}{1.73770t^{0.5}} - 2.88\right)^2\right\} dt} \exp(-0.04082X_{11})$$

- 2) The survival analysis result using Cox non-proportional hazard model gives a conclusion that the factor which affects the patient's endurance time to experience healing is respiration rate. This suggests that there are other factors that affect patient resilience in addition to therapy performed and patients with high respiration characteristics have a greater risk of failure than patients with lower respiration rates.

ACKNOWLEDGMENT

I would like to thank my lecturer, Mr. H. Sugiyarto, Ph.D. as my tutor lecturer in completing this article. I am also grateful to my colleagues from Ahmad Dahlan University who provide insight and expertise in research, although they may disagree with all interpretations /conclusions from this paper.

REFERENCES

1. Cox, D. R. 1972. Regression Model and Life Table. J Roy Stat Soc B, 34, 187-202.
2. Collet, D. 1994. Modelling Survival Data in Medical Research. (London: Chapman and Hall)
3. Collet, D. 2003. Modelling Survival Data In Medical Research. 2nd Edition. (London: Chapman and Hall)
4. Kleinbaum, D.G. dan Klein, M. 2005. Survival Analysis: A Self Learning, 2nd Edition. (New York: Springer)
5. Kleinbaum, D. G., & Klein, M. 2012. Survival Analysis: A Self-Learning Text. (London: Springer)
6. Law, A. M., & Kelton, D. W. 2000. Simulation Modelling Analysis. 3rd Edition. (New York: MacGraw-Hill)
7. Lawless, J.F. 1982. Statistical Models and Methods for Lifetime Data. (USA: John Wiley & Sons, Inc.)
9. Pardeshi, G. 2009. Survival Analysis And Risk Factor For Death In Tuberculosis Patients On



10. Directly Observed Treatment-Short Course. Indian Journal of Medical Sciences, 63, 180- 186.
11. Riska, dkk. 2012. Analisis Survival Faktor-Faktor yang Mepengaruhi Laju Kesembuhan
12. Pasien Penderita Dmam Berdarah Dengue (DBD) RSU Haji Surabaya dengan Regresi Cox.
13. Journal Institut Teknologi Sepuluh November.

AUTHORS PROFILE



Dr Sugiyarto is the Senior Lecture at the Faculty of Mathematics and Natural Sciences, Ahmad Dahlan University, Indonesia. His areas of interest is Applied Statistics and Mathematics, Optimization and Operational Research.



Madam Umami Aulia Septiani is a Master of Computing Science student at the Brawijaya University Malang Indonesia. Her areas of interest is in computing.



: Ts Dr Suliadi Sufahani is a Senior Lecturer and the Head Panel of Statistics in the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia. His areas of interest are Statistical Analysis, Operational Research, Mathematical Programming, Statistical Modeling, Optimization and Data Mining.