International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 8 Issue 6 June - 2024, Pages: 27-32

Nonparametric Estimation of Type 1 Cencored Survival and Hazard Functions in Leukemia Cancer Patients with Kaplan-Meier Method

[1]Evi Wijayawati, [1]Alda Fauziah Afifah, [1]Michelle Adelia Suwarno, [2]Toha Saifudin*

[1]Student of Study Program of Statistics, Faculty of Science and Technology, Airlangga University Jl. Mulyorejo, Kampus C UNAIR, Surabaya, 60115, Indonesia

[2]Department of Mathematics, Faculty of Science and Technology, Airlangga University Jl. Mulyorejo, Kampus C UNAIR, Surabaya, 60115, Indonesia

*Corresponding Author : tohasaifudin@fst.unair.ac.id

***Abstract**: Health is one of the basic necessities for the life of the elderly (lansia). Health development is also one of the goals in the Sustainable Development Goals (SDGs), precisely the third goal, namely ensuring a healthy life and promoting the well-being of all ages. In this goal, the SDGs have set several targets that must be achieved by each country. Cancer is currently still one of the world's biggest health problems. Cancer is a disease caused by the abnormal growth of cells in a particular organ, which can cause clinical symptoms for the sufferer. Based on data from the World Health Organization (WHO) in 2015, cancer is the second leading cause of death after cardiovascular disease, causing 8.8 million deaths. To date, the exact cause of leukemia is unknown, but there are predisposing factors that cause leukemia, such as genetic factors; certain viruses cause changes in gene structure (Tcell Leukemia-Lhymphoma Virus/HLTV), radiation, immunosuppressive drugs, cardiogenic drugs such as diethylstilbestrol, hereditary factors; in monozygotic twins, chromosomal abnormalities; down syndrome. Leukemia usually affects white blood cells. In this healing process, one of the methods used is placebo treatment. Based on this, in this study, the probability of survival and the risk of death of placebo method recipients up to a certain time limit will be estimated using the Kaplan-Meier statistical method. Result of the discussion the chance of life for patients receiving placebo treatment for the first week after placebo treatment is 95.2% with a decrease for the next time interval until the 25th week showing a constant chance of 18.9% until the last observation week.*

Keywords—placebo; Kaplan-Meier;Leukemia; survival; Hazard

1. Introduction

Health is one of the basic necessities for the life of the elderly (lansia). Health development is also one of the goals in the Sustainable Development Goals (SDGs), precisely the third goal, namely ensuring a healthy life and promoting the well-being of all ages. In this goal, the SDGs have set several targets that must be achieved by each country.

Cancer is currently still one of the world's biggest health problems. Cancer is a disease caused by the abnormal growth of cells in a particular organ, which can cause clinical symptoms for the sufferer. Based on data from the World Health Organization (WHO) in 2015, cancer is the second leading cause of death after cardiovascular disease, causing 8.8 million deaths.1 Cancer can be categorized based on the organ where the cancer cells grow [1]. Among the various types of cancer, leukemia is a type of cancer with a high incidence and mortality rate in the world today.

Cancer is a non-communicable disease characterized by abnormal/continuous and uncontrolled growth of cells that can damage surrounding tissues and can spread to distant sites from their origin called metastasis. Cancer cells are malignant and can originate or grow from any type of cell in the human body [2]. Cancer can attack anyone regardless of age, including children, adolescents, and adults. Each cancer has different risk factors. In adults, these risk factors are generally due to unhealthy lifestyles. However, similar causes are relatively rare in children [3].

To date, the exact cause of leukemia is unknown, but there are predisposing factors that cause leukemia, such as genetic factors; certain viruses cause changes in gene structure (Tcell Leukemia-Lhymphoma Virus/HLTV), radiation, immunosuppressive drugs, cardiogenic drugs such as diethylstilbestrol, hereditary factors; in monozygotic twins, chromosomal abnormalities; down syndrome. Leukemia usually affects white blood cells.

The cause of most types of leukemia is unknown. Exposure to radiation and certain chemicals (e.g., benzene) and the use of anticancer drugs increase the risk of leukemia. People with certain genetic abnormalities (Down syndrome and Fanconi syndrome) are also more susceptible to leukemia [4].

In this healing process, one of the methods used is placebo treatment. Placebo treatment is a process in which the patient feels sure that they are receiving medical treatment, even though in reality no medical treatment is being given to them. The purpose of placebo treatment is to test and prove the effectiveness of drugs and treatments. Placebo is a substance that does nothing, or a treatment method that does not affect the body in any way, but has a healing effect. To improve patient healing after placebo treatment, the factors

that affect post-treatment survival must be understood. One of the latest innovations for analyzing the survival of leukemia cancer patients.

Survival analysis is an analysis that involves observing the time of occurrence (time-to-event analysis). In this analysis, the time of occurrence of an expected event is called failure time or survival time. The survival function is a function that indicates the probability of an individual surviving to time t. In addition, there is a hazard function, which is a function that indicates the probability of an individual experiencing failure or death at time t on condition that the individual has survived to time t [5]. As for the approach, there are two estimation methods, namely parametric and non-parametric methods. The parametric method is done by assuming the distribution of the population first, while the non-parametric method does not require an assumption of the distribution of the population. The non-parametric method that is commonly used to estimate the survival function on incomplete (censored) data is the Kaplan-Meier method.

Based on this, in this study, the probability of survival and the risk of death of placebo method recipients up to a certain time limit will be estimated using the Kaplan-Meier statistical method.

1. THERORICIAL BASIS

2.1 Leukimia Cancer

Cancer is a broad term for a group of diseases characterized by uncontrolled and abnormal cell growth. These cells divide and multiply rapidly, invading and damaging healthy tissues. According to WHO, cancer is a large group of diseases that can affect any part of the body with other terms being malignant tumors and neoplasms with abnormal growth of new cells that develop beyond normal limits by invading other parts of the body and spreading to other organs called the metasis process [6]. Cancer is also the disease that causes the highest number of deaths.

Leukemia is a malignant and progressive disease of the blood-forming organs characterized by abnormal proliferation and development of leukocytes and their predecessors in the blood and bone marrow. Based on the Big Indonesian Dictionary, Leukemia is an acute or chronic disease due to the presence of one type of immature leukocytes that multiply malignantly in the bone marrow or lymph nodes, which then spreads to other parts of the body; blood cancer.

2.2 Placebo Treatment

Placebo treatment is a process in which a patient believes that he is receiving medical treatment, despite the fact that no medical treatment has been given to him. The purpose of placebo treatment is to test and prove the effectiveness of drugs and treatments. In some cases, the placebo effect can be as good as the actual treatment.

A placebo is a substance that produces nothing, or a treatment method that does not affect the body in any way, but has a healing effect. Placebos produce therapeutic effects mainly due to belief and the desire to get what one wants.

2.3 Survival Function

The variable T is a non-negative random variable that represents the survival time of a person. Random variable T with probability density function f(t) has a cumulative distribution function expressed in the following equation (Lawless, 2003):

$$F(t) = P(T \leq t) = \int f(u)\, du \qquad (1)$$

The survival function is defined as the chance of an individual surviving until time t and is expressed as follows:

$$S(t) = P(T > t) = \int_0^{\infty} f(u)\, du \quad t > 0 \qquad (2)$$

Based on the above definition, a new equation is obtained that states the relationship between the survival function and the cumulative distribution function, namely :

$$S(t) = 1 - P(T \leq t) \qquad (3)$$

$$S(t) = 1 - F(t) \qquad (4)$$

The survival function (t) is a monotonous (non-increasing) continuous function with time t with the following properties

When (0) = 1, because at the beginning of the observation no individual has experienced an event. Meanwhile, if at (∞) = 0, namely $lim\ S(t) = 0$ shows that in theory, if the observation time is infinite and at the end of time no individual survives. In theory, the survival function can be described by a survival curve that shows the chance of survival of an individual at a time between 0 and ∞.

The probability density function (t) is defined by the probability in the time interval t and $t = \Delta t$, the survival function equation is obtained as follows:

$$S(t) = \int_0^{\infty} f(u)\, du \qquad (5)$$

2.4 Hazard Function

In a distribution, the properties of mean and variance are very important, as well as in survival analysis. Not only the survival function, but it is necessary to add the hazard function which focuses on the occurrence of an event so that it is seen as an informant as opposed to the survival function. The hazard function is defined as the failure rate at time t h(t) until an individual survives until time t. The hazard function is defined in the following equation:

$$h(t) = \lim_{\Delta t \to \infty} \frac{(t \leq T < t + \Delta t)}{\Delta t} \qquad (6)$$

Based on this equation, we get :

$$h(t) = \frac{f(t)}{S(t)} \qquad (7)$$

The probability density function is the first negative derivative of survival, expressed by the equation :

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 8 Issue 6 June - 2024, Pages: 27-32

$$f(t) = -S(t) \quad (8)$$

2.5 Kaplan-Meier Method

This method is used to estimate nonparametric curves on the survival function to overcome incomplete (censored) data and can be used on samples that are not large [7]. The model of the Kaplan-Meier method can be written in the following equation :

$$S(t) = \frac{n_i}{d_i} \quad (9)$$

with the information that :
(t) : Probability surviving since the beginning of the study
n : Number of things that are at risk but still survive at time i
d : The number of things that experience events at time i

As for curve fitting, Kaplan-Meier is a nonparametric estimate for events from the survival distribution [8]. The cumulative proportion of something experiencing an event is represented by the time symbol t for each point.

2. MATERIAL AND METHODS

2.1 Data and Data Sources

This study uses a type of quantitative research. In this study, the type of data used is secondary data regarding remission time data or signs of cancer starting to decrease or disappear in weeks for Leukemia patients obtained from the Kaggle website.

2.2 Data collection method

A research variable is an attribute, trait or value of people, objects or activities that have certain variations set by researchers to study and then draw conclusions [9]. The variable survival time of Leukemia patients who have received placebo treatment and the status of Leukemia patients after receiving placebo treatment are used. To minimize errors in estimating the variables used, the variables will be defined as follows:

a. Survival Time

That is the length of time a patient remains alive after receiving placebo treatment at a hospital. This information is obtained based on data.

b. Status (Censorship)

Namely the condition or state of the patient after receiving placebo treatment, whether it still survives or has not survived (died).

2.3 Analysis Technique

The data analysis technique applied in the study is the life test data analysis method using the survival function and hazard function as follows.

A. Manual Calculation Method

1. Define the variables used, namely survival time or notated with ti and status (censorship) with the notation d_i.
2. Accumulate censoring into the number of patients who experience death/failure at time j $(t_{(i)})$ by making a table.
3. Calculate the survival function using the formula

$$\hat{S}(t) = G\left(\frac{n_i - d_i}{n_i}\right), i:(t)_i \leq t \quad (10)$$

4. Calculate the hazard function using the formula

$$\hat{h}(t) = \frac{d_i}{n_i} \quad (11)$$

5. Interpret the results of the calculation of the survival function and hazard function.

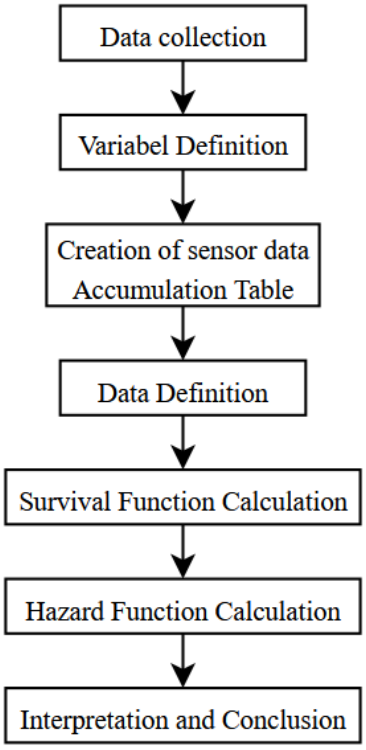

Fig. 1. Research Flowchart

3. RESULT AND ANALYSIS

3.1 Data Description

The survival data used in this study is the survival data of 42 leukemia cancer patients who received placebo treatment sourced from Kaggle.com: Emmanuel Masavo Djegou. This data shows that patients who are in remission until the end of the study time limit are included in the censored data. Here are the data for the 42 leukemia cancer patients who were received placebo treatment.

Table 1 : Data of 42 Leukemia Cancer Patients who received Placebo treatment

Patient	Time	Time	Time	Patient	Time
1	1	15	8	29	16
2	1	16	8	30	17
3	2	17	8	31	17*
4	2	18	8	32	19*
5	3	19	9*	33	20*
6	4	20	10	34	22
7	4	21	10*	35	22

8	5	22	11	36	23
9	5	23	11	37	23
10	6	24	11*	38	25*
11	6	25	12	39	32*
12	6	26	12	40	32*
13	6*	27	13	41	34*
14	7	28	15	42	35*

Based on the data, descriptive statistics of 42 leukemia cancer patients who received placebo treatment were obtained as follows.

Table 2 : Descriptive statistics of 42 Leukemia Cancer Patients

Variable	Total Number
Patients died (d_i)	12
Total number of patients (n_i)	42
Live Patients	30

Table 3 : Mean and Median for Survival time

Means and Medians for Survival Time			
Mean[a]		Median	
Estimate	Std. Error	Estimate	Std. Error
15.339	1.860	12.000	1.717

Table 2 shows that in the data of leukemia cancer patients who received placebo treatment, the total number of patients observed was 42 people. **Table 2** also shows that 12 people died or experienced survival failure. As for the average patient survival time is 15.39 weeks or if rounded up 16 weeks. It can be concluded that of the 42 leukemia cancer patients who received placebo treatment, 12 patients experienced survival failure with an average survival time of 16 weeks.

Next, to analyze Type 1 Censored using the Kaplan-Meier method, a normality and exponential test is first performed. The Kaplan Meier method is a type of non-parametric method that can be used to estimate the survival function in survival test data analysis. One of the characteristics of data that can be analyzed with non-parametric methods is that the data is not normally distributed and not exponential. The results of the normality and exponential test for the survival data of leukemia cancer patients are presented in **Table 4**.

Table 4 : Normality Test of Leukimia Cancer Data

One-Sample Kolmogorov-Smirnov Test		
		Status
N		42
Normal Parameters[a,b]	Mean	.71
	Std. Deviation	.457
Most Extreme Differences	Absolute	.448
	Positive	.266
	Negative	-.448
Test Statistic		.448
Asymp. Sig. (2-tailed)		.000[c]

Table 5 : Exponential Test of Leukimia Cancer Data

One-Sample Kolmogorov-Smirnov Test 2		
N		42[c]
Exponential parameter.[a,b]	Mean	1.00
Most Extreme Differences	Absolute	.768
	Positive	.768
	Negative	-.232
Kolmogorov-Smirnov Z		4.206
Asymp. Sig. (2-tailed)		.000

Based on the normality test in **Table 4**, it is known that Asymp. Sig 0.000 < α = 5%. While the exponential test also shows in **Table 5** that asymp.sig 0.000 < < α = 5%. Thus, it can be concluded that the data is not normally distributed and exponential, so it can be continued with the analysis of type 1 censored data using the Kaplan-Meier method.

3.2 Estimation of Survival chances of Leukimia Cancer Patients using Survival Function

The estimation of the survival function using the Kaplan-Meier method is based on the patient's survival time with the assumption that the censored data is independent based on survival. The calculation of the Kaplan-Meier survival function is as follows:

$$\hat{S}(t) = \prod_{j:t_{(i)} \le t} \frac{n_i - d_i}{n_i} \qquad (12)$$

With,

$\hat{S}(t)$: Estimation of the survival function at time t

$f(t_{(i)})$: Probability of survival at time $t_{(i)}$

n_i: Number of patients at risk of experiencing failure at $t_{(i)}$

d_i : Number of patients who experience death/failure at $t_{(i)}$

Based on the formula above, the estimated survival function is as follows:

Table 6 : Calculation of Survival Function Estimation for Patients Receiving Placebo Treatment

$t_{(i)}$	n_i	d_i	$\hat{S}(t)$
1	42	2	0,952380
2	40	2	0,904761
3	38	1	0,880952
4	37	2	0,833333
5	35	2	0,785714
6	33	3	0,714285
7	29	1	0,689654
8	28	4	0,591132
9	24	0	0,591132
10	23	1	0,565431
11	21	2	0,511580

12	18	2	0,454738
13	16	1	0,426317
15	15	1	0,397896
16	14	1	0,369475
17	13	1	0,341054
19	11	0	0,341054
20	11	0	0,341054
22	9	2	0,265264
23	7	2	0,189474
25	5	0	0,189474
32	5	0	0,189474
34	5	0	0,189474
35	5	0	0,189474

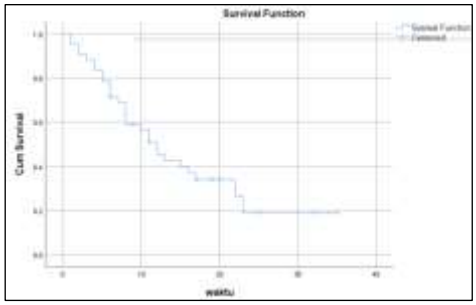

Fig. 2. Cumulative Graph of Survival Function

Table 6 shows that the column indicates the number of patients at risk of failure, and the column indicates the number of leukemia cancer patients undergoing placebo treatment. Based on **Table 6**, it is shown that leukemia cancer patients undergoing placebo treatment have a 95.23% chance of survival in the first week after placebo treatment. For the second week, patients have a 90.47% chance of survival. For subsequent time periods, the patient's chance of survival decreases according to **Figure 2**, with the addition of time inversely proportional to the chance of survival of leukemia cancer patients undergoing placebo treatment**.**

3.3 Etimation of Survival Chances of Leukimia Cancer Patients using Hazard Function

The hazard function is defined as the failure rate of the observation at time-t if the observation object survives until time-t. The hazard function is formulated as follows.

$$\hat{h}(t) = \frac{d_i}{n_i} \qquad (13)$$

with,

$\hat{h}(t)$: estimated hazard function
d_i : number of patients who died at time i $(t_{(i)})$
n_i : many patients "at risk" of death at time i $(t_{(i)})$

The results of the calculation of the hazard function estimate for patients receiving placebo treatment can be seen in the following table.

Table 7 : Calculation of Hazard Function Estimation for Patients Receiving Placebo Treatment

t_i	n_i	d_i	$\hat{h}(t)$
1	42	2	0.047619
2	40	2	0.05
3	38	1	0.026316
4	37	2	0.054054
5	35	2	0.057143
6	33	3	0.090909
7	29	1	0.034483
8	28	4	0.142857
9	23	0	0
10	23	1	0.043478
11	21	2	0.095238
12	18	2	0.111111
13	16	1	0.0625
15	15	1	0.066667
16	14	1	0.071429
17	13	1	0.076923
19	11	0	0
20	11	0	0
22	11	2	0.181818
23	9	2	0.222222
25	7	0	0
32	7	0	0
34	7	0	0
35	7	0	0

Based on **Table 7** it is shown that the results of the calculation of the estimated chance of patient death using the hazard function in patients receiving placebo treatment. This is evidenced in the first week after placebo treatment, the patient's chance of death is 4.8% and tends to fluctuate. However, in the seventh week the chance of death was at the smallest chance of 3.5%. This fluctuation continued in the following weeks. The 25th week showed a constant chance of 0% week of the last observation day.

The cumulative hazard function has the following calculations

Table 8 : Calculation of Cumulative Hazard Function Estimation for Patients Receiving Placebo Treatment

t_i	n_i	d_i	$\hat{h}(t)$
1	42	2	0,04879
2	40	2	0,10008
3	38	1	0,12675
4	37	2	0,18232
5	35	2	0,24116
6	33	3	0,33647
7	29	1	0,37156
8	28	4	0,52571
9	23	0	0,52571
10	23	1	0,57017
11	21	2	0,67025

12	18	2	0,78803
13	16	1	0,85257
15	15	1	0,92156
16	14	1	0,99567
17	13	1	1,07571
19	9	0	1,07571
20	9	0	1,07571
22	9	2	1,32703
23	7	2	1,66350
25	5	0	1,66350
32	5	0	1,66350
34	5	0	1,66350
35	5	0	1,66350

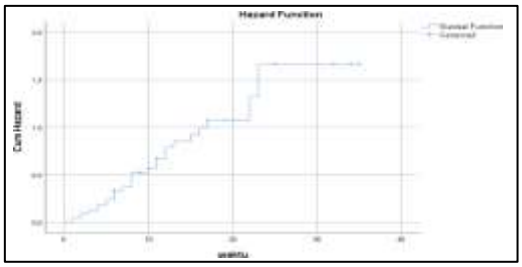

Fig. 3. Cumulative Graph of Hazard Function against Time

Based on **Table 8** it can be identified that the cumulative results of the estimated probability of death of patients using the hazard function of patients receiving placebo treatment tend to increase. The chance of patient death in the first week until week 8 is 0.52571 and tends to be constant in week 9. However, in the following days the chance increases again until the 17th week which is 1.07571. From week 19 to week 20, the chance was constant again and increased from week 22 to week 23 until the last week when the cumulative hazard chance reached 1.66350, indicating a very high chance of patient death.

4. CONCLUSIONS

Based on the discussion in the previous chapter, the following conclusions are drawn.

1. A total of 12 out of 42 placebo method recipients died with an average survival time of 14 weeks.
2. The chance of life for patients receiving placebo treatment for the first week after placebo treatment is 95.2% with a decrease for the next time interval until the 25th week showing a constant chance of 18.9% until the last observation week.
3. The chance of death of placebo-treated patients in the first week was 4.8% and tended to fluctuate. However, in the seventh week the chance of death was at the smallest chance of 3.5%. This fluctuation continued in the following weeks. The 25th week showed a constant chance of 0% week of the last observation day with a cumulative hazard chance of 1.66350 indicating a very high chance of patient death.

5. REFERENCES

[1] World Health Organization. Cancer: Fact Sheet 2015. http://www.who.int/mediacentre/factsheets/fs297/en.
[2] Departemen Kesehatan RI. (2009). Data dan Informasi Kesehatan
[3] American Cancer Society. (2013). Cancer fact and figures https://www.cancer.org/cancerfactstatistic/2013
[4] Amin and Hardhi. (2015). Aplikasi AsuhanKeperawatan Berdasarkan Diagnosa Medis & NANDA NIC-NOC Jilid 2. Jogjakarta: MediAction
[5] Kleinbaum, D. G., & Klein, M. (2005). Computer appendix: survival analysis on the computer. Statistics for Biology and Health. Survival Analysis. A Self- Kleinbaum, D. G., & Klein, M. (1996). Survival analysis a self-learning text. Springer, 508-42.
[6] WHO. (2008). Fact Sheet Cancer. www.who.int/mediacentre/factsheet/f529/7/eh
[7] Kaplan, E. L., & Meier, P. (1958, June). Nonparametric Estimation from Incomplete Observations. 53(282), 457-481. doi:10.2307/2281868
[8] Cleophas, T. J., Zwinderman, A. H., Cleophas, T. F., Cleophas, T. J., Zwinderman, A. H., & Cleophas, T. F. (2002). Proportional data analysis: Part 2. *Statistics applied to clinical trials: Self-assessment book*, 73-102.
[9] Ridha, N. (2017). Proses penelitian, masalah, variabel dan paradigma penelitian. *Hikmah*, *14*(1), 62-70.