INDONESIAN COUNCIL OF PREMIER STATISTICAL SCIENCE

Indonesian Council of Premier Statistical Science
ISSN: 3030-9956/https://dx.doi.org/10.24014/icopss.v4i1.37519

Volume 4 Issue 1, 36 – 44, February 2025

Estimation of Hazard Cumulative Function Using the Nelson-Aalen Method on Covid-19 Patient Data in Jember Regency

Hilvania Ramadhani[1], Rini Pauziah[2]

[1,2]*Program Study of Mathematics, Faculty of Science & Technology, Sultan Syarif Kasim State Islamic University Riau*
Email: hilvaniaramadhani@uin-suska.ac.id

Received: 9 June 2025 Revised: 14 June 2025 Accepted: 15 June 2025 Published: 18 June 2025

Abstract - *The Covid-19 pandemic presents a major challenge in the health sector, especially related to understanding patient recovery patterns. This study aims to estimate the cumulative hazard function using the Nelson-Aalen method on the length of treatment data of Covid-19 patients who have recovered in Jember Regency. The Nelson-Aalen method is a non-parametric approach that does not require certain distribution assumptions and is suitable for survival data, especially those subjected to right censorship. In this study, all patient data was complete without sensors. The analysis was conducted with Microsoft Excel software, resulting in a cumulative hazard curve that showed an increased risk of recovery as the treatment time increased. The results of this study provide an empirical picture of patient recovery patterns and serve as a basis for evaluating health service efficiency and hospital capacity planning during the pandemic. In addition, the application of the Nelson-Aalen method reinforces the contribution of non-parametric statistical methods in epidemiological studies.*

Keywords: *Survival Analysis, Nelson-Aalen, Cumulative Hazard, Covid-19, Length of Treatment.*

1. Introduction

The Covid-19 pandemic has become a major challenge for health systems around the world, including in Indonesia. One of the important aspects of pandemic management is understanding the patient's healing pattern, including the length of treatment time needed until the patient is declared cured. This information is not only important for hospitals, but also the basis for policy making by local governments.

Survival analysis is a statistical approach used to analyze the time until an event occurs, such as death, equipment failure, or in this context – the patient's recovery [1]. In contrast to ordinary descriptive statistics, survival analysis is able to account for cumulative risk over time, and often involves censored data, even though in this study all patients have recovered.

The Nelson-Aalen method is one of the non-parametric estimators used to estimate the cumulative hazard function, i.e. the probability of events (in this case healing) accumulating over a period of time. This method is quite simple and effective in describing risk patterns without certain distribution assumptions, so it is suitable for application to observational data such as Covid-19 patient data.

Research on the estimation of cumulative hazard function is essential to describe the change in risk over time, especially in the context of infectious diseases such as Covid-19. Using data on patients who have been declared cured, the Nelson-Aalen method can provide an idea of when the peak risk of recovery occurs and how quickly the recovery process occurs in the population. This is very useful for hospital capacity planning, evaluation of health service efficiency, and anticipation of future surge in cases.

In addition to being important for clinical purposes, the analysis of the length of treatment of Covid-19 patients also contributes to the study of statistics in the field of public health. Cumulative hazard estimation can help identify groups of patients with longer recovery times, so that more targeted interventions can be made based on specific demographic characteristics or comorbidities [2]. In addition, the use of statistical methods such as Nelson-Aalen is important because it does not require specific distribution assumptions, in contrast to parametric approaches that require an explicit form of distribution of event time. Thus, this study not only provides an empirical picture of the healing process of Covid-19 patients in Jember Regency, but also expands the application of non-parametric methods in epidemiological studies

2. Theoretical Foundations

2.1. *Survival Analysis*

Rain Survival analysis is a statistical method used to study the time it takes until an important event occurs or referred to as Event, such as death, recovery, recurrence of illness, or equipment damage. The data used is called survival data or survival data, which is data that records the time interval between the beginning of observation and the occurrence of the event. This time can be measured in units of days, weeks, months, or years depending on the type of study being conducted. In a medical context, for example, survival analysis is widely used to study the length of time a patient survives or recovers after receiving treatment [3].

In the analysis of survival, there are several important components:

1) Survival time, which is the duration from the beginning of observation until the subject experiences an event.
2) Event status, which indicates whether the subject has experienced an event (for example, recovered = 1) or not yet (censored = 0).
3) Censorship (censoring), which is a condition in which complete information about the time of the event cannot be obtained. This censorship can occur if the subject quits the study before the event occurs, or if the study ends before all subjects experience the event [4].

There are three main functions in the most relevant survival analysis:

1) Probability density function (f(t), which indicates the probability of an event occurring at a given time
2) Survival function (S(t)), i.e. the probability that the subject will survive longer than time t.
3) The hazard function (h(t)), which expresses the rate or risk of an event occurring at time t, assuming that the subject has not experienced an event up to that time.

The relationship between these three functions is mathematical, where if one of the functions is known, then the other functions can be derived from it [3].

One of the important aspects of survival analysis is the concept of censorship. There are three main types of censorship:

1) Right censoring, where the subject has not experienced an event until the end of the study period.
2) Left censoring, where the event has already occurred before observation begins.
3) Interval sensor, which is when the exact time an event occurs is unknown, but is known to occur within a certain time range.

This censorship makes survival analysis different from regular statistical analysis, as this method can still be used even if the data are incomplete as a whole [4].

The main purpose of survival analysis is to estimate survival functions and hazard functions, compare these functions between two or more groups, and assess the relationship between free variables and survival time. One of the advanced methods of analysis in survival is Cox Proportional Hazard regression, which links independent variables to the risk of event occurrence. In addition, for data that do not follow a specific distribution, non-parametric approaches such as the Kaplan-meier and Nelson-Aalen methods are used [3].

2.2. *Cumulative Hazard Function*

In survival analysis, the cumulative hazard function is one of the key concepts used to describe the accumulated risk of an event occurring over time. This function is particularly important in the context of survival data because it provides an idea of how likely it is that someone or something will experience a particular event (e.g. death or recovery) over a certain period of time [5].

Mathematically, the cumulative hazard function H(t) is defined as the integral of the hazard function over time, namely:$h(t)$

$$H(t) = \int_0^t h(u)du$$

Or it can also be written as , where is the survival function $\boldsymbol{H(t) = -\ln S(t)}$ $\boldsymbol{S(t)}$[6]. This relationship shows that the cumulative hazard function represents the "total amount of risk" that has accumulated over time, and is used as a basis for estimating the survival function indirectly through an exponential formula.$\boldsymbol{tS(t) = e^{-H(t)}}$

The interpretation of the cumulative hazard function is not as simple as the survival function that indicates the chances of survival. However, this function has a deep meaning, namely as a force of mortality or total risk that a person faces in a certain period of time. In a clinical context, this function helps doctors and researchers understand how much the patient's cumulative risk of disease recurrence or death is over observation time [7].

To estimate the cumulative hazard function, one of the popular non-parametric methods used is the Nelson-Aalen method. This estimator works by summing the ratio of the number of events that occurred at a given time to the number of individuals who are still at risk at this time. Nelson-Aalen's estimator is written as:

$$\boldsymbol{\widehat{H}(t) = \sum_{t_i \le t} \frac{d_i}{n_i}}$$

Where is the number of events at the time, and is the number of individuals who are still at risk at that time. This approach is particularly beneficial, especially on right-censored data, and provides results that are stable enough to describe the patterns of risk accumulated in the population $\boldsymbol{d_i t_i n_i}$[8].

Practically, the cumulative hazard function can be visualized in the form of curves that continue to increase over time, since the risks are accumulative. When the curve goes up steeply, it indicates a period of time where the risk of events is very high. Conversely, if the curve is sloping, it means that the risk is relatively stable or low. Therefore, the cumulative hazard function is not only theoretically important, but also has great applicative value in risk modeling in fields such as medicine, actuarial, and engineering engineering [5].

2.3. *Method Nelson-Aalen*

The Nelson-Aalen method is a nonparametric approach used to estimate the *cumulative hazard* from the survival data that has been sensory, in particular *right-censored* data [9]. This estimator was introduced by Nelson (1969, 1972) and further developed by Aalen (1978) using the *counting process* and martingal theory, which allows for in-depth statistical analysis of survival data and other stochastic processes.

Mathematically, the Nelson-Aalen estimator for *the cumulative hazard* $\boldsymbol{A(t)}$ function is formulated as:

$$\widehat{\boldsymbol{A}}(\boldsymbol{t}) = \sum_{\boldsymbol{t_j \le t}} \frac{\boldsymbol{d_j}}{\boldsymbol{r_j}}$$

Where is the number of events (e.g. deaths) at the time , and is the number of individuals who are still "at risk" (have not experienced an event and have not been censored) shortly before . This estimator is in the form of $\boldsymbol{d_j t_j r_j t_j}$*a step function* that increases discretely at each time of the event.

The main advantage of the Nelson-Aalen estimator lies in its nonparametric nature, so it does not require distribution assumptions about the time of occurrence. This estimator is also almost unbiased and has variances that can be estimated with the formula:

$$\widehat{\boldsymbol{\sigma}}^{\boldsymbol{2}}(\boldsymbol{t}) = \sum_{\boldsymbol{t_j \le t}} \frac{(\boldsymbol{r_j - d_j})\boldsymbol{d_j}}{(\boldsymbol{r_j - 1})\boldsymbol{r^2}_j}$$

In large samples, the Nelson-Aalen estimator is uniformly consistent and follows the asymptotic normal distribution, which allows the construction of confidence intervals for *the cumulative hazard function.* One commonly used form of trust interval is log-normal transformation:

$$\widehat{\boldsymbol{A}}(\boldsymbol{t})\boldsymbol{exp}\left(\pm \boldsymbol{z_{1-\alpha/2}} \cdot \frac{\widehat{\boldsymbol{\sigma}}(\boldsymbol{t})}{\widehat{\boldsymbol{A}}(\boldsymbol{t})}\right)$$

Application

The Nelson-Aalen estimator is not only used in the context of simple survival data, but also in more complex models such as:

- Multistate model: for estimating the intensity of transitions in the Markov process, e.g. the transition from healthy to sick or sick to dead.
- Data with left truncation: when the subject enters the study at varying times.
- Epidemiological model: to estimate the cumulative rate of infection in the spread of infectious diseases.
- Data with *relative mortality*: when individual hazards are calibrated against an external standard population.

Relationships with Other Methods

The Nelson-Aalen estimator also has a close relationship with the famous Kaplan-Meier estimator. Theoretically, the survival function can be estimated from the Nelson-Aalen estimator through:$\boldsymbol{S(t)}$

$$\widehat{\boldsymbol{S}}(\boldsymbol{t}) = \boldsymbol{exp}\left[-\widehat{\boldsymbol{A}}(\boldsymbol{t})\right]$$

While Kaplan-Meier is a direct estimator of the survival function, Nelson-Aalen provides a robust theoretical approach to the cumulative hazard function, and is more flexible in the context of martingal-based stochastic process models [10].

3. Research Methods

3.1 Types of Research

This study is a quantitative research with *a non-parametric survival analysis* approach which aims to estimate the cumulative hazard function in Covid-19 patients treated in Jember Regency using the Nelson-Aalen method

3.2 Data Analysis Techniques

The analysis was carried out using the Nelson-Aalen method to estimate the cumulative hazard function . $\widehat{\boldsymbol{H}}(\boldsymbol{t})$The analysis steps are as follows:

1) Survival time calculation: calculates the length of treatment of each patient from the date of admission to the date of recovery

2) Timing sequencing: determining the order in which events occur (healing)
3) Determine the number of events and individuals at risk:
   - d_j: Number of patients recovered at the time t_j
   - r_j: The number of patients who are still in care shortly before the time t_j
4) Calculating the Nelson-Aalen estimate:

$$\hat{H}(t) = \sum_{t_i \le t} \frac{d_i}{r_i}$$

5) Visualization of results: create a cumulative hazard function graph over time to observe the pattern of recovery risk over time.
6) Interpretation: interpreting the cumulative hazard curve to find out when the risk of recovery is highest and how the recovery process unfolds in aggregate.
   The calculation is done using statistical software such as Microsoft Excel for analysis to be done manually.

4. Results and Discussion

4.1 Deskriptif Data

Table 1. Actual Data

Yes	The patient	Status	Duration of Treatment Until Recovery (Days)	Gender
1	1	Recover	12	*
2	2	Recover	46	*
3	10	Recover	14	*
4	11	Recover	14	*
5	3	Recover	38	*
6	5	Recover	41	*
7	6	Recover	44	*
8	8	Recover	36	*
9	26	Recover	14	*
10	18	Recover	22	Woman
11	21	Recover	24	Man
12	23	Recover	24	Man
13	19	Recover	31	*
14	20	Recover	31	Man
15	22	Recover	31	Man
16	15	Recover	38	Man
17	48	Recover	13	*
18	49	Recover	12	Woman
19	50	Recover	12	Man
20	57	Recover	11	*
21	63	Recover	8	*
22	65	Recover	7	*
23	36	Recover	18	*
24	37	Recover	18	*
25	47	Recover	14	*
26	66	Recover	10	*
27	91	Recover	3	*
28	34	Recover	23	Woman

29	35	Recover	23	*
30	58	Recover	17	*
31	61	Recover	16	*
32	28	Recover	2	*
33	60	Recover	20	*
34	67	Recover	15	Woman
35	84	Recover	13	Woman
36	87	Recover	11	*
37	88	Recover	11	*
38	38	Recover	27	*
39	70	Recover	16	Man
40	78	Recover	15	Man
41	79	Recover	15	Woman
42	92	Recover	11	*
43	99	Recover	8	*
44	62	Recover	23	*
45	80	Recover	19	Woman
46	89	Recover	16	*
47	96	Recover	11	*
48	104	Recover	9	*
49	86	Recover	20	Woman
50	110	Recover	7	*
51	14	Recover	28	*
52	83	Recover	24	Man
53	103	Recover	14	*
54	39	Recover	38	*
55	54	Recover	33	Man
56	130	Recover	4	*
57	109	Recover	15	*
58	68	Recover	31	*
59	33	Recover	47	*
60	158	Recover	7	*
61	29	Recover	51	*
62	108	Recover	22	*
63	127	Recover	13	*
64	30	Recover	52	*
65	32	Recover	52	*
66	46	Recover	45	*
67	52	Recover	44	*
68	53	Recover	44	*
69	56	Recover	43	*

70	166	Recover	6	*
71	169	Recover	6	*
72	167	Recover	7	*
73	173	Recover	7	*
74	174	Recover	7	*
75	175	Recover	7	*
76	176	Recover	7	*

This data shows several samples of COVID-19 patients in Jember Regency who have recovered [11]. The main focus of the data was on how long they were treated until they recovered. Although the gender column exists, the information has not been filled in completely

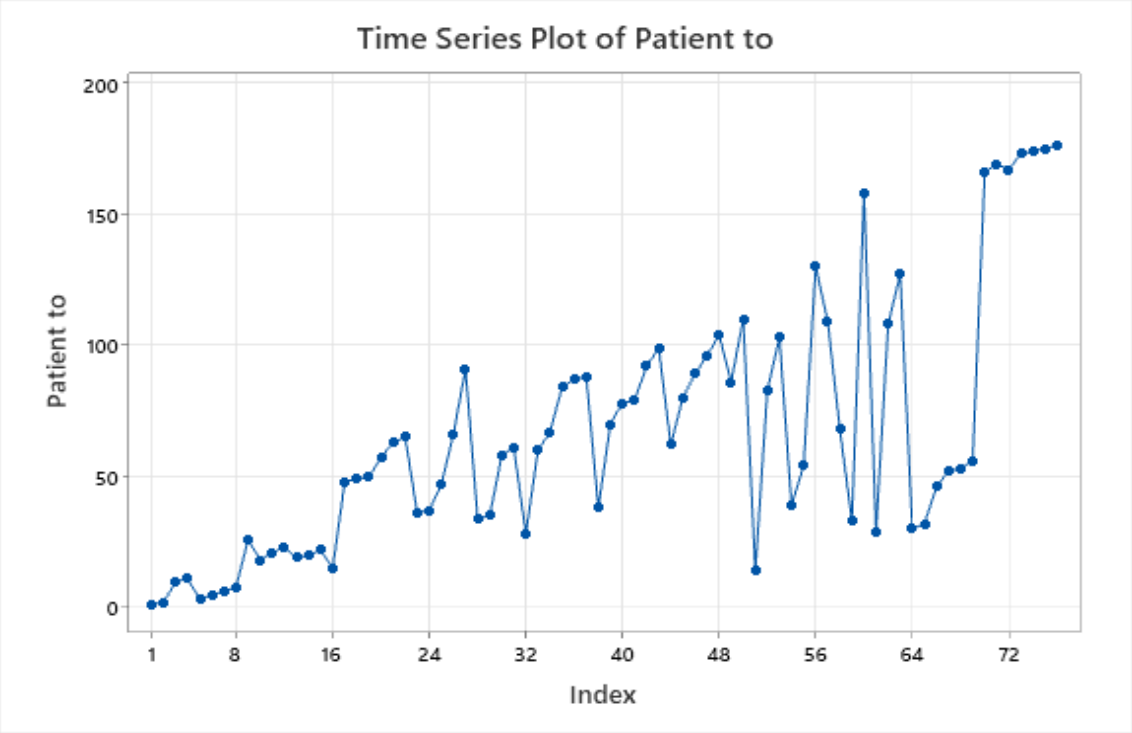

Figure 1. Actual Data Graph

The "Patient-to-" time series graph shows the order in which COVID-19 patients are recorded who have recovered in Jember Regency. It is seen that the order of patients is not recorded sequentially by time, characterized by fluctuations up and down on the graph. This is likely due to the process of collecting data from different sources or times. At the end of the chart, there is a fairly sharp and steady rise, which indicates that the data is starting to be recorded more regularly. This graph illustrates irregularities in the recording of patient sequences, rather than the development of the number of cases over time.

4.2 Application of the Nelson-Aalen Method

The application of the Nelson-Aalen method in this study aims to estimate the cumulative hazard function from the length of treatment data of COVID-19 patients in Jember Regency. All patients in the data have been declared cured, so there is no censored data, and all events are considered to have occurred. Therefore, the Nelson-Aalen approach is very appropriate because it does not require the assumption of a specific event time distribution and is able to handle survival data in a nonparametric manner.

Table 2. Implementation of the Nelson-Aalen Method with Microsoft Excel

Leaks	Fustat	The Number Of Observations At Risk	Hazard	Cumulative Hazard
12	1	76	0,013158	0,013157895
46	1	75	0,013333	0,026491228

14	1	74	0,013514	0,040004742
14	1	73	0,013699	0,053703372
38	1	72	0,013889	0,067592261
41	1	71	0,014085	0,081676768
44	1	70	0,014286	0,095962482
36	1	69	0,014493	0,110455236
14	1	68	0,014706	0,125161118
22	1	67	0,014925	0,140086491
24	1	66	0,015152	0,155238006
24	1	65	0,015385	0,170622622
31	1	64	0,015625	0,186247622
31	1	63	0,015873	0,202120637
31	1	62	0,016129	0,21824967
38	1	61	0,016393	0,234643112
13	1	60	0,016667	0,251309779
12	1	59	0,016949	0,268258932
12	1	58	0,017241	0,285500311
11	1	57	0,017544	0,303044171
8	1	56	0,017857	0,320901313
7	1	55	0,018182	0,339083132
18	1	54	0,018519	0,35760165
18	1	53	0,018868	0,376469575
14	1	52	0,019231	0,395700344
10	1	51	0,019608	0,415308187
3	1	50	0,02	0,435308187
23	1	49	0,020408	0,45571635
23	1	48	0,020833	0,476549684
17	1	47	0,021277	0,497826279
16	1	46	0,021739	0,51956541
2	1	45	0,022222	0,541787632
20	1	44	0,022727	0,564514905
15	1	43	0,023256	0,587770719
13	1	42	0,02381	0,611580242
11	1	41	0,02439	0,635970486
11	1	40	0,025	0,660970486
27	1	39	0,025641	0,686611512
16	1	38	0,026316	0,712927301
15	1	37	0,027027	0,739954328
15	1	36	0,027778	0,767732106
11	1	35	0,028571	0,796303535
8	1	34	0,029412	0,8257153

23	1	33	0,030303	0,85601833
19	1	32	0,03125	0,88726833
16	1	31	0,032258	0,919526394
11	1	30	0,033333	0,952859728
9	1	29	0,034483	0,987342486
20	1	28	0,035714	1,023056772
7	1	27	0,037037	1,060093809
28	1	26	0,038462	1,098555348
24	1	25	0,04	1,138555348
14	1	24	0,041667	1,180222014
38	1	23	0,043478	1,223700275
33	1	22	0,045455	1,269154821
4	1	21	0,047619	1,316773868
15	1	20	0,05	1,366773868
31	1	19	0,052632	1,419405447
47	1	18	0,055556	1,474961003
7	1	17	0,058824	1,533784532
51	1	16	0,0625	1,596284532
22	1	15	0,066667	1,662951199
13	1	14	0,071429	1,73437977
52	1	13	0,076923	1,811302847
52	1	12	0,083333	1,89463618
45	1	11	0,090909	1,985545271
44	1	10	0,1	2,085545271
44	1	9	0,111111	2,196656382
43	1	8	0,125	2,321656382
6	1	7	0,142857	2,464513525
6	1	6	0,166667	2,631180192
7	1	5	0,2	2,831180192
7	1	4	0,25	3,081180192
7	1	3	0,333333	3,414513525
7	1	2	0,5	3,914513525
7	1	1	1	4,914513525

This table displays the results of estimating hazard and cumulative hazard functions based on the patient's treatment time until recovery. Each row represents one time of occurrence, i.e. when at least one patient recovers at a given time. The Futime column indicates the days on which the patient recovered, while the Fustat column is valued at 1 to indicate that a recovery event has occurred at that time (no data were censored in this study). The Number Of Observations At Risk column shows the number of patients who are still in treatment (at risk) shortly before that time.

The Hazard column shows the value of the hazard rate at a given time, calculated as the ratio of the number of incidents to the number of individuals who are still at risk. This value is an estimate of the probability of recovery on that day. The smaller the number of patients remaining (at risk), the greater the hazard value in general, reflecting the greater the likelihood of the patient being cured in the near future.

The last column, cumulative hazard, is the result of the cumulative sum of the hazard value up to a certain time, which is the estimate of the cumulative hazard function $\hat{H}(t)$. This function describes the accumulation of healing risks that increase over time. As can be observed, the cumulative hazard value continued to increase from the beginning to the end of the observation time, reaching a maximum value of about 4.91 by the time all patients had recovered.

It was seen that at the beginning of the period, the hazard value was relatively small because many patients were still under treatment (e.g. 0.013 on day 12). However, the hazard value increases sharply in the final hours, especially when there are only a few patients left in care. For example, when only 3 patients are left (at risk = 3), the hazard value reaches 0.333, and when there is 1 patient left, the hazard becomes 1, which means that the only remaining patient will be cured at that time. This reflects a common characteristic of the Nelson-Aalen cumulative hazard function, which is that the curve increases exponentially near the end of the observation period, as the event becomes more certain.

Overall, this table shows that the Nelson-Aalen method provides an accurate and gradual picture of the increased risk of recovery in COVID-19 patients, from the slow initial days to the sharp acceleration towards the end of the entire patient's treatment period.

5. Conclusion

This study successfully applied the Nelson-Aalen method as a non-parametric approach to estimate the cumulative hazard function in the long-term data of Covid-19 patient care in Jember Regency. The estimated results show that the cumulative risk of recovery increases over time, with some periods showing an acceleration of the recovery process. The cumulative hazard curve provided a clear picture of the dynamics of patient healing, so that it could be a basis for consideration in evaluating health service efficiency and hospital capacity planning.

The advantages of the Nelson-Aalen method lie in its simplicity and its ability to handle survival data without certain distribution assumptions, making it a flexible analytical tool in the context of epidemiology and public health. In the future, similar research can be further developed by considering factors such as age, gender, or comorbidities to produce more comprehensive and useful models for decision-making in the health sector.

Bibliography

[1] D. Turner, “The Electrical Conductivity of Blood and Urine in Health and in Disease and as a Test of the Functional Efficiency of the Kidney,” *Am. J. Med. Sci.*, vol. 134, no. 3, p. 440, 1907, doi: 10.1097/00000441-190709000-00027.

[2] M. Wiederkehr, “Brown tumor complicating end-stage kidney disease,” *Clin. Nephrol. – Case Stud.*, vol. 8, no. 01, pp. 72–79, 2020, doi: 10.5414/CNCS110195.

[3] C. Feng, Y. Han, P. Wu, and X. Tu, *Survival analysis*. 2013. doi: 10.1007/978-1-4614-4322-3_1.

[4] R. Adolph, "CHAPTER II LITERATURE REVIEW," pp. 1–23, 2016.

[5] T. Hermawan, D. Nurrohmah, and I. F. Jannah, "Estimation of Survival Function and Cumulative Hazard Function in Survival Data of Multiple Myeloma Sufferers and Factors Affecting Their Survival Time," *Intersections*, vol. 4, no. 2, pp. 33–43, 2019, doi: 10.47200/intersections.v4i2.507.

[6] Stanford University, “Survival Distributions , Hazard Functions , Cumulative Hazard,” *Report*, no. 0, pp. 1–14, 2000.

[7] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman, “Survival Analysis Part I: Basic concepts and first analyses,” *Br. J. Cancer*, vol. 89, no. 2, pp. 232–238, 2003, doi: 10.1038/sj.bjc.6601118.

[8] I. Y. Chen, "Lecture 5 : Survival Analysis," pp. 1–7, 2018.

[9] Ø. Borgan, “ N elson– A alen Estimator ,” *Wiley StatsRef Stat. Ref. Online*, no. 2, 2014, doi: 10.1002/9781118445112.stat06045.

[10] B. P. Hobbs, “On Nonparametric Hazard Estimation,” *J. Biom. Biostat.*, vol. 06, no. 01, 2015, doi: 10.4172/2155-6180.1000232.

[11] W. Hariadi and S. Sulantari, "Analysis of the Long Survival of Covid-19 Patients with the Kaplan-Meier Method and Log-Rank in Jember Regency," *Transform. J. Educators. Matt. and Matt.*, vol. 5, no. 1, pp. 415–425, 2021, doi: 10.36526/tr.v5i1.1000.