P-ISSN: 2986-4178
E-ISSN: 2988-4004

Emerging Statistics and Data Science Journal
Vol 3, No.3, Tahun 2025

Survival Analysis of Kidney Failure Patients Using the Kaplan Meier Method and Log-Rank Test

Fatika Lovina Febrianti[1], Firdatul Fahria[2], Siti Hamidah Ardhy[3], Rahma Nurmalita[4], Suci Laeliyah[5], Muhammad Rifqy Ardiansyah[6], Ihsan Fathoni[7,*] M. Al Haris[8]

[1,2,3,4,5,6,8] Departement of Statistics, Universitas Muhammadiyah Semarang, Kedungmundu Raya Street No.18, Tembalang, Semarang City, 50275, Indonesia
[7] Departement of Sains Data, Universitas Muhammadiyah Semarang, Kedungmundu Raya Street No.18, Tembalang, Semarang City, 50275, Indonesia

*Corresponding author: ihsanfathoni@unimus.ac.id

P-ISSN: 2986-4178
E-ISSN: 2988-4004

Riwayat Artikel
Dikirim: 25 Januari 2025
Direvisi: 08 Mei 2025
Diterima: 27 Oktober 2025

ABSTRACT
Kidney failure is a global health issue that continues to rise, impacting patients' quality of life and placing significant pressure on healthcare systems. This study aims to analyze the factors influencing the survival probability of kidney failure patients using the Kaplan-Meier method and the log-rank test. Medical records from 106 kidney failure patients treated at Hasanuddin Hospital between 2018 and 2020 were used to examine the effects of age, gender, and disease severity on survival outcomes. The Kaplan-Meier analysis revealed that patients aged ≤ 50 years, females, and those with chronic conditions had better survival probabilities. However, the log-rank test indicated that survival differences based on these three variables were not statistically significant (p-value > 0,05). These findings provide an initial understanding of the survival patterns of kidney failure patients and highlight the need for further research with larger sample sizes or more advanced methods to support the development of personalized care strategies and improve patients' quality of life.

Keywords: Kaplan-Meier, Log-Rank, Survival Analysis, Kidney Failure

1. Introduction

Kidney failure is one of the global health challenges that is increasingly alarming, with its prevalence steadily rising due to the aging population and risk factors such as diabetes and hypertension. According to a report from the World Health Organization (WHO), more than 850 million people worldwide suffer from impaired kidney function. This condition not only affects the quality of life of patients but also places a significant burden on healthcare systems [1]. Both acute and chronic kidney failure can trigger serious complications that may be life-threatening. Therefore, understanding patient prognosis is crucial for designing more effective treatments [2].

Survival analysis has become a primary method for evaluating patient survival time based on various risk factors. One of the most used techniques in this analysis is the Kaplan-Meier method, first introduced by Kaplan and Meier in 1958. This method enables the estimation of survival probabilities in incomplete data [3]. On the other hand, the log-rank test is used to compare survival curves between two or more groups to determine whether there are statistically significant differences [4]. This study focuses its analysis on several key variables, namely age, gender, and the severity of kidney failure [5].

Age is known to be one of the main factors affecting patient prognosis, as older patients tend to have lower survival rates compared to younger patients. Additionally, gender is also considered influential, with some studies reporting differences in outcomes between men and women in the context of kidney disease. The severity of the disease, whether chronic or acute, is another critical aspect of this analysis. Chronic kidney failure is often associated with poorer prognosis compared to acute kidney failure, especially due to factors such as delayed diagnosis and limited access to adequate medical interventions [6]. By examining the relationships between these three variables, this study aims to provide deeper insights into the factors influencing the survival of kidney failure patients. Previous studies also support the importance of survival analysis in understanding the prognosis of kidney failure patients. For instance, a study by Ramadhani (2023) conducted at Hasanuddin University Hospital showed that demographic and clinical factors significantly affect patient outcomes. Another study by Tampake & Doho [7] revealed that age and gender influence the success of hemodialysis therapy in chronic kidney failure patients, with younger patients showing higher survival rates. Similar findings were reported by Sulistyaningrum et al. [8], who identified a relationship between age and fatigue as well as the quality of life in patients undergoing hemodialysis.

The log-rank test has been widely used to compare survival curves between groups with different characteristics. For example, a study by [9] utilized this method to analyze the survival of Covid-19 patients, which is also relevant in the context of kidney failure. Through a survival analysis approach using the Kaplan-Meier method and log-rank test, this study aims to contribute to a better clinical understanding of the prognosis of kidney failure patients. The results are expected to support the development of more effective treatment strategies and promote personalized patient management to improve their overall quality of life [10].

Although previous studies have explored factors influencing kidney failure prognosis, most of them either focus on a single variable or rely on multivariate models such as the Cox proportional hazards model. Research that uses a purely descriptive approach like Kaplan-Meier curves combined with log-rank testing to compare survival functions based on all three factors simultaneously (age, gender, and disease severity), especially in a local or general population context, is still limited. This study addresses that gap by using a straightforward but informative method to visually and statistically explore group differences. This approach not only enriches existing literature with more accessible tools but also offers a foundational reference for future research using larger samples or more advanced techniques.

2. Methods

2.1. Data

This study uses secondary data sourced from Hasanuddin University Hospital. The analyzed data consists of medical records of kidney failure patients recorded during the period from January 2018 to December 2020. The dataset includes 106 samples and contains information such as duration of care, number of patients, patient status, severity of the disease, gender, and age group.

2.2. Survival Analysis

Survival originates from the word "survive", which means endurance or survival. In this context, the term survival analysis is interpreted as an analysis of endurance. Generally, survival analysis is defined as a set of statistical methods used to analyze data with an outcome variable represented by the time until a certain event occurs [11]. This definition aligns with Muhajir and Palupi, who state that survival analysis aims to study data where the observed outcome variable is the time until a specific event occurs, also known as survival time [12].

In survival analysis, three main terms need to be understood:

a. Survival time: The duration of time an individual survives during the observation period until a specific event occurs.
b. Event: The main variable of interest in the study. This event is often associated with negative occurrences, such as death or the onset of a disease.
c. Censored data: Occurs when information about an individual's survival time is available, but the exact survival time is unknow.

Censored data can be categorized into three types:

a. If no failure occurs from the start of the study to the end of the observation, this is called right censoring.
b. If failure occurs before the study begins, this is called left censoring.
c. If failure occurs during the observation period but is not recorded, this is called interval censoring.

2.3. Survival Time Distribution

The survival time distribution represents a probability pattern related to the duration until a specific event occurs. In survival analysis, this distribution is used to analyze data characteristics, such as the median survival time, the probability of survival up to a certain time, or the likelihood of an event occurring within a certain time frame [13]. Survival time distribution is typically explained using the following key functions:

2.3.1. Continuous Survival Time Distribution Model

The survival function represents the probability that an individual can survive beyond a certain time threshold. This function indicates the probability that the random variable T is greater than a specific time t. If T is a non-negative random variable in the range $[0, \infty]$ representing the time of an event in a population, then $f(t)$ is the probability density function of T. Thus, the probability that an individual does not experience the event until time t is expressed through the survival function $K(t)$. [14]

$$\begin{aligned} K(t) &= P(T \geq t) \\ &= \int_t^\infty f(x)dx \end{aligned} \tag{1}$$

Based on the definition of the survival function related to the cumulative distribution of T, the survival function can be formulated as follows.

$$\begin{aligned} K(t) &= P(T \geq t) \\ &= 1 - P(T \geq t) \\ &= 1 - F(t) \\ F(t) &= 1 - K(t) \end{aligned} \tag{2}$$

Where:

$K(t)$: Survival Function (Life expectancy function)

T : Remaining lifespan

t : Age

P : Probability Function

The characteristics of the survival function $K(t)$ are as follows:

a. The survival function $K(t)$: is inversely related to the value of t, as t increases $K(t)$ decreases.
b. For $t = 0$, $K(t) = K(\infty) = 0$ this occurs at the beginning of the study when no subjects have experienced a failure event $P(T > 0) = 1$.
c. For $t = \infty$, $K(t) = K(\infty) = 0$, This theoretically occurs if the study period is extended to infinity (∞), where no samples can survive.

After understanding the characteristics of the survival function, we will discuss the probability density function of object failure within a certain time interval.

a. Density Function Object Failures in Intervals
The density function represents the probability rate of an event occurring at a specific time t. This function is used to measure the relative probability of an event within a very small-time interval. The formula for the density function is:

$$f(t) = \lim_{\Delta t \to 0} \frac{P[Object\ failures\ in\ intervals\ t\ (t + \Delta t)]}{\Delta t} \quad (3)$$

$$= \lim_{\Delta t \to 0} \frac{P(t \leq T < (t + \Delta t))}{\Delta t}$$

b. Hazard Function
The hazard function represents the probability that an individual experiences failure or death, given that the individual has survived up to time t. The hazard function, denoted as $h(t)$ is formulated as follow:

$$h(t) = \frac{f(t)}{R(t)} \quad (4)$$

Meanwhile, the cumulative hazard function describes the total disruption or risk experienced by a product over the time interval $[0, t]$. his cumulative hazard function is denoted as $H(t)$ and its formula is as follows:

$$H(t) = -ln\ R(t) \quad (5)$$

2.3.2. Survival Time Distribution for the Discrete Model

The survival function in a discrete model can be defined as follows [15].

$$K(t) = P(T \geq t) \quad (6)$$
$$= \sum_{j:t_i \geq t} f(t_i)$$

a. Density Function
The probability density function for a discrete model at $t_1, t_2, \ldots$, where $0 \geq t_1 < t_2 < \ldots$, can be defined as follows:

$$f(t_i) = K(t_i) - K(t_i + 1) \quad (7)$$

b. Hazard Function
The hazard function is a monotonically decreasing function with values $K(0)$ and $K(\infty) = 0$. Therefore, the hazard function in the discrete model is as follows:

$$h(t_i) = \frac{f(t_i)}{k(t_i)} \tag{8}$$

2.4. Kaplan-Meier Survival Analysis

The Kaplan-Meier method is used to estimate the survival function over time based on the data obtained. The survival function $S(t)$ is defined as the probability that an object survives without experiencing an event. $S(t)$ represents the estimated survival function or the probability of survival time being greater than t.

The Kaplan-Meier estimation formula:

$$S(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right) \tag{9}$$

Where:

d_i : Number of events at time t_i

n_i : Number of individuals still under observations before time t_i

2.5. Log-Rank Test

The Log-Rank test is used to compare survival functions between groups. The hypotheses tested are:

H_0 : There is no difference in survival functions between groups.

H_1 : There is no difference in survival functions between groups

Log-Rank Test Statistic Formula:

$$x^2 = \frac{(\sum(O_i - E_i))^2}{\sum V_i} \tag{10}$$

Where:

O_i : Observed number of events in group i

E_i : Expected number of events in group i

V_i : Variance of the difference $O_i - E_i$

2.6. Survival Time Distribution

Kidney failure occurs when the kidneys experience a progressive decline in function, rendering them unable to filter body electrolytes and maintain fluid and chemical balance in the blood or urine. This condition can result from various factors, including the narrowing of arteries leading to the kidneys, which causes inflammation and kidney damage. Kidney disease can also be influenced by other illnesses affecting the patient [16]. Patients with kidney failure may undergo various therapies to slow the progression of the disease, such as hemodialysis, peritoneal dialysis, or kidney transplantation. Although these therapies can help, they cannot fully cure kidney disease [17]. The groups most vulnerable to kidney damage include elderly individuals, those with obesity, people of Black ethnicity, and individuals from the Indian subcontinent. Mortality rates for kidney failure patients range between 30-70%, depending on the patient's age. Most surviving patients suffer from kidney impairment, while a small proportion progresses to end-stage kidney disease [18].

2.7. Analysis Stages

1. Data Input
   The dataset used in this study is imported into RStudio.
2. Descriptive Analysis
   Descriptive analysis is performed to summarize and explore the dataset. This includes calculating measures such as frequencies, proportions, means, medians, and standard deviations, depending on the variable types.
3. Kaplan-Meier Estimation
   The Kaplan-Meier estimator is utilized to estimate the survival probabilities of kidney failure patients. This method calculates the probability of survival beyond a specific time point for different categories of variables such as severity, gender, and age.
4. Log Rank Test
   The Log-Rank test is conducted to compare the survival distributions between different groups within each variable. This test assesses whether there is a statistically significant difference in survival probabilities.

2.8. Flowchart

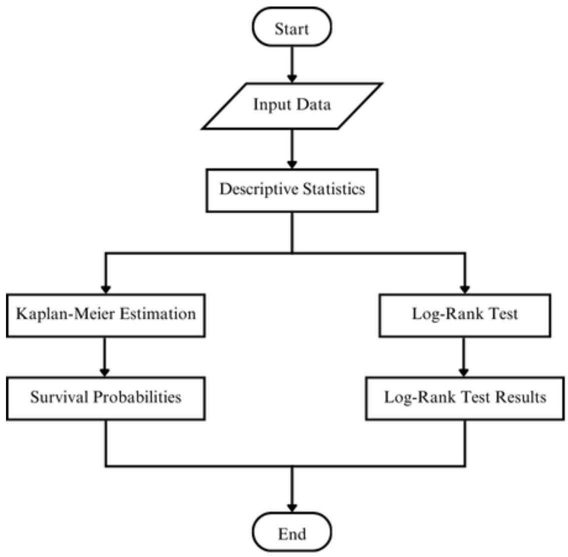

Figure 1. Research Flowchart

3. Result and Discussion

3.1. Descriptive Analysis

Table 1 Descriptive Statistics of Kidnry Failure Patients

Variable	Count
Patients who died from kidney failure	16
Censored patients with kidney failure	90
Total patients with kidney failure	106

Source: Medical Records Data from Universitas Hasanuddin Hospital, 2018-2020

Based on Table 1, the dataset consists of 106 patients, where the majority are censored, with 90 patients (85%) and 16 patients (15%) who have died. The variables analyzed in this study include the duration of kidney failure treatment (in days), patient status (censored: alive, and uncensored: deceased), and the total number of kidney failure patients recorded from January 2018 to December 2020.

3.2. Relationship between variables and patient status

3.2.1.The Relationship between Disease Severity and Patient Status

Table 2 Kidney Failure Patients Based on Disease Severity

Disease Severity	Status	
	Censored	Deceased
Chronic	31	3
Acute	59	13

Source: Medical Records Data from Hasanuddin University Hospital, 2018-2020

Based on Table 2, there are 34 patients with chronic severity, with 31 patients (91.18%) censored and 3 patients (8.82%) deceased. Meanwhile, of the 72 patients with acute severity, 59 patients (81.94%) are censored and 13 patients (18.06%) are deceased. This shows that patients with chronic severity have a lower mortality rate than patients with acute severity. However, patients with acute severity are more likely to be censored, indicating a difference in the severity levels or the effectiveness of care between the two groups.

3.2.2.The Relationship between Gender and Patient Status

Table 3 Kidney Failure Patients Based on Gender

Gender	Status	
	Censored	Deceased
Male	46	9
Female	44	7

Source: Medical Records Data from Hasanuddin University Hospital, 2018-2020

Based on Table 3, there are 55 male patients, with 46 patients (83.64%) censored and 9 patients (16.36%) deceased. On the other hand, of the 51 female patients, 44 patients (86.27%) are censored and 7 patients (13.73%) are deceased. This shows that females have a slightly lower mortality rate than males, although the difference is not statistically significant.

3.2.3.The Relationship between Age and Patient Status

Table 4 Kidney Failure Patients Based on Age

Age	Status	
	Censored	Deceased
≤ 50 Years	34	4
> 50 Years	56	12

Source: Medical Records Data from Hasanuddin University Hospital, 2018-2020

Based on Table 4, the group aged ≤ 50 years includes 38 patients, with 34 patients (89.47%) censored and 4 patients (10.53%) deceased. Meanwhile, the group aged > 50 years consists of 68 patients, with 56 patients (82.53%) censored and 12 patients deceased. This indicates that patients aged > 50 years have a higher mortality rate than aged ≤ 50 years.

3.3. Kaplan Meier Method Analysis

The following is a graph estimating the severity of kidney failure patients at Hasanuddin University Hospital, analyzed using the Kaplan-Meier method with the assistance of RStudio software.

3.3.1.Kaplan Meier Analysis Based on Disease Severity

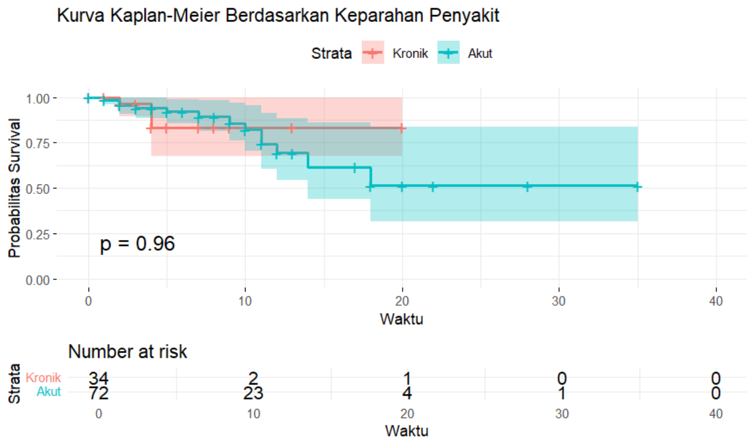

Figure 2. Kaplan Meier Curve for Disease Severity Factor

Based on Figure 2, the Kaplan Meier analysis shows that patients with acute disease severity have a lower survival probability compared to patients with chronic disease severity. This is evident from the Kaplan Meier graph, where the survival curve for acute patients is below the curve for chronic patients. This indicates that patients with acute severity are at a higher risk of death during the observation period. The survival curve for chronic patients tends to stabilize after a certain point, whereas the acute patients' curve shows a gradual decline, with the survival probability reaching around 51.3% on day 18.

3.3.2.Kaplan Meier Analysis Based on Gender

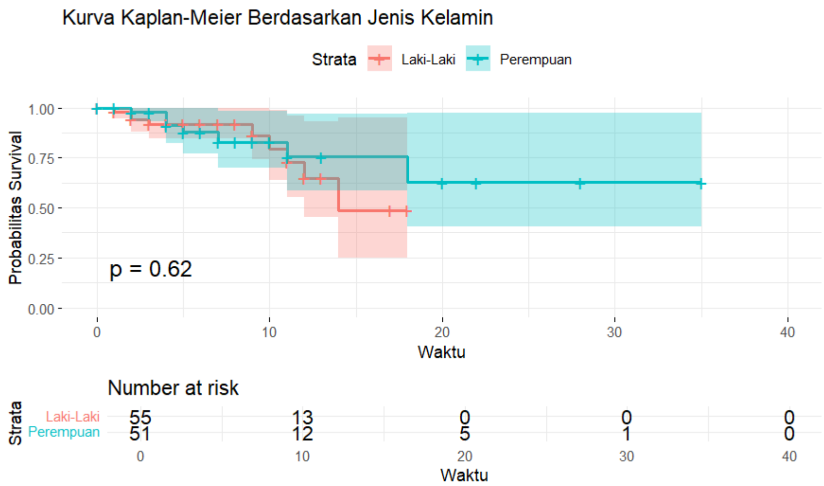

Figure 3. Kaplan Meier Curve for Gender Factor

Based on Figure 3, the Kaplan Meier graph shows that female patients have a slightly higher survival probability compared to male patients. The survival curve for female patients is above the curve for male patients, indicating that females tend to have a lower risk of death. However, the difference between the two group curves appears small, and the survival patterns of both groups tend to converge at longer observation times.

3.3.3. Kaplan Meier Analysis Based on Age

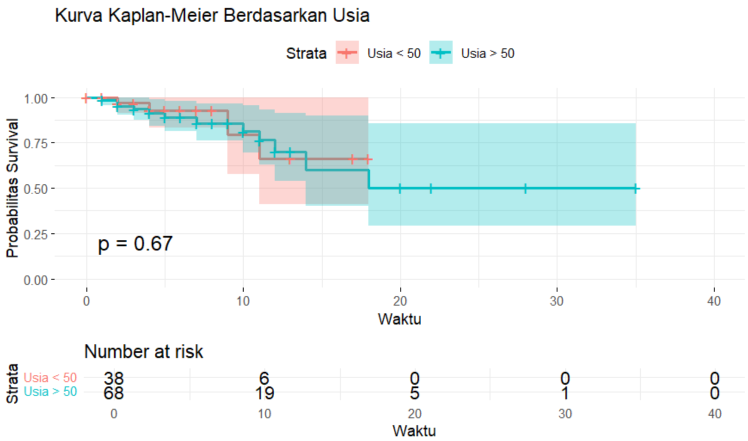

Figure 4. Kaplan Meier Curve for Age Factor

Based on Figure 4, the Kaplan Meier curve for age groups shows that patients aged ≤ 50 years have a higher survival probability compared to patients aged > 50 years. The survival curve for patients aged ≤ 50 years is above the curve for patients aged > 50 years, meaning younger patients have a lower risk of death. On day 12 of observation, the survival probability for patients aged ≤ 50 years is still around 70.1%, whereas for patients aged > 50 years, it has decreased to 50.1%.

3.4. Log-Rank Test

The Log-Rank test is used to analyze data between two related groups, with subjects observed in two different conditions. The results of the Log-Rank test analysis are shown in Table 5.

Table 5 Log-Rank Test Results for Each Independent Variable

Variable	Log-Rank		Conclusion
	P-Value	Sig	
Desease Severity	0,96	0,05	Fail to Reject H_0
Gender	0,62	0,05	Fail to Reject H_0
Age	0,67	0,05	Fail to Reject H_0

Based on Table 5, the Log-Rank test results show that there are no significant differences in survival functions based on disease severity, gender, or age. For disease severity, the p-value of 0.96 > α (0.05) indicates that the survival probabilities for acute patients are not significantly different from those with chronic severity. For the gender variable, the p-value of 0.62 > α (0.05) shows that the survival probabilities between male and female patients also do not differ significantly. Similarly, for the age variable, the p-value of 0.67 > α (0.05) indicates that there is no significant difference in survival between patients aged ≤ 50 years and > 50 years. Although the Kaplan-Meier graph visually shows different survival patterns for each variable, the Log-Rank test results confirm that these differences are not statistically significant. This could be due to the sample size or the relatively homogeneous data distribution.

4. Conclusion

This study aimed to analyze factors affecting the survival probability of kidney failure patients using the Kaplan-Meier method and Log-Rank test. Descriptively, patients with acute disease severity had lower survival probabilities compared to those with chronic severity, females had slightly higher survival probabilities than males, and patients aged ≤ 50 years tended to survive longer than those aged > 50 years. However, the Log-Rank test indicated no statistically significant differences in survival functions across these groups. One possible explanation is the limited number of observed events (deaths) during the study period, which may reduce the statistical power of the analysis. These findings suggest that while certain patterns are observable descriptively, further studies with larger samples and more events are necessary to draw stronger and more generalizable conclusions about the factors influencing survival in kidney failure patients.

5. References

[1] D. Novianty, "Gagal Ginjal Akut Sebagai Komplikasi Covid-19: Literature Review," *Indones. J. Nurs. Heal. Sci.*, vol. 2, no. April, pp. 15–24, 2021.

[2] C. J. G. Paath, G. Masi, and F. Onibala, "Study Cross Sectional : Dukungan Keluarga dengan Kepatuhan Hemodialisa pada Pasien Gagal Ginjal Kronis," *J. Keperawatan(JKp)*, vol. 8, no. 1, pp. 106–112, 2020.

[3] E. W. Sary and N. N. Ghina, "Korelasi Antar Dukungan Keluarga, Efikasi Diri dan Kepatuhan Hemodialisa pada Pasien GGK," *J. Nurs. Invent.*, vol. 3, no. 1, pp. 7–15, 2022.

[4] Diyono, S. Ediyono, and B. Santoso, "Edukasi dan Deteksi Dini Faktor Risiko Penyakit Gagal Ginjal dengan Aplikasi Q-Kidney," *Abdimas Kosala*, vol. 2, no. 1, pp. 1–7, 2023.

[5] F. Yuliana and A. Pitayanti, "Hubungan Dukungan Keluarga dengan Kepatuhan Jadwal Menjalani Terapi Hemodialisis pada Pasien Gagal Ginjal Kronik," *J. Pengemb. Ilmu dan Prakt. Kesehat.*, vol. I, no. 2, pp. 39–47, 2022.

[6] N. N. Yudani, K. Lisnawati, and N. L. P. D. Puspawati, "Hubungan Pengetahuan dan Dukungan Keluarga dengan Kepatuhan Pembatasan Asupan Cairan Pasien Gagal Ginjal Kronis yang Menjalani Hemodialisis di RSUD Sanjiwani Gianyar," *J. Nurs. Res. Publ. Media*, vol. 1, no. 3, pp. 133–143, 2022, doi: 10.55887/nrpm.v1i3.22.

[7] R. Tampake and A. D. S. Doho, "Karakteristik Pasien Gagal Ginjal Kronik Yang Menjalani Hemodialisa," *Lentora Nurs. J.*, vol. 1, no. 2, pp. 39–43, 2021.

[8] D. P. Sulistyaningrum, M. C. A. Septianingtyas, and P. Indriani, "Hubungan Self Acceptance dengan Interdialytic Weight Gain Penderita Gagal Ginjal Terminal yang Menjalani Hemodialisa," *J. Rumpun Ilmu Kesehat.*, vol. 2, no. 3, pp. 100–105, 2022.

[9] L. C. Kasih, N. Safuni, Jufrizal, and I. Nurhidayah, "Persepsi Terhadap Vaksinasi COVID-19 Pada Pasien Gagal Ginjal Kronik yang Menjalani Hemodialisis," *J. Ilm. Permas J. Ilm. STIKES Kendal*, vol. 13, no. 3, pp. 1117–1126, 2023.

[10] U. Hasanah, N. R. Dewi, Ludiana, A. T. Pakarti, and A. Inayati, "Analisis Faktor-Faktor Risiko Terjadinya Penyakit Ginjal Kronik Pada Pasien Hemodialisis," *J. Wacana Kesehat.*, vol. 8, no. 2, pp. 96–103, 2023, doi: 10.52822/jwk.v8i2.531.

[11] D. Gayatri, "Mengenal Analisis Ketahanan (Survival Analysis)," *J. Keperawatan Indones.*, vol. 9, no. 1, pp. 36–40, 2005.

[12] M. Muhajir and Y. D. Palupi, "Analisis Survival terhadap Pasien Diare Anak Mengunakan Metode Kaplan Meier dan Uji Log Rank," *Eksakta J. Ilmu-Ilmu MIPA*, vol. 18, pp. 74–84, 2017, doi: 10.20885/eksakta.vol18.iss1.art8.

[13] V. D. Murti, Sudarno, and Suparti, "Kajian Data Ketahanan Hidup Tersensor Tipe I Berdistribusi Eksponensial dan Six Sigma," *J. Gaussian*, vol. 1, no. 1, pp. 241–248, 2012.

[14] A. P. Ramadhani, N. Rahmi, P. Nadira, S. Hidayah, and T. A. H. Kisty, "Analisis Survival Terhadap Pasien Penderita Gagal Ginjal Menggunakan Metode Kaplan Meier," *Indones. Counc. Prem. Stat. Sci.*, vol. 2, no. 1, p. 7, 2023, doi: 10.24014/icopss.v2i1.25323.

[15] I. N. Ilahi, W. Alwi, and A. Sauddin, "Analisis Survival Terhadap Pasien Penderita Penyakit Gagal Ginjal Menggunakan Metode Kaplan-Meier," *J. Mat. dan Stat. serta Apl.*, vol. 12, no. 1, pp. 52–59, 2024.

[16] L. P. Arub and N. Siyam, "Kejadian Penyakit Ginjal Kronik pada Penderita Hipertensi," *HIGEIA (Journal Public Heal. Res. Dev.*, vol. 8, no. 1, pp. 63–73, 2024, doi: 10.15294/higeia.v8i1.68655.

[17] N. R. N. Insyira, R. M. Atok, and I. S. Ahmad, "Pemodelan Waktu Survival Pekerja dengan Menggunakan Regresi Cox Proportional Hazard," 2019. doi: 10.12962/j23373520.v7i2.34697.

[18] J. M. Baroleh, T. B. Ratag, and F. L. F. G. Langi, "Faktor-Faktor Yang Berhubungan Dengan Penyakit Ginjal Kronis Pada Pasien Di Instalasi Rawat Jalan RSU Pancaran Kasih Manado," *Kesmas J. Kesehat. Masy. Univ. Sam Ratulangi*, vol. 8, no. 7, p. 8, 2019.