Mathematical Journal of Modelling and Forecasting
Vol. 3, No. 2, December 2025, pp. 1-8

Survival Analysis of Patients with Chronic Kidney Disease at Dr Achmad Darwis Regional General Hospital Using the Kaplan-Meier Method

Mei Syarah Anis[1,*] and Devni Prima Sari[2]

[1,2] Department of Mathematics, Universitas Negeri Padang, Indonesia
[*] syarahanismei@gmail.com

Abstract. Chronic kidney disease (CKD) is a global health problem with a steadily increasing incidence. In West Sumatra, its prevalence is recorded at 0.2%, with the highest proportion occurring in individuals aged 45–54 years. Although relatively low, CKD still has a substantial impact on patients' quality of life. To date, data on the survival probability of CKD patients at the regional level, particularly in the district where this study was conducted, remain limited and require further investigation. This study aims to determine the overall survival probability of CKD patients and to examine the factors that influence survival, such as age, sex, disease stage, history of diabetes mellitus, hypertension, anaemia, heart disease, and smoking. This study employs survival analysis using the Kaplan-Meier method and the log-rank test to compare differences between groups. The results show that the overall survival probability of the 140 patients declined significantly over the four-year observation period. The Kaplan-Meier analysis revealed a sharp decline in survival probability within the first 12 days of treatment, with only 20.9% of patients remaining alive by day 12. Based on the log-rank test, the factors significantly associated with survival were age, history of diabetes mellitus, hypertension, and heart disease. These findings underscore the importance of early detection and integrated management of comorbidities in clinical practice, as they may help improve survival outcomes and guide healthcare planning for CKD patients in regional settings.

Keywords: Chronic Kidney Disease, Kaplan Meier Method, Log-Rank Test, Survival Analysis.

1 Introduction

In the human excretory system, metabolic waste products such as urea, uric acid, creatinine, excess salts, and water are filtered from the blood, processed, and excreted through excretory organs, primarily the kidneys. The filtration of waste substances takes place in the kidneys, the primary organs of the excretory system. After undergoing a series of processes, the waste materials are converted into urine, which is then transported through the ureters to the bladder, where it is temporarily stored before being excreted from the body during urination [1].

Chronic Kidney Disease (CKD) is a condition characterised by a progressive and irreversible decline in kidney function due to chronic damage to the renal parenchyma. A person is diagnosed with chronic kidney disease (CKD) when kidney damage or dysfunction persists for more than three months [2], characterised by a decline in kidney function of approximately 78–85% or a reduction in the Glomerular Filtration Rate (GFR) to less than 60 ml/min/1.73 m². This means that the kidneys are filtering less than 60 millilitres of blood per minute, adjusted to the standard body surface area of 1.73 square meters. If the GFR continues to decline and falls below 15 ml/min/1.73 m², the condition is classified as End-Stage Renal Disease (ESRD), indicating a severe impairment of kidney function that significantly affects overall organ function [3].

CKD has become a global public health issue, marked by a continuous rise in new cases, incidence rates, and mortality. The prevalence of chronic kidney disease (CKD), based on WHO data 2018, indicates that CKD is a health problem affecting approximately 1 in 10 people worldwide [4]. Additionally, it is estimated that between 5 to 10 million deaths occur annually due to this disease, while around 1.7 million deaths per year are attributed to acute kidney injury [5]. The World Health Organisation (WHO) projects that the number of individuals affected by CKD will increase by 41.4% between 1995 and 2025 [6]. In West Sumatra, the prevalence of CKD is recorded at 0.2% of the total population. The highest incidence occurred in the 45-54-year age group, with a prevalence of 0.79% [7]. In general, CKD is more common in men than in women. In

2013, the prevalence in West Sumatra was recorded at 0.3% in men and 0.2% in women. This figure increased in 2018 to 0.42% in men and 0.35% in women. Although the prevalence rate appears relatively low, CKD remains a critical concern due to its long-term treatment requirements and its substantial impact on patients' quality of life.

CKD is generally caused by intrinsic kidney diseases that are diffuse and long-lasting. In addition, several factors suspected to contribute to the increasing incidence of CKD include smoking habits, the use of analgesics and non-steroidal anti-inflammatory drugs, hypertension, and the consumption of energy supplement drinks [8]. Kidney failure may also be attributed to factors such as age, sex, and a medical history of diseases like diabetes or other metabolic disorders that can lead to decreased kidney function [9].

One of the main challenges in managing CKD is the lack of local data on patient survival, particularly at the district level. Such information is crucial for estimating the survival probabilities of CKD patients in specific regions and analysing the factors that influence them, which can serve as a basis for improving the quality of healthcare services. Beyond its scientific contribution, this study is also expected to provide practical value for hospitals and health insurance providers. To date, there has been limited research specifically examining the survival prospects of CKD patients in areas such as Lima Puluh Kota District.

In the field of healthcare, one of the most commonly used statistical methods for assessing patient survival is survival analysis, where the outcome variable of interest is the time until a specific event occurs. This event typically refers to a failure condition or endpoint, such as death or disease recurrence [10]. One approach to estimating the survival function, applicable to both censored and uncensored data, is the Kaplan-Meier method [11], which was first introduced by Kaplan and Meier in 1958 [11]. To compare survival functions across two or more groups, the log-rank test is employed, aiming to determine whether differences in Kaplan-Meier curves are statistically significant [10]. The Kaplan-Meier method, also known as the Product-Limit Method, is commonly used in the field of health sciences. It is suitable for analysing censored data, requiring relatively simple calculations, and is particularly well-suited for small sample sizes. Survival data are typically subject to censoring when study subjects are lost to follow-up, withdraw from the study, or have not experienced the defined event by the end of the study period [12].

Moreover, the Kaplan-Meier method provides a graphical representation of the survival distribution. Ilahi [13] reported that the survival probability of patients with kidney failure was 54% based on Kaplan-Meier analysis, while Hidayat [14] compared the Kaplan-Meier and Life Table methods and found that the Kaplan-Meier method was more accurate, as it accounted for survival time precisely, rather than using interval classes. Nevertheless, both studies were limited to methodological aspects without further examining the influence of clinical factors and patient characteristics on survival outcomes. This highlights the need for research that not only estimates survival probability but also explores the factors that may affect the survival of patients with kidney failure. Therefore, this study aims to determine the survival probability of patients with kidney failure using the Kaplan-Meier method and to identify the factors that influence their survival.

2 Research Methods

The type of data used in this study is secondary data. The data was obtained directly by the researchers from the medical records department of Dr Achmad Darwis Regional General Hospital and consists of medical records of inpatients admitted between 2021 and 2024. The patients selected were patients who were treated between 2021 and 2024 and had a history of diabetes mellitus, heart disease, anaemia, and a history of smoking based on previous diagnoses.

Let T be a positive random variable representing the survival time of each subject, such that the possible values for T are $T \geq 0$. If T is a positive random variable defined on the interval $[0, \infty)$, then $F(t)$ denotes the cumulative distribution function (CDF) of T, representing the probability that an event occurs at or before time t. Based on the definition of cumulative probability, the survival function $S(t)$ can be expressed as [15]:

$$\begin{aligned} S(t) &= P(T > t) \\ &= 1 - P(T \leq t) \\ &= 1 - F(t) \end{aligned} \quad (1)$$

The data analysis technique refers to the steps taken to resolve the research problem. The analysis stages conducted in this study are as follows:

1) Collecting inpatient medical record data of kidney failure patients at Dr. Achmad Darwis Regional Hospital during the period 2021–2024. The event investigated from these data is the condition when a kidney failure patient is declared deceased during hospitalisation.
2) Constructing distribution tables based on clinical and demographic characteristics, as well as distribution tables for each factor presumed to influence survival, to describe the survival patterns of patients at each observation time point.
3) Censoring is conducted using Type I censoring. Given n objects: $x_1, x_2, \dots, x_n$ and an observation period of T time units, during the observation period r survival times will be obtained, denoted as $t_1 \leq t_2 \leq \cdots \leq t_r \leq T$. The observation ends at time T with $n - r$ data points that do not experience the event [16].
4) Constructing survival tables for all kidney failure patients and for each suspected influencing factor. These tables include the time of the event (t_j), number of individuals at risk (n_j), number of events (d_j), estimated survival $\left(\hat{S}(t_j)\right)$.

   The estimate for surviving up to time *t* can be written as Equation [16]:

$$\hat{S}(t) = \begin{cases} 1 & if\ t < t_j \\ \prod_{t_j \leq t} \left[1 - \frac{d_j}{n_j}\right] & if\ t_j \leq t \end{cases} \quad (2)$$

   d_j: represents the number of events (failures) at time t_j
   n_j: denotes the number of individuals at risk just prior to time t_j.
   To perform inference about $S(t)$ using the Kaplan-Meier estimator $\hat{S}(t)$, it is necessary first to calculate the standard error or the variance of $S(t)$. The variance of the Kaplan-Meier estimator $\hat{S}(t)$ is commonly referred to as Greenwood's formula [16]:

$$var\left[\hat{S}(t)\right] = \hat{S}(t)^2 \sum_{t_j \leq t} \frac{d_j}{n_j\left(n_j - d_j\right)} \quad (3)$$

   To obtain the standard error of the Kaplan-Meier survival function estimate, the square root of the variance given by Greenwood's formula is used [16], which is:

$$Std.Err\left(\hat{S}(t)\right) = \sqrt{Var\left(\hat{S}(t)\right)} \quad (4)$$

5) Plotting Kaplan-Meier survival curves for all kidney failure patients, as well as by factors suspected to affect survival.
6) Performing log-rank tests to compare survival curves across different groups for each suspected influencing factor, using R software. The log-rank test is a statistical method used to compare two or more survival functions and to determine whether the Kaplan-Meier survival curves for two or more groups are statistically equivalent [10].

   The hypotheses used in the log-rank test are as follows:
   H_0: There is no difference in survival curves between the groups.
   H_1: There is a difference in survival curves between the groups.
   The test used in this log-rank test is the Chi-Square test [16]:

$$X^2{}_{calculated} = \sum_{i=1}^{G} \frac{\left(O_i - E_i\right)^2}{E_i} \quad (5)$$

   O_i: the observed number of events (failures) in each group.
   E_i: the expected number of events (failures) in each group.

To calculate E_i, let m_{ij} represent the number of failure events at time t_j, and n_{ij} denote the number of patients still at risk in group i just before time t_j. Thus, the calculation is as follows:

$$0_i - E_i = \sum_{j=1}^{h} (m_{ij} - e_{ij}) \quad (6)$$

With

$$e_{ij} = \left(\frac{n_{ij}}{\sum_{i=1}^{G} n_{ij}}\right)\left(\sum_{i=1}^{G} m_{ij}\right) \quad (7)$$

Where:

G : the number of groups
0_i : the observed number of failures in group i
E_i : the expected number of failures in group i
m_{ij} : the number of subjects who died in group i at time $t_{(j)}$
n_{ij} : the number of subjects at risk of death in group i at time $t_{(j)}$
e_{Ij} : the expected number of failures in group i at time $t_{(j)}$

The decision rule is to reject H_0 for group i if $\chi^2_{(calculated)} > \chi^2_{(critical)}$ or if the *p-value* < α = 0,05. In this study, the rejection of H_0 is based on the *p-value*. If the *p-value* is less than α (0.05), then H_0 is rejected, indicating that the observed data provide sufficient evidence to conclude that the survival curves differ significantly from what is stated under H_0

7) Concluding.

3 Results and Discussion

3.1 Data Description

The data analysed in this study consist of records of kidney failure patients who were hospitalised at Dr Achmad Darwis Regional General Hospital and discharged either alive or deceased.

Table 1. Characteristics of Patients with Kidney Failure

Symbol	Category	Description	Status: Alive (0)	Status: Deceased (1)	Total
X_1	Age	0 = ≤ 40 years	12	0	12
		1 = > 40 years	82	46	128
X_2	Sex	0 = Male	50	22	72
		1 = Female	44	24	68
X_3	Stage	0 = Early stage (1, 2, and 3)	2	1	3
		1 = Advanced stage (4,5, and 9)	92	45	147
X_4	Diabetes Mellitus	0 = No history	63	20	83
		1 = With history	31	26	57
X_5	Hypertension	0 = No history	20	2	22
		1 = With history	74	44	118
X_6	Anemia	0 = No history	40	23	63
		1 = With history	54	23	77
X_7	Heart Disease	0 = No history	36	6	42
		1 = With history	58	40	98
X_8	Smoking	0 = Non Smoker	67	24	91
		1 = Smoker	27	22	49

Patients who died were coded as 1, while those who were still alive at the end of the study period or who discontinued or transferred treatment during the study were considered censored data and coded as 0. The study utilised data collected from 2021 to 2024, comprising a total of 140 patients, of whom 46 were reported deceased.

3.2 Data Analysis

3.2.1 Overall Kaplan-Meier Survival Curve and Log-Rank Test

Descriptive analysis using the Kaplan-Meier survival curve is employed to provide an overview of the general survival patterns of patients with kidney failure. After grouping the data at each observation time point, a distribution table of the number of surviving and deceased patients was obtained as follows.

Table 2. Overall Patient Distribution Table

t_j	n_j	d_j	Censored Data
0	140	4	4
1	132	10	7
2	115	9	10
3	96	9	13
4	74	3	20
5	51	3	18
6	30	4	7
7	19	1	8
8	10	1	3
9	6	1	3
12	2	1	0
13	1	0	1

Before describing the Kaplan-Meier survival curve, the probability of survival of renal failure patients for four years was first calculated. At t_{13}, it is not necessary to calculate the survival estimate because no patients experienced an event or are called censored data.

Table 3. Kaplan-Meier Survival Estimates for Kidney Failure Patients

t_j	n_j	d_j	$\hat{S}(t_j)$	Standard Error	Lower Bound	Upper Bound
0	140	4	0.971	0.0141	0.9442	0.999
1	132	10	0.898	0.0259	0.8485	0.950
2	115	9	0.828	0.0328	0.7657	0.894
3	96	9	0.750	0.0386	0.6780	0.830
4	74	3	0.720	0.0408	0.6439	0.804
5	51	3	0.677	0.0451	0.5943	0.772
6	30	4	0.587	0.0574	0.4845	0.711
7	19	1	0.556	0.0622	0.4467	0.692
8	10	1	0.500	0.0769	0.3703	0.676
9	6	1	0.417	0.0995	0.2613	0.666
12	2	1	0.209	0.1556	0.0483	0.900

Based on the results shown in Table 3, the estimated survival probability at time $t = 12$ is $\hat{S}(12) = 0.209$. indicating that only 20.9% of kidney failure patients are expected to survive up to day 12. This estimate is calculated with a 95% confidence interval. Ranging from a lower bound of 0.0483 to an upper bound of 0.900. The total Kaplan-Meier survival curve for individuals with kidney failure is shown in Fig 1. During the first 12 days, a noticeable drop in the survival curve is observed. indicating a crucial time frame with a high risk of death right after starting treatment. The existence of severe comorbidities may be the cause of this sharp decline—limited access to prompt and appropriate therapy, or late-stage disease at diagnosis. To improve short-term survival outcomes and gain a better understanding of the underlying causes of early mortality, further research on these aspects is necessary.

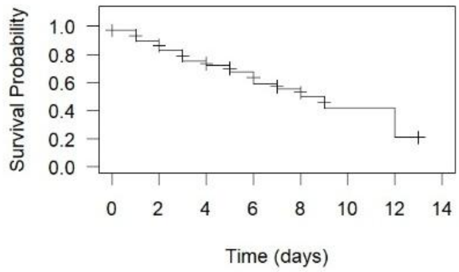

Fig 1. Survival Curve of Patients with Kidney Failure

3.2.2 Kaplan-Meier Survival Curves by Factors and Log-Rank Test

Several factors are thought to affect the survival of patients with kidney failure, such as age, gender, disease stage, and medical histories including diabetes mellitus, hypertension, anaemia, heart disease, and smoking. The Kaplan–Meier analysis revealed significant variations in patient survival across several of these variables.

1. Age

Patients under the age of 40 had a higher and more stable survival probability throughout the observation period compared to those aged 40 and above. At month 12, the survival probability for patients aged ≥ 40 years was only 18.2% (95% CI: 0.047–0.813). The log-rank test indicated a statistically significant difference between the two age groups (χ^2 = 5.3. df = 1. $p - value$ = 0.02).

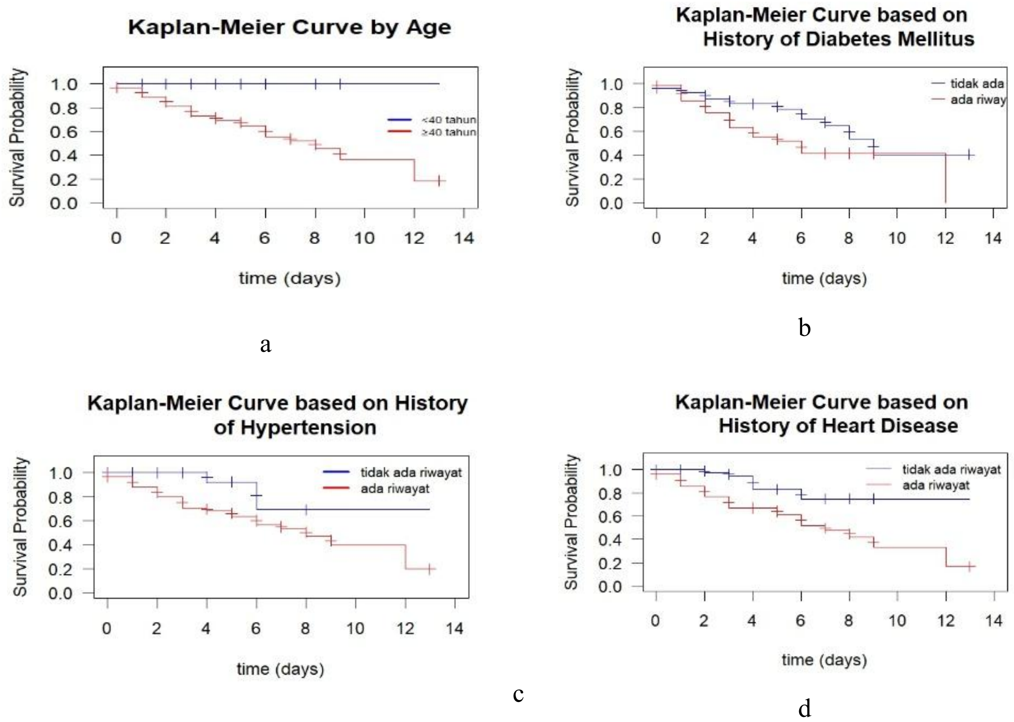

Fig 2. Kaplan-Meier survival curves stratified by (a) age. (b) history of diabetes mellitus. (c) history of hypertension. and (d) history of heart disease.

2. History of Diabetes Mellitus

Patients without a history of diabetes had better survival outcomes (40.4% at month 9; 95% CI: 0.197–0.828) compared to those with a history of diabetes. All of whom had died before the end of the first month. The log-rank test showed a significant difference (χ^2 = 7.4. df = 1. $p - value$ = 0.01).

3. History of Hypertension

Patients without a history of hypertension had a higher survival probability (69.2% at month 6; 95% CI: 0.385–1.000) than those with a history of hypertension (19.7% at month 12; 95% CI: 0.0454–0.857). The log-rank test revealed a significant difference (χ^2 = 4.3. df = 1. $p - value$ = 0.04).

4. History of Heart Disease

Patients without a history of heart disease had a higher survival probability (74.7% at month 6; 95% CI: 0.573–0.974) compared to those with a history of heart disease (16.7% at month 12; 95% CI: 0.0369–0.754). The log-rank test indicated a significant difference (χ^2 = 7.6. df = 1. $p - value$ = 0.01).

4 Conclusion

Based on the Kaplan-Meier analysis using R software, the overall survival probability of patients with kidney failure at Dr. Achmad Darwis Hospital from 2021 to 2024 showed a significant decline throughout the observation period. However, the proportion of patients discharged alive (67.1%) was higher than that of those who died (32.9%). the Kaplan-Meier analysis revealed that the survival probability was very low. with a sharp decrease observed within the first 12 days of hospitalisation. leaving only 20.9% survival by day 12. This finding indicates the initial phase of hospitalisation. particularly the first two weeks. represents a critical period that significantly determines the survival outcomes of patients with kidney failure. The Kaplan-Meier survival analysis and log-rank test identified several significant factors influencing the survival of kidney failure patients, including age ≥ 40 years, and medical histories of diabetes mellitus. hypertension. and heart disease. The non-significance of sex. disease stage. history of anaemia. and smoking history in influencing the survival of kidney failure patients in this study is likely attributable to the limited sample size. the unequal distribution of factors. and the dominant effects of other stronger clinical variables.

This study emphasises the importance of hospitals in strengthening early screening and integrated care for high-risk patients with kidney failure, as well as for health insurers to adjust their financing policies and utilise survival data for long-term planning, thereby improving outcomes and ensuring budget sustainability. Future studies are advised to apply multivariate regression methods. such as Cox regression, to better quantify the effect of each factor on survival. Including additional variables (e.g, nutritional status, dialysis adherence, social support, and lifestyle) and expanding the sample size. observation period. and study sites would improve the comprehensiveness. external validity. and generalizability of the findings.

References

[1] T. Nababan. "Pengaruh Musik Instrumental Terhadap Kualitas Hidup Menjalani Hemodialisa." *J. Keperawatan Priority. Vol 4. No. 1. Januari 2021 ISSN 2614-4719*. vol. 4. no. 1. pp. 125–134. 2021.

[2] A. Damayanti. "Asuhan Keperawatan Pada Tn.S Dengan Chronic Kidney Disease (CKD) di Ruang Baitul Izzah 1 Rumah Sakit Islam Sultan Agung Semarang." 2022. [Online].

[3] S. A. Makmur. M. Madania. and N. Rasdianah. "Gambaran Interaksi Obat Pada Pasien Gagal Ginjal Kronik Dalam Proses Hemodialisis." *Indones. J. Pharm. Educ.*. vol. 2. no. 3. pp. 218–229. 2022. doi: 10.37311/ijpe.v2i2.13333.

[4] A. Harisa *et al.*. "Penanganan Kecemasan Pada Pasien Chronic Kidney Disease Dengan Terapi Spiritual Emotional Freedom Technique (Seft) di Ruang Hemodialisa Rsptn Universitas Hasanuddin." *J. Pengabdi. Masy. Med.*. pp. 125–132. 2023. doi: 10.23917/jpmmedika.v3i2.2139.

[5] E. Syahputra. "Dukungan Keluarga Berhubungan Dengan Kualitas Hidup Pasien Gagal Ginjal Kronik Yang Menjalani Terapi Hemodialisa." *J. Penelit. Perawat Prof.*. vol. 4. no. 3. 2022. doi: 10.33084/jsm.v9i3.6463.

[6] R. Melianna and W. Wiarsih. "Hubungan Kepatuhan Pembatasan Cairan Terhadap Terjadinya Overload Pada Pasien Gagal Ginjal Kronik Post Hemodialisa di Rumah Sakit Umum Pusat Fatmawati." vol. 3. no. 1. pp. 37–46. 2019.

[7] Riskedas RI. "Laporan Riskesdas 2018 Nasional.pdf." 2018.

[8] A. F. Rahmadani. N. Rahmi. P. Nadira. S. Hidayah. T. Amelia. and H. Kisty. "Analisis Survival Terhadap Pasien Penderita Gagal Ginjal Menggunakan Metode Kaplan Meier." *Indones. Counc.*

Prem. Stat. Sci.. vol. 2. no. 1. pp. 7–13. 2023.

[9] R. Pranandari and W. Supadmi. “Faktor Risiko Gagal Ginjal Kronik di Unit Hemodialisis RSUD Wates Risk Factors Cronic Renal Failure On.” vol. 11. no. 2. pp. 316–320. 2015.

[10] J. Harlan. Analisis Surviva*l*. vol. 6. no. 1. Jakarta: Gunadarma. 2017. [Online].

[11] M. Muhajir and Y. D. Palupi. “Survival Analysis of Child Patient Diarrhea Using Kaplan Meier Method and Rank Log Test.” *EKSAKTA J. Sci. Data Anal.*. vol. 18. pp. 74–84. 2018. doi: 10.20885/eksakta.vol18.iss1.art8.

[12] F. Putri. B. Martadiputra. P. Studi Matematika. and F. Pendidikan Matematika dan Ilmu Pengetahuan. “Log-Rank Test as A Continuation of the Kaplan-Meier Method in Survival Analysis Using R Language (Case Study: D-penicillamine Treatment on the Probability of Survival in Primary Biliary Cirrhosis (PBC) Patients at the Mayo Clinic).” vol. 21. no. 1. pp. 285–306. 2024. doi: 10.20956/j.v21i1.36073.

[13] Ilahi. Ilham Nur. “Analisis Survival Terhadap Pasien Penderita Penyakit Gagal Ginjal Menggunakan Kaplan-Meier.” vol. 12. no. 8.5.2017. pp. 2003–2005. 2022.

[14] R. Hidayat. “Penggunaan Metode Kaplan-Meier dan Life Table Analisis Survival Untuk Data Tersensor.” *J. Din.*. vol. Vol. 07. N. pp. 1–8. 2016.

[15] S. Tahir. Model Frailty Spasial Survival Untuk Pemetaan Penyakit Demam Berdarah Dengue di RSU Wahidin Sudirohusodo Kota Makassar. 2017.

[16] Y. W. Di Asih I Maruddani. Tarno. Abdul Hoyyi. Rita Rahmawati. Survival Analysi*s*. 2021.