

A COMPARISON OF KAPLAN-MEIER AND TARONE-WARE METHODS IN ANALYZING STUNTING DATA BY GENDER AT PUSKESMAS LEGOK

Naufal Fadhlullah^{1*}, Asthagina Delia Putri², Wiwik Wiyanti³

^{1,2,3} Faculty of Science, Computer, and Mathematics, Matana University, Indonesia

*e-mail: naufal.fadhlullah@student.matanauniversity.ac.id

Article Info:

Received: February 24, 2026

Accepted: May 25, 2026

Available Online: June 2, 2026

Keywords:

Stunting; Kaplan-Meier; Tarone-Ware; Survival Analysis

Abstract: Stunting is one of the problems of chronic growth problem due to the consumption of unnutritious food for a long period, at least the first 1000 days of life. Biological differences between males and females can be a risk factor in the detection of stunting time. The time it took to detect stunting was measured from when a child was born until the first time stunting was officially diagnosed. Toddlers who were not found to be stunted by the end of the study period were considered as censored cases. Therefore, this study was conducted to determine whether there is a significant difference in the detection time of stunting in male and female toddlers at Puskesmas Legok. Comparison of data between groups by sex were analyzed using Kaplan-Meier and Tarone-Ware test. The results of the study showed that there was a significant difference in the detection time of stunting in toddlers based on gender. The average stunting detection time in female toddlers is longer than in male toddlers. The Tarone-Ware test also showed significant results with a Chi-Square value of 7.802 and $p = 0.005$, indicating that the time to stunting discovery is influenced by gender.

1. INTRODUCTION

Stunting is a health problem that inhibits growth, characterized by a child's physical and cognitive development that falls below the standard for their age group. This condition stems from deficiency caused by the feeding of food that is not in accordance with nutritional needs [1], manifesting over a long period of time, particularly during the first 1,000 days of life (HPK) [2]. Beyond affecting physical height, stunting also impairs cognitive abilities and increases the risk of contracting diseases in the future [3].

The global burden of stunting remains substantial, affecting more than 149 million children under five worldwide [4]. In Indonesia, data from the 2024 Indonesian Nutrition Status Survey recorded a national stunting prevalence of 19.8%, a figure that sits well above the target of 14.2% set by the Medium-Term Development Plan (RPJMN), which was assembled in coordination with the Vice President's Secretariat and Bappenas [5]. This figure places the country far from the long-term national goal of reducing stunting to 5% by 2045, underscoring the urgency of strengthening detection and prevention efforts at all levels of the health system. Stunting remains a persistent obstacle in efforts to create superior human resources, given that its long-term impact not only on health conditions but also on social and economic outcomes and adult productivity [6].

Among the factors associated with stunting risk, nutritional intake in the first 1,000 days of life has been consistently identified as a primary driver. Research by [7] on stunting incidence among toddlers in Bandung confirmed that the risk of stunting is strongly influenced by children's nutritional intake during this critical window. A separate study by [8] at the X health center in Banyuanyar Village demonstrated that exclusive breastfeeding can meaningfully reduce the risk of stunting in toddlers. Beyond nutritional factors, biological differences between males and females may also constitute a risk factor in the manifestation of stunting. Physiological and hormonal differences between the sexes may influence growth trajectories and metabolic responses to nutritional deficiency, suggesting that male and female toddlers may differ not only in whether stunting occurs, but also in when it first becomes clinically detectable.

Understanding when stunting is first detected, rather than merely whether it occurs, is therefore of practical importance for early intervention programs. In this study, stunting detection time is operationally defined as the age of the toddler at the time of first confirmed stunting diagnosis, measured in months from birth to the point at which the child's height-for-age index fell below the stunting threshold during routine health monitoring at Puskesmas Legok. Analyzing this time-to-event outcome requires a statistical framework, capable of modeling the elapsed time from a defined origin, until the occurrence of a specified event, which in this study is the first clinical detection of stunting. Toddlers who had not been detected as stunted by the end of the integrity of the time-based analysis [9], statistical analysis more broadly contributes to the identification, understanding of dynamics, and evaluation of health interventions [10], and survival analysis in particular enables examination of differences in the timing of stunting onset across subgroups.

Within survival analysis, the Kaplan-Meier method is a standard non parametric estimator that describes research objects based on the time contained in each observation, producing graphical visualizations of the probability of remaining free from the event of interest across a time span [9]. Published by Kaplan and Meier in 1958, this method modifies the survival function to handle data with varying sample sizes while generating interpretable survival curves. However, the Kaplan-Meier method produces only a visual illustration and does not itself provide a formal test of whether differences between groups are statistically significant [11]. The Tarone-Ware test complements this limitation by serving as a hypothesis testing tool that assigns greater weight to early time points in the survival curve, making it particularly appropriate when group differences are expected to emerge early in the observation period [12]. This combination allows both the visualization and formal statistical comparison of survival patterns between groups.

Previous research has demonstrated the utility of combining Kaplan-Meier and Tarone-Ware analyses in health settings. A study by [12] assessed factors influencing the response time of complaints by medical scheme members at the Council for Medical Schemes in South Africa, identifying analysts who took longer to resolve complaints and establishing that the nature of the complaint is an important determinant of responses time. Another study by [6] suggests that future research should increase data volume and consider machine learning approaches to improve accuracy. These applications demonstrate that the Kaplan-Meier method provides meaningful time-based insights in health research contexts. However, the application of these methods to stunting data, particularly for gender-based comparisons at the primary health care level remains limited.

This study therefore uses data from toddler records at Puskesmas Legok to analyze whether a statistically significant difference exists in the time to stunting detection between

male and female toddlers, and to compare survival patterns between the two groups using the Kaplan-Meier method and the Tarone-Ware test. The hypotheses of this study are as follows:

H0: There is no significant difference in the detection time of stunting between male and female toddlers.

H1: There is a significant difference in the detection time of stunting between male and female toddlers.

2. LITERATURE REVIEW

2.1. Survival Analysis

Survival analysis is a collection of statistical methods used to analyze time-to-event data, where the primary outcome of interest is the elapsed time from a defined starting point until the occurrence of a specified event[13]. More specifically, survival analysis examines variables that may influence the transition from the beginning of an observation period to the occurrence of an event, providing results in the form of time-based probability estimates[14].

In survival analysis, the survival function $S(t)$ represents the probability that a subject survives, that is has not yet experienced the event of interest beyond time t , and is expressed as follows:

$$S(t) = P(T > t) = 1 - F(t) \quad (1)$$

Where T is the survival time of a subject, $S(t)$ is the probability that survival time exceeds t , and $F(t)$ is the cumulative distribution function of T [14][15]. In addition to the survival function, the hazard function $h(t)$ describes the instantaneous rate at which an event occurs at time t , given that the subject has survived up to that point. The hazard function over a time interval from t to $t + \Delta t$ is formally defined as:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t | T \geq t)}{\Delta t} \quad (2)$$

The hazard function $h(t)$ focuses specifically on the risk of failure or event occurrence at a given moment, and provides information that is complementary to, but distinct from the survival function[14]. The relationship between the survival function $S(t)$, the hazard function $h(t)$, and the probability density function $f(t)$ can be expressed as:

$$h(t) = \frac{f(t)}{S(t)} \quad (3)$$

2.2. Kaplan Meier

The Kaplan-Meier method is a non-parametric estimation technique used to estimate the survival function from time-to-event data, and is particularly suitable for describing the survival between two or more population [16]. The Product-Limit Estimator, proposed by Kaplan and Meier in 1958, remains the standard approach for estimating survival probabilities at each observed failure time[17]. The Kaplan-Meier survival probability at the j -th failure time $t_{(j)}$ is defined as:

$$\hat{S}_{(t_{(j)})} = \hat{S}_{(t_{(j-1)})} \times \hat{Pr}(T > t_{(j)} | T \geq t_{(j)}) \quad (4)$$

This recursive formula states that the probability of surviving beyond $t_{(j)}$ equals the probability of surviving beyond the previous failure time $t_{(j-1)}$, multiplied by the conditional probability of surviving past $t_{(j)}$ given survival up to that point. Extending this across all observed failure times yields the Product-Limit form of the estimator:

$$\hat{S}_{(t_{(j-1)})} = \prod_{i=1}^{j-1} \hat{Pr}(T > t_{(i)} | T \geq t_{(i)}) \quad (5)$$

The complete Kaplan-Meier estimator can thus be written as:

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_{(1)} \\ \prod_{j:t_{(j)} \leq t} (1 - \frac{d_j}{n_j}) & \text{if } t \geq t_{(1)} \end{cases} \quad (6)$$

Where:

d_j : the number of events (stunting detections) occurring at time $t_{(j)}$

n_j : the number of individuals at risk immediately before time $t_{(j)}$ [15].

2.3 Tarone Ware Test

The Tarone-Ware test is a weighted log-rank test used to formally compare survival functions between two or more groups [11]. It shares structural similarities with the Wilcoxon test, in that both assign greater weight to early time points in the observation period. However, while the Wilcoxon test weights each time point by the total number of individuals at risk n_j , the Tarone-Ware test uses the square root of the number at risk as its weight function, placing it intermediate between the Log-Rank test and the Wilcoxon test in terms of early-time sensitivity [11]. According to Tarone and Ware (1977), Klein et al. (2001), Kleinbaum and Klein (2005), and Allison (2010) as cited in [11], the weight assigned at each failure time in the Tarone-Ware test is:

$$\omega_j = \sqrt{r_j} \quad (7)$$

Where r_j denotes the number of individuals at risk at time $t_{(j)}$. the Tarone-Ware test statistic is then computed as:

$$Tarone - Ware = \frac{\left(\sum_j \sqrt{r_j} \left(d_{1j} - d_j \frac{r_{1j}}{r_j} \right) \right)^2}{\sum_{j=1}^k \frac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}} \quad (8)$$

Where:

d_{1j} = observed number of events in group 1 at time $t_{(j)}$

$\hat{e}_{1j} = d_j \frac{r_{1j}}{r_j}$ = expected number of events in group 1 under the null hypothesis

r_{1j}, r_{2j} = number at risk in group 1 and group 2 at time $t_{(j)}$

d_j = total events across both groups at time $t_{(j)}$

r_j = total individuals at risk at time $t_{(j)}$

The weight $\sqrt{r_j}$ ensures that the Tarone-Ware test is more sensitive to differences that emerge early in the follow-up period compared to the Log-Rank test, while being less dominated by early observations than the Wilcoxon test [11]. This property makes the

Tarone-Ware test appropriate for this study, as gender-based differences in stunting detection are expected to be more pronounced in the earlier months of a toddler's life.

3. METHODOLOGY

3.1 Research Procedure

This study employs a quantitative approach using secondary data obtained from Puskesmas Legok. The dataset represents a cross-sectional record of toddlers who were confirmed as stunted based on anthropometric measurements conducted during the May 2025 posyandu activity period. Cross-sectional data was chosen because this study focused on observing the stunting condition of toddlers at a single point in time, without examining changes in their condition over time [12]. Since all 569 observations in the dataset consist of toddlers who had already been confirmed as stunted at the time of measurement, this study does not model the probability of stunting occurrence in the classical survival analysis sense. Instead, the Kaplan-Meier method is applied as a descriptive tool to characterize and compare the distribution of age at stunting detection across gender groups that is, to examine whether male and female toddlers differ in the age at which stunting was first identified through routine anthropometric screening. In this framework, the age of each toddler at the time of confirmed stunting detection serves as the primary time variable (t), and all observations are treated as complete events ($\delta = 1$).

The Tarone-Ware test is subsequently applied to formally assess whether the observed difference in age-at-detection distributions between male and female toddlers is statistically significant. The total dataset consisted of 569 toddlers 335 male and 234 female, all of whom had been confirmed as stunted during the study period, with analysis focused on three variables: age at measurement as the time variable (t), stunting confirmation as the event indicator (δ), and gender as the group differentiating variable.

3.2 Data Collection

The data collection technique in this study uses secondary data documentation from Puskesmas Legok. The data represents a cross-sectional snapshot of toddlers who participated in posyandu activities and were confirmed as stunted during the May 2025 measurement period. The data was recorded by health workers through the posyandu routine recording system and contained the basic identity of toddlers, nutritional status classifications, anthropometric measurement results, and related supporting variables. The use of this data was conducted with the permission of Puskesmas Legok, and all data were handled in accordance with applicable confidentiality standards to protect the identity of individual toddlers.

The data collection process was carried out by accessing and organizing the official dataset from the health center to suit the needs of the research. This stage included checking the completeness of variables, verifying the consistency of the data, and selecting the variables used in the analysis: age at the time of stunting detection as the time variable, stunting confirmation as the event indicator, and gender as the group differentiator. This approach of data collection through documentation was chosen because it is considered

appropriate to obtain standardized numerical data and allow for further statistical analysis. In addition, the use of secondary data from health facilities provides an objective record of stunting conditions among toddlers in the Puskesmas Legok service area.

3.3. Data Analysis

The data obtained through documentation from Puskesmas Legok were processed through several sequential stages. The first stage was data cleaning, which included checking the completeness of the data, handling missing values, and checking for inconsistencies in the variables used. Data that did not meet the analysis criteria or contained recording errors were excluded so as not to affect the results. The second stage was the coding and selection of variables. The age-at-measurement variable was designated as the time variable (t), representing the age in months at which stunting was first confirmed for each toddler. Since all toddlers in the dataset had been confirmed as stunted at the time of measurement, the event indicator (δ) was set to 1 for all observations, indicating that the event of interest stunting detection had occurred for every subject. The coding of $\delta = 0$ as "toddlers who did not experience the event" is therefore not applicable in this dataset, as the data exclusively comprises toddlers with confirmed stunting status. The gender variable was converted into a categorical variable for group comparison purposes. All variables were then adjusted to the requirements of the analysis.

The third stage was comparative analysis using the Kaplan-Meier method and the Tarone-Ware test. The Kaplan-Meier estimator was used as a descriptive nonparametric tool to visualize and compare the distribution of age at stunting detection between male and female toddler groups, rather than to model survival probabilities in the classical sense. To formally test whether the difference in age-at-detection distributions between the two gender groups was statistically significant, the Tarone-Ware test was selected. The Tarone-Ware test was chosen over the Log-Rank test and the Breslow (Wilcoxon) test because its weighting function $\sqrt{r_j}$ at each time point provides greater sensitivity to differences that emerge early in the observation period while maintaining balance relative to later time points. This property is appropriate for this study, as gender-based differences in stunting detection age are expected to be more pronounced in the earlier months of a toddler's life.

4. RESULTS AND DISCUSSION

This data used in this study were analyzed descriptively to provide an overview of the distribution and central tendency of the key variables. Since the dataset consists entirely of toddlers confirmed as stunted, all 569 observations are complete events. The descriptive statistics for the time variable, age at stunting detection in months, as well as the distribution of gender and event status are presented in **Table 1** below,

Table 1. Descriptive Statistics

	Age at Stunting Detection
Min	0.2
1 st quartile	20.77

Median	31.02
Mean	30.77
3 rd quartile	40.62
Max	60.03
Varians	186.69
Gender (n%)	
Male	335 (58.9%)
Female	234 (41.1%)
Event Status	
Stunting confirmed ($\delta = 1$)	569 (100%)
Censored ($\delta = 0$)	0

The table shows the spread and concentration of the age-at-detection variable. The mean age at stunting detection across all toddlers was 30.77 months, with values ranging from 0.02 to 60.03 months, indicating considerable variability in the age at which stunting was first identified. The dataset comprised 335 male toddlers (58.9%) and 234 female toddlers (41.1%), all of whom had been confirmed as stunted at the time of measurement.

Distribusi Jenis Kelamin

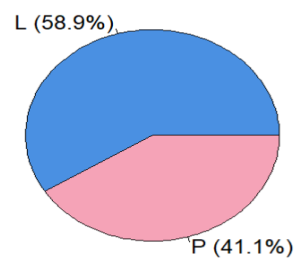


Fig 1. Sex Distribution Pie Chart

A comparative analysis of age-at-detection distributions between gender groups was conducted using the Kaplan-Meier method as a descriptive nonparametric tool. The resulting Kaplan-Meier curves are presented in **Figure 2**.

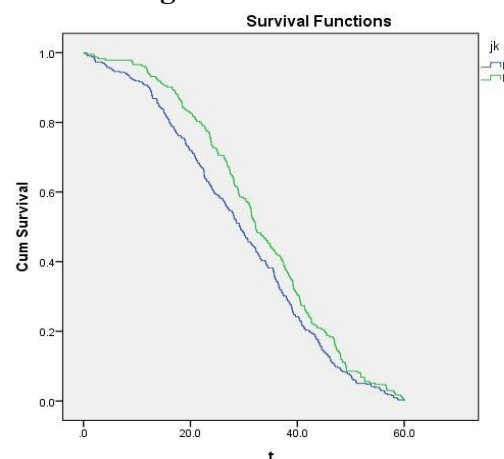


Fig 2. Kaplan-Meier Graph

The Kaplan-Meier curves indicate that male toddlers tend to reach the stunting detection point at earlier ages compared to female toddlers. This pattern suggest that, within this dataset, stunting was identified at a younger age in male toddlers on average. It should be noted, however, that this finding reflects differences in the *age at which stunting was detected* during routine posyandu screening, and does not necessarily indicate differences in the biological onset of stunting or in the underlying risk of stunting between genders.

To formally assess whether the observed difference in age-at-detection distributions between amle and female toddlers was statistically significant, the Tarone-Ware test was applied. The results are presented in **Table 2**.

Table 2. Tarone-Ware test results

	Chi-Square	df	Sig.
Tarone-Ware	7.802	1	0.005

Table 2 shows that the Tarone-Ware test produced a Chi-Square value of 7.802 with a significance value of 0.005, which is below the 5% significance level. This result indicates rejection of H_0 , confirming that there is a statistically significant difference in the age at stunting detection between male and female toddler groups.

The mean and median age-at-detection estimates by gender, along with their 95% confidence intervals, are presented in **Table 3**.

Table 3. Mean and Median Table with 95% Confidence Interval by Gender

JK	Mean Estimate	95% CI (mean)	Median Estimate	95% CI (median)
L (laki-laki)	29.359	27.870 – 30.847	29.267	26.926 – 31.607
P (Perempuan)	32.931	31.270 – 34.592	32.167	29.939 – 34.395

The female toddler group shows a higher mean age at detection (32.931 months) compared to the male toddler group (29.359 months), a difference of approximately 3.6 months. Similarly, the median age at detection was higher in female toddlers (32.167 months) than in male toddlers (29.267 months). These estimates are consistent with the Kaplan-Meier curves and support the Tarone-Ware test result.

These findings indicate that, within the Puskesmas Legok dataset, stunting was detected at a statistically significantly younger age in male toddlers compared to female toddlers. This pattern is consistent with existing literature suggesting that male children are generally more biologically vulnerable to growth faltering in early life due to differences in hormonal regulation and immune response, which may accelerate the manifestation of nutritional deficiency into measurable stunting. From a public health perspective, the earlier age of stunting detection in male toddlers suggests that screening and nutritional intervention programs at Puskesmas Legok may benefit from prioritizing male toddlers for earlier assessment, particularly in the first two years of life.

However, several important caveats must be acknowledged in interpreting these findings. The observed difference in age at detection may reflect not only biological differences in stunting onset between genders, but also differences in posyandu attendance patterns, health-worker screening practices, or parental health-seeking behavior between families of male and female toddlers. Furthermore, because the dataset consists exclusively of toddlers who were already confirmed as stunted, with no non-stunted comparison group, the analysis cannot be interpreted as modeling the probability or risk of stunting occurrence. The findings therefore describe differences in when stunting was detected among those who experienced it, rather than differences in whether stunting is more likely to occur in one gender over the other.

5. CONCLUSION

The application of the Kaplan-Meier method to stunting detection data from Puskesmas Legok indicates that male toddlers had a statistically significantly lower age at stunting detection compared to female toddlers. This finding is supported by the Tarone-Ware test result (Chi-Square = 7.802; $p = 0.005$), which confirms a significant difference in the age-at-detection distributions between the two gender groups. The mean age at stunting detection was approximately 3.6 months earlier in male toddlers (29.359 months) than in female toddlers (32.931 months).

These results suggest that gender may be a relevant factor in the timing of stunting detection during routine health monitoring, and that male toddlers at Puskesmas Legok may benefit from earlier and more intensive nutritional screening in the first two years of life. However, it should be noted that this study was conducted using a dataset comprising only toddlers with confirmed stunting status, and the findings should be interpreted as describing differences in age at detection among stunted toddlers, rather than differences in overall stunting risk between genders. Future studies should incorporate data from both stunted and non-stunted toddlers to enable a full survival analysis with appropriate censoring, and should consider multiple locations to allow for broader generalization of findings.

REFERENCES

- [1] K. Rahmadhita, "Permasalahan Stunting dan Pencegahannya," *J. Imiah Kesehat. sandi husada*, vol. 11, no. 1, pp. 225–229, 2020, doi: 10.35816/jiskh.v10i2.253.
- [2] J. Kharisa, "Peran Dinas Kesehatan Dalam Penurunan Stunting Di Kabupaten Kotawaringin Timur Provinsi Kalimantan Tengah," Institut Pemerintahan Dalam Negeri, 2025. [Online]. Available: <http://eprints.ipdn.ac.id/id/eprint/21016>
- [3] E. Lestari, A. Siregar, A. K. Hidayat, and A. A. Yusuf, "Stunting and its association with education and cognitive outcomes in adulthood: A longitudinal study in Indonesia," *PLoS One*, vol. 19, no. 5, pp. 1–18, 2024, doi: 10.1371/journal.pone.0295380.
- [4] PAUDPedia, "149 Juta Anak di Dunia Alami Stunting, Sebanyak 63 Juta di Indonesia; Wapres Minta Keluarga Prioritaskan Kebutuhan Gizi." [Online]. Available: <https://paudpedia.kemendikdasmen.go.id/berita/149-juta-anak-di-dunia-alami-stunting-sebanyak-63-juta-di-indonesia-wapres-minta-keluarga-prioritaskan-kebutuhan-gizi?do=MTY2NC01YjRhOGZkNA==&ix=MTEtYmJkNjQ3YzA=>
- [5] K. BKKP, "SSGI 2024: Prevalensi Stunting Nasional Turun Menjadi 19,8%." [Online]. Available: <https://www.badankebijakan.kemkes.go.id/ssgi-2024->

- prevalensi-stunting-nasional-turun-menjadi-198/
- [6] W. Wiyanti, A. P. Zenklinov, J. S. Seleki, V. Jimy, and S. Arifin, “Stunting Risk Factor Analysis Using Seemingly Unrelated Regression (SUR) Integrated with Machine Learning,” *J. Math. Comput. Stat.*, vol. 8, no. September, pp. 647–657, 2025, doi: 10.35580/jmathcos.v8i2.9630.
- [7] A. Sutriyawan and C. C. Nadhira, “Kejadian Stunting Pada Balita Di UPT Puskesmas Citarip Kota Bandung,” *J. Kesehat. Masy. Khatulistiwa*, 2020.
- [8] S. Sumardiyono, “Pengaruh Usia, Tinggi Badan dan Riwayat Pemberian Asi Eksklusif terhadap Stunting pada Balita,” *J. Ilm. Kesehat.*, vol. 15, no. 1, pp. 1–8, 2020.
- [9] F. P. Damayanti and A. Wibowo, “Analisis Survival Penyakit jantung Koroner Berulang,” *J. Biometrika dan Kependud.*, vol. 6 No.1, pp. 43–51, 2017, doi: <https://doi.org/10.20473/jbk.v6i1.2017.43-51>.
- [10] BARAKA, “Statistika Terapan dalam Ilmu Kesehatan: Studi Kasus Epidemiologi,” *Universitas Medan Area*, Medan, 2024. [Online]. Available: <https://baraka.uma.ac.id/epidemiologi-adalah/>
- [11] P. G. Karadeniz and I. Ercan, “EXAMINING TESTS FOR COMPARING SURVIVAL CURVES WITH RIGHT CENSORED DATA,” vol. 18, no. 2, pp. 311–328, 2017, doi: 10.21307/stattrans-2016-072 EXAMINING.
- [12] M. M. Willie, “Analysis of factors related to complaints in medical schemes,” vol. 3, no. July, pp. 337–345, 2011.
- [13] A. R. Faisal, M. N. Bustan, and S. Annas, “Analisis Survival Dengan Pemodelan Regresi Cox Proportional Hazard Menggunakan Pendekatan Bayesian,” vol. 1, no. 2, pp. 1–8, 2019.
- [14] D. G. Kleinbaum and M. Klein, *Statistics for Biology and Health*, 2nd ed. Springer Science+Business Media, Inc, 2005.
- [15] H. A. Mutiah and J. Purwadi, “Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Splines Pada Kasus Penyakit Ginjal Kronis,” Universitas Ahmad Dahlan, 2019.
- [16] R. Pradika and B. A. P, “Aplikasi Metode Kaplan Meier Sebagai Penduga Ketahanan Hidup Penderita Kanker Payudara,” vol. 9, no. 1, pp. 35–44, 2021, doi: <https://doi.org/10.17509/jem.v7i2.22133>.
- [17] J. P. Klein and M. L. Moeschberger, *SURVIVAL ANALYSIS Techniques for Censored and Truncated Data Second*, 2nd ed. ohio, 2003.