

## Estimation of Survival Function in Head and Neck Cancer Patients Using the Kaplan-Meier Method

Ardi Kurniawan<sup>(1)</sup>, Adelia Frielady Yosifa<sup>(2)</sup>, Azizatul Kholidiyah<sup>(3)</sup>, Vrisca Natalia Putri Wardhani<sup>(4)</sup>

<sup>1,2,3,4</sup>Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Airlangga

Jalan Dr. Ir. H. Soekarno, Mulyorejo, Kec. Mulyorejo, Surabaya, Jawa Timur

e-mail: [ardi-k@fst.unair.ac.id](mailto:ardi-k@fst.unair.ac.id)<sup>(1)</sup>, [adelia.frielady.yosifa-2021@fst.unair.ac.id](mailto:adelia.frielady.yosifa-2021@fst.unair.ac.id)<sup>(2)</sup>,  
[azizatul.kholidiyah-2021@fst.unair.ac.id](mailto:azizatul.kholidiyah-2021@fst.unair.ac.id)<sup>(3)</sup>, [vrisca.natalia.putri-2021@fst.unair.ac.id](mailto:vrisca.natalia.putri-2021@fst.unair.ac.id)<sup>(4)</sup>

### ABSTRAK

Analisis data uji hidup atau analisis survival adalah kumpulan metode statistik yang digunakan untuk menjawab pertanyaan yang berkaitan dengan apakah dan kapan suatu peristiwa terjadi. Data daya tahan hidup yang diperoleh berupa data tersensor tipe I yang dianalisis menggunakan Metode Kaplan-Meier atau Metode Produk Limit. Sampel tersensor tipe I adalah cara pengambilan sampel secara statistik didasarkan pada lama pengamatan yang telah ditentukan sebelumnya. Metode Kaplan-Meier merupakan salah satu metode yang digunakan untuk mengestimasi fungsi survival dalam analisis data uji hidup. Penelitian ini bertujuan untuk mengestimasi daya tahan hidup seorang penderita kanker kepala dan leher. Kanker kepala dan leher adalah jenis kanker yang paling umum di dunia dan terdiri dari berbagai kelompok tumor yang mempengaruhi saluran aerodigestive bagian atas. Penelitian dimulai dengan menentukan data tersensor dan tipe sensor, kemudian menghitung estimasi fungsi survivor. Fungsi survivor merupakan probabilitas individu atau suatu objek tertentu untuk tetap bertahan setelah melewati suatu satuan waktu. Data pada penelitian ini adalah Data Kanker Kepala dan Leher Northern California Oncology Group. Kelompok pasien kanker kepala leher diobati dengan terapi radiasi plus kemoterapi (RCT) dengan 45 pasien. Hasil penelitian menunjukkan bahwa pasien kanker kepala dan leher yang menjalani pengobatan RCT memiliki peluang bertahan hidup semakin kecil seiring berjalannya waktu.

**Kata kunci:** Kanker Kepala; Kanker Leher; Analisis Survival; Metode Kaplan-Meier; Tersensor Tipe I

### ABSTRACT

*Life test data analysis or survival analysis is a collection of statistical methods used to answer questions relating to whether and when an event occurred. The survival data obtained were type I censored data which were analyzed using the Kaplan-Meier Method or the Product Limit Method. Type I censored sample is a statistical sampling method based on a predetermined length of observation. The Kaplan-Meier method is one of the methods used to estimate the survival function in the analysis of survival test data. This study aims to estimate the survival of a patient with head and neck cancer. Head and neck cancer is the most common type of cancer worldwide and comprises a diverse group of tumors that affect the upper aerodigestive tract. The research begins by determining the censored data and the type of sensor, then calculating the estimated survivor function. The survivor function is the probability of an individual or a certain object to survive after passing a certain time. The data in this study is the Head and Neck Cancer Data from the Northern California Oncology Group. The head and neck cancer patient group was treated with radiation therapy plus chemotherapy (RCT) with 45 patients. The results showed that head and neck cancer patients undergoing RCT treatment had a smaller chance of survival over time.*

**Keywords:** Head Cancer; Neck Cancer; Survival Analysis; Kaplan-Meier method; Censored Type I

## INTRODUCTION

Analysis of survival test data is one of the statistical analyzes that is useful for testing the survival of a component or measuring the survival of a patient in the treatment of a disease [1]. In analyzing the survival test data, survival data is needed which includes the survival time and survival time status of the object or individual to be studied. Survival time is recorded as data about the duration of an event from start to finish [2]. Life time data obtained can be in the form of complete samples and incomplete samples (censored). The sample is said to be complete if all the individuals observed during a certain research period experience the desired event (failure). Meanwhile, the sample is said to be incomplete or censored if not all observation units observed during a certain period fail so that the actual survival time of some observations is unknown.

There are two types commonly used of censored samples, namely type I censored samples and type II censored samples [3]. Type I censored sample is a statistical sampling method based on a predetermined length of observation. If the life span of the sample exceeds the time allotted, the observation of the sample will be stopped or a censored sample will be obtained. Survival modeling is divided into two, namely the parametric model and the nonparametric model. The parametric survival model is used if the population distribution assumption is made first, while the nonparametric model is a model that does not depend on the population distribution assumption. There are several ways to estimate the survival function with nonparametric models, including life tables, Kaplan-Meier estimation or the Product Limit method. The Kaplan-Meier method is a method that can be used to estimate the survival function in the analysis of survival test data. The benefit of using this method is that it can provide an estimate of the chances of survival. The Kaplan-Meier method provides a graphical representation of the survival distribution [4].

One of the problems related to survival in the health sector is cancer. Cancer is a disease caused by abnormal growth of body tissue cells that turn into cancer cells [5]. There are several types of cancer including lung cancer, breast cancer, cervical cancer, head and neck cancer and so on. There are several factors that cause cancer including heredity, unhealthy lifestyle, radiation and others. In a previous study conducted by Pradika and Priatna (2019), explained the probability of survival for breast cancer patients with the Kaplan Meier method and it was concluded that the survival time of breast cancer patients who receive chemotherapy treatment has a greater chance of survival compared to those who do not receive chemotherapy [6].

Head and neck cancer is the most common type of cancer worldwide and comprises a diverse group of tumors that affect the upper aerodigestive tract [7]. The risk of this cancer is strongly associated with smoking and alcohol consumption. Head and neck cancer patients have been reported to show high rates of depression.

According to Ashraf-Ul-Alam (2021) in a previous study, head and neck cancer patients were given 2 different treatments, namely radiation therapy (RT) and radiation and chemotherapy (RTC) [8]. The research focuses on comparing the most suitable Bayesian models. It was concluded that the log-normal model fits the data better than the log-logistics and Weibull models.

From previous study, there has been no explanation regarding the estimation of survival chances. Therefore, it is necessary to know the factors that influence the survival of head and neck cancer patients. Adaptations in statistics about the factors that influence the survival time of head

and neck cancer patients can be analyzed using type I censored survival estimation using the Kaplan-Meier method.

Based on the explanation above, the authors are interested in estimating the survival of a patient with head and neck cancer. Estimating the survival of head and neck cancer patients begins by calculating the survivor function or  $S(t)$  and continues with the hazard function or  $h(t)$ . The survivor function is the probability of an individual or a particular object surviving after a certain amount of time has passed [9]. The hazard function is the probability that an individual will reach a particular event at time  $t$ , provided that he has survived until that time. This function is used to express momentary opportunities then, it is known that a system has aged  $t$  [10]. After calculating the two functions, the results will be interpreted. Based on the background above, the authors wanted to conduct research with the title Estimation of Functional Survival in Head and Neck Cancer Patients Using the Kaplan-Meier Method.

## METHOD

This type of research is quantitative research, namely a systematic research on a phenomenon by collecting data that can be measured using statistical, mathematical, or computational techniques [11]. with secondary data. The secondary data used is Type 1 Censored Head and Neck Cancer Data from the Northern California Oncology Group. The method used in this article is the Kaplan-Meier method, which is a nonparametric method used to estimate the survival function in the analysis of time to event data or survival data. The Kaplan-Meier method in head and neck cancer patients is very useful in estimating the probability of survival. This method requires research steps, including:

1. Collection of head and neck cancer patient data that involves information about the time of occurrence (such as death or disease progression) and the patient's status at any given point in time ensuring the data is complete and accurate.
2. Dataset preparation, by compiling a dataset that includes variable time of event, status (event or censoring), and other relevant variables such as patient characteristics or risk factors by ensuring the dataset is ready for analysis in the right format.
3. Conducted a descriptive analysis of type 1 censored head and neck cancer patient data from the Northern California Oncology Group.
4. Calculation of the survival function and hazard function, using statistical software that supports survival analysis to estimate the survival function using the Kaplan-Meier method where this software usually has a special function to calculate the survival function and produce estimates at each observation time point.
5. Visualization of the survival curve, by creating a graph of the survival curve that shows the development of the survival function over time where this graph will visualize the survival rate of head and neck cancer patients during the observation period.
6. Interpretation of the results, by analyzing the Kaplan-Meier and survival curves will provide an estimate of the survival function at each time point by interpreting these results with care, describing the survival rate of head and neck cancer patients and the factors that might affect survival.

**RESULT AND DISCUSSION**

The head and neck cancer patient group was treated with radiation therapy plus chemotherapy (RCT) with 45 patients. The patient's survival data is presented in Table 1 below.

**Table 1.** Survival Data (in days) of Head and Neck Cancer Patients treated with RCTs

Patient	Lifespan (Days)	Patient	Lifespan (Days)	Patient	Lifespan (Days)
1	37	16	173	31	633
2	84	17	179	32	725
3	92	18	194	33	759*
4	94	19	195	34	817
5	110	20	209	35	1092*
6	112	21	249	36	1245*
7	119	22	281	37	1331*
8	127	23	319	38	1557*
9	130	24	339	39	1642*
10	133	25	432	40	1771*
11	140	26	469	41	1776
12	146	27	519	42	1897*
13	155	28	528*	43	2023*
14	159	29	547*	44	2146*
15	169*	30	613*	45	2297*

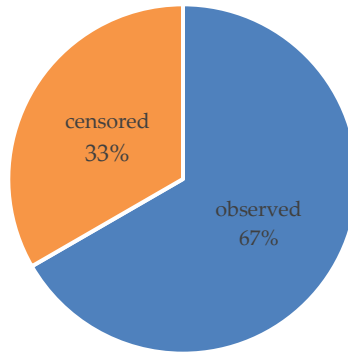
**Descriptive Statistics**

Based on the data in table 1, the descriptive statistics are obtained in table 2 as follows.

**Table 2.** Descriptive Statistics

N Total	Mean	StDev	Min.	Median	Max.
45	639,20	668,859	37	319	2297

Based on Table 2, an average value of 639.2 is obtained with a standard deviation of 668.859. The basis for calculating the standard deviation is desire to know a diversity data group. In addition, the lowest (minimum) data obtained was 37 and the highest (maximum) data was 2297. The median value of 45 head and neck cancer patient data was 319. Data is classified as censored and observed (uncensored) data. There were 45 data with 30 (67%) being observed data and 15 (33%) being censored data. The proportion of censored and observed data is presented graphically in Figure 1 as follows.



**Figure 1.** Graph of the Proportion of Senior and Observed Data

**Normality Assumption Test**

To continue the analysis of type 1 censored data using the Kaplan-Meier method, a normality test is performed first. The Kaplan-Meier method is a type of nonparametric method that can be used to estimate the survival function in the analysis of survival test data. One of the characteristics of the data that can be analyzed using nonparametric methods is that the data is not normally distributed [12]. The results of the normality test on the survival data of head and neck cancer patients are shown in table 3 as follows.

**Table 3.** Kolmogorov-Smirnov Normality Test

N		45
Normal Parameters <sup>a,b</sup>	Mean	639,20
	Std. Deviation	668,859
Most Extreme Differences	Absolute	0,207
	Positive	0,207
	Negative	-0,184
Test Statistic		0,207
Asymp. Sig. (2-tailed)		0,000 <sup>c</sup>

Based on the normality test in table 3 above, it is found that the Asymp. Sig.  $0,000 < \alpha = 5\%$ . It can be concluded that the data is not normally distributed, so it can be continued with type 1 censored data analysis using the Kaplan-Meier method.

**Estimation of Survival and Hazard Functions with the Kaplan-Meier Method**

To estimate the product limit or Kaplan-Meier estimator, it can be calculated in the following way.

1. Column  $t_i$  is the survival of head and neck cancer patients undergoing treatment with radiation plus chemotherapy which includes censored and uncensored data.
2. Column  $R_i$  is the number of living individuals in the  $i$ th observation.
3. Column  $\delta_i$  is the number of individuals that died in the  $i$ -th observation.
4. Column  $C_i$  is the number of individuals censored in the  $i$ -th observation.

5. Column  $q_i$  is an estimate of the individual's probability of death in the  $i$ -th observation. The  $q_i$  column can also be called a hazard function. The calculation of  $q_i$  is as follows. The calculation of  $q_i$  is as follows.

$$q_i = \frac{\delta_i}{R_i} \quad (1)$$

6. Column  $p_i$  is an estimate of the probability of individual survival in the  $i$ -th observation. The calculation of  $p_i$  is as follows.

$$p_i = 1 - \frac{\delta_i}{R_i} \quad (2)$$

7. Column  $S(t)$  is the estimated survival function for each individual included in the observation. The calculation of  $S(t)$  is as follows.

$$S(t) = \prod_{i=1}^{37} p_i \quad (3)$$

Some of Kaplan Meier's estimation calculations are as follows.

1. For a life time of 37 days

- a) Hazard function ( $q_i$ )

It is known that the number of individuals who died on the 37th day of observation ( $\delta_i$ ) was 1 patient and the number of individuals who lived on the 37th day of observation ( $R_i$ ) was 45 patients, so the hazard function calculation can be done as follows.

$$q_1 = \frac{\delta_1}{R_1} = \frac{1}{45} = 0,022222$$

The estimated probability of individual death (hazard function) on the 37th day is 0,022222.

- b) Estimated Chance of Individual Resilience ( $p_i$ )

It is known that the value of the hazard function is 0,022222, so the estimated chance of individual survival on the 37th day of observation is as follows.

$$p_1 = \left(1 - \frac{\delta_1}{R_1}\right) = (1 - q_1) = 1 - 0,022222 = 0,977778$$

The estimated probability of individual survival on the 37th day of observation was 0,977778.

- c) Survival Function Estimation ( $S(t)$ )

$$S(37) = \prod_{i=1}^1 p_i = 0,977778$$

The estimated survival function for each individual included on the 37th day of observation was 0,977778.

2. For a life time of 155 days

a) Hazard function ( $q_i$ )

It is known that the number of individuals who died on the 155th day of observation ( $\delta_i$ ) was 1 patient and the number of individuals who lived on the 155th day of observation ( $R_i$ ) was 33 patients, so the hazard function calculation can be done as follows.

$$q_{13} = \frac{\delta_{13}}{R_{13}} = \frac{1}{33} = 0,030303$$

The estimated probability of individual death (hazard function) on the 155th day is 0,030303.

b) Estimated Chance of Individual Resilience ( $p_i$ )

It is known that the value of the hazard function is 0,030303, so the estimated chance of individual survival on the 155th day of observation is as follows.

$$p_{13} = \left(1 - \frac{\delta_{13}}{R_{13}}\right) = (1 - q_{13}) = 1 - 0,030303 = 0,969697$$

The estimated probability of individual survival on the 155th day of observation was 0,969697.

c) Survival Function Estimation ( $S(t)$ )

$$S(155) = \prod_{i=1}^{13} p_i = 0,711111$$

The estimated survival function for each individual included on the 155th day of observation was 0,711111.

3. For a life time of 1776 days

a) Hazard function ( $q_i$ )

It is known that the number of individuals who died on the 1776th day of observation ( $\delta_i$ ) was 1 patient and the number of individuals who lived on the 1776th observation day ( $R_i$ ) was 5 patients, so the hazard function calculation can be done as follows.

$$q_{41} = \frac{\delta_{41}}{R_{41}} = \frac{1}{5} = 0,2$$

The estimated probability of individual death (hazard function) on day 1776 was 0,2.

b) Estimated Chance of Individual Resilience ( $p_i$ )

It is known that the value of the hazard function is 0.2, so the estimated probability of individual survival on the 1776th day of observation is as follows.

$$p_{41} = \left(1 - \frac{\delta_{41}}{R_{41}}\right) = (1 - q_{41}) = 1 - 0,2 = 0,8$$

The estimated probability of individual survival on the 1776th day of observation was 0,8.

c) Survival Function Estimation (S(t))

$$S(1776) = \prod_{i=1}^{41} p_i = 0,262696$$

It was found that the estimated survival function for each individual included on the 1776th day of observation was 0,262696.

Using the Kaplan-Meier method, the results of the product limit estimator or Kaplan-Meier survival data for type 1 censored head and neck cancer patients can be seen in table 4 as follows.

**Table 4.** Table of Estimation of the Survival Function of the Kaplan-Meier Method

ti	Ri	δi	Ci	qi	pi	S(t)
37	45	1	0	0,022222	0,977778	0,977778
84	44	1	0	0,022727	0,977273	0,955556
92	43	1	0	0,023256	0,976744	0,933333
94	42	1	0	0,02381	0,97619	0,911111
110	41	1	0	0,02439	0,97561	0,888889
112	40	1	0	0,025	0,975	0,866667
119	39	1	0	0,025641	0,974359	0,844444
127	38	1	0	0,026316	0,973684	0,822222
130	37	1	0	0,027027	0,972973	0,8
133	36	1	0	0,027778	0,972222	0,777778
140	35	1	0	0,028571	0,971429	0,755556
146	34	1	0	0,029412	0,970588	0,733333
155	33	1	0	0,030303	0,969697	0,711111
159	32	1	0	0,03125	0,96875	0,688889
169	31	0	1	0	1	0,688889
173	30	1	0	0,033333	0,966667	0,665926
179	29	1	0	0,034483	0,965517	0,642963
194	28	1	0	0,035714	0,964286	0,62
195	27	1	0	0,037037	0,962963	0,597037
209	26	1	0	0,038462	0,961538	0,574074
249	25	1	0	0,04	0,96	0,551111
281	24	1	0	0,041667	0,958333	0,528148
319	23	1	0	0,043478	0,956522	0,505185

339	22	1	0	0,045455	0,954545	0,482222
432	21	1	0	0,047619	0,952381	0,459259
469	20	1	0	0,05	0,95	0,436296
519	19	1	0	0,052632	0,947368	0,413333
528	18	0	1	0	1	0,413333
547	17	0	1	0	1	0,413333
613	16	0	1	0	1	0,413333
633	15	1	0	0,066667	0,933333	0,385778
725	14	1	0	0,071429	0,928571	0,358222
759	13	0	1	0	1	0,358222
817	12	1	0	0,083333	0,916667	0,32837
1092	11	0	1	0	1	0,32837
1245	10	0	1	0	1	0,32837
1331	9	0	1	0	1	0,32837
1557	8	0	1	0	1	0,32837
1642	7	0	1	0	1	0,32837
1771	6	0	1	0	1	0,32837
1776	5	1	0	0,2	0,8	0,262696
1897	4	0	1	0	1	0,262696
2023	3	0	1	0	1	0,262696
2146	2	0	1	0	1	0,262696
2297	1	0	1	0	1	0,262696

Table 4, shows that patients with head and neck cancer who are undergoing treatment with an RCT have a 37-day chance of survival of 0,977778, while the chance for patients with head and neck cancer who are undergoing treatment with an RCT is 0,2206, and so are so on. The value of the hazard function (failure rate) of head and neck cancer patients undergoing treatment with RCTs has a 37-day survival chance of 0,022222, while the patient's survival for 84 days is 0,044444, and so on.

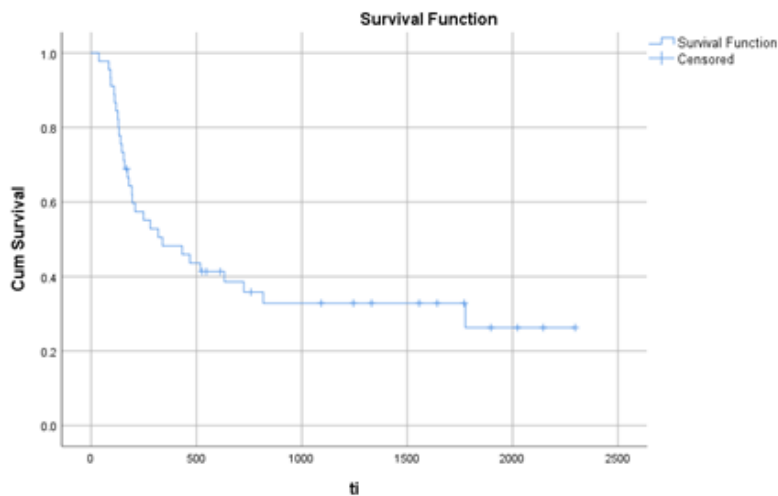


Figure 2. Survival Data Plot of Head and Neck Cancer Patients

Figure 2 shows that the survival time of head and neck cancer patients undergoing RCT treatment in the first 500 days decreased dramatically, after 500 days the plot slowly decreased, which means that the chances of a patient surviving are getting smaller as time goes by. Based on this figure, the survival chance of head and neck cancer patients undergoing RCT treatment for 130 days is 0,8.

## CONCLUSION

Based on the results and discussion, head and neck cancer patients have an average life span of 639,2 days. The results of the calculation of the product limit or Kaplan-Meier estimator and the survival plot of head and neck cancer patients tend to decrease, which means that the chances of a patient surviving are getting smaller as time goes by. By using the Kaplan-Meier method, it can be seen that the chance of head and neck cancer patients undergoing RCT treatment surviving 37 days is 0,977778, while the failure rate for head and neck cancer patients surviving 37 days is 0,022222.

## REFERENCE

- [1] R. Nia and A. Agoestanto, "Estimator Bayes untuk Rata-Rata Tahan Hidup dari Distribusi Rayleigh pada Data Disensor Tipe II," *Unnes Journal of Mathematics*, vol. 3, no. 2, 2014.
- [2] S. Mulyani and A. Maftukhin, "Analysis Survival Therapy Hemodialysis pada Pasien Gagal Ginjal Kronik di RS Kelas B Dr. Sosodoro Djatikoesoemo Bojonegoro," *Asuhan Kesehatan Jurnal Ilmiah Ilmu Kebidanan dan Keperawatan*, vol. 7, no. 1, 2016.
- [3] A. S. Yadav, S. K. Singh, and U. Singh, "Bayes estimator of the parameter and reliability function of Marshall-Olkin extended exponential distribution using hybrid Type-II censored data," *Journal of Statistics and Management Systems*, vol. 19, no. 3, pp. 325-344, 2016.
- [4] B. Audina and M. Fatekurohman, "Analisis Survival pada Data Pasien Covid 19 di Kabupaten Jember," *Berkala Sainstek*, vol. 8, no. 4, pp. 118, 2020.
- [5] S. S. Utami and Mustikasari, "Aspek Psikososial Pada Penderita Kanker Payudara: Studi Pendahuluan," *Jurnal Keperawatan Indonesia*, vol. 20, no. 2, pp. 65-74, 2017.
- [6] R. Pradika, and B. A. Priatna, "Aplikasi Metode Kaplan Meier Sebagai Penduga Ketahanan Hidup Penderita Kanker Payudara," *Jurnal EurekaMatika*, vol. 9, no.1, pp. 35-44, 2019.
- [7] M. D. Mody, J. W. Rocco, S. S. Yom, R. I. Haddad, and N. F. Saba, "Head and neck cancer," *The Lancet*, vol. 398 no. 10318, pp. 2289-2299, 2021.
- [8] M. Ashraf-Ul-Alam and A. A. Khan, "Comparison of accelerated failure time models: A Bayesian study on head and neck cancer data," *Journal of Statistics Applications and Probability*, vol. 10, no. 3, pp. 715-738, 2021.
- [9] R. Hidayat, M. Sam, R. Y. Wardi, and M. I. Iskandar, "Pemodelan Survival Pasien Covid-19 dengan Hazard Non-Proporsional," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 10, no. 1, pp. 120-130, 2022.
- [10] M. Gustina, "Identifikasi Karakteristik Hazard Rate Distribusi Generalized Exponential," Thesis, Universitas Lampung, Lampung, 2016.
- [11] M. Ramdhan, *Metode penelitian*. Cipta Media Nusantara, 2021.
- [12] S. Santoso, *Mahir statistik parametrik*. Elex Media Komputindo, 2019.