



A Closer Look at the Kaplan-Meier and Life Table Models in Survival Analysis

Anthony Joe Turkson

Department of Mathematics, Statistics and Actuarial Science, Takoradi Technical University, Sekondi-Takoradi, Ghana

Email: anthony.turkson@ttu.edu.gh

How to cite this paper: Turkson, A.J. (2021) A Closer Look at the Kaplan-Meier and Life Table Models in Survival Analysis. *Open Access Library Journal*, 8: e8104. <https://doi.org/10.4236/oalib.1108104>

Received: October 20, 2021

Accepted: November 15, 2021

Published: November 18, 2021

Copyright © 2021 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Survival analysis comprises a set of statistical methods deployed in studying the timing and occurrence of events. This paper studied survival functions with particular reference to Kaplan-Meier (K-M) estimators and Life Tables. Secondary data of In-patients who reported cases of malaria (origin state) to time until death or recovering (censored) was extracted and analyzed using Kaplan-Meier and Life Table functions in SPSS. Through this discourse, we showed how survival probabilities could be obtained and graphed. We inferred from the data provided in this paper that the mean years of life left for a new born child e_0 was 61.9 years. The expectation of life was equal to 0.016, which translates to 160 deaths per 100,000 person-years. Again, the number of new born children dying before age twenty (20) was given by $I_0 - I_{20} = 100000 - 97127 = 2873$. The probability of new born children dying before age twenty (20) years was 0.0287. The population survival curves for the two classes of users of ITN, after adjusting for gender, gave us a p-value of $0.002 < 0.05$ which was highly significant, implying that there was a statistically significant difference between the population survival curves. For the patients who subscribed to ITNs, the probability that they would survive for at least a day after admission was 0.9, while for non-users of ITNs, the probability was 0.8, again, the probability that patients who use ITNs will survive for at least 30 days after admission was 0.6, while for non-users the probability was 0.2. It should be noted that survival analysis is suitable for studies that has to do with time until the occurrence of events. It could also be used to identify factors which significantly influence an event.

Subject Areas

Statistics

Keywords

Insecticide Treated Nets, Kaplan-Meier, Life Table, Malaria, Survival Analysis

1. Introduction

Survival Analysis derived its name from biostatistics, medical science and epidemiology; in engineering it is referred to as reliability analysis; in economics it is called duration analysis; and in sociology or demography, the method is known as event history analysis. It is the field of statistics that models time-to-event, time-to-death, or time-to-failure data. Survival analysis contains an extremely powerful class of predictive analytics that can be used with any time-to-event predictions. Methods of survival analysis are concerned with the time from a known origin to either an event or a censoring point. It may deal with events such as the time from diagnosis of a disease to death, it can also be used in time-dependent phenomenon, such as time in hospital or time until a disease recurs. The best way to display survival data is by a Kaplan-Meier survival curve and could be complemented with Life Table methods. With this curve, the flow of events is seen graphically as a step function. Each time an event occurs, the survival curve is re-calculated. Additionally, Life Tables calculate survivals at fixed points in time. This paper is devoted to the study of survival functions with particular reference to Kaplan-Meier Estimators and Life Tables. Special emphasis is placed on how Kaplan-Meier curves and Life Tables are generated, analyzed and interpreted; how their values are not affected by incomplete data, and how each of the procedures could be applied in real life situations. Applying the theories behind Kaplan Meier Estimators and Life Tables reveals hidden trends in health data which invariably leads to the introduction of interventions to curb further occurrences. These procedures are useful for predictions in maintaining optimal policies in medical and biomedical studies. It also has the added advantage of being used as a research design for other research works. The purpose of this paper is threefold: To have a closer look at the dynamics of Kaplan Meier estimators lacking the discussions with examples; to give an overview of the Life Table model using the actuarial method and the Kaplan Meier Method; and finally, to analyze a real case scenario using the Kaplan-Meier Methods and the Life Table methods.

1.1. Theoretical Underpinning

Hosmer *et al.* [1] have mentioned that the most common non-parametric approach in the literature is the Kaplan-Meier (K-M) estimator which estimates the unadjusted probability of surviving beyond a certain time point. The K-M estimator works by breaking up the estimation of $S(t)$ into a series of intervals based on observed event times. These observed data contribute to the estimation of $S(t)$ until the event occurs or are censored. According to Hosmer *et al.*

(2008) [1] and Hosmer *et al.* (1999) [2] for each interval, the probability of surviving until the end of the interval is calculated, with the prior information that subjects were at risk at the beginning of the interval. In medical research for instance, it is used to measure the proportion of patients living for a certain amount of time after treatment. In employment studies, Kaplan-Meier estimators are used to measure the length of time graduates remain unemployed after graduation from college, then, in agricultural studies, researchers use this model to study how long it takes fruits to mature before they are plucked by harvesters.

Kaplan and Meier [3] published a paper on how to deal with incomplete observations. Thereafter, the Kaplan-Meier curves and estimates of survival data became very influential in dealing with various time-to-event data especially for cases of incomplete information. There are many real-life cases where time-to-events may be important end-point variables, these cases include cancer survival times, onset of malaria, time period between contracting HIV and progression to clinical AIDs, and quite recently, the duration from time a person gets infected with COVID-19, till death or recovery following use of medication. In these few examples, survival times may relate to actual survival with or without death being the event. The event of interest may vary. The Kaplan-Meier method resembles a modified form of the life table method. In the life table method, the time axis is divided into many discrete time intervals, usually years. The number at the beginning of the year, the number dying in the year, and the number censored or lost to follow-up in the year are all tabulated. The Kaplan-Meier method also divides the time axis into many discrete intervals. However, in this case the intervals are not defined by a fixed length but by the occurrence of an event. The probability of death during any of these intervals is simply the number who died divided by the number of individuals who started the interval. The probability of surviving the interval is one minus the probability of dying in the interval as indicated by Rich *et al.* and Turkson [4] [5].

In preparing for Kaplan-Meier analysis, each subject must be characterized by three variables namely:

- The time variable which should be sequential;
- The status variable at the end of the time (uncensored or censored); and
- The treatment group variable.

These characteristics have been displayed in **Table 1**. For the construction of survival time probabilities and curves, the sequential times for each subject are arranged from the shortest to the longest time, without regard to when they entered the study. By this maneuver, all subjects within the group begin the analysis at the same point, they will be studied till the occurrence of the event of interest or censoring.

1.2. Kaplan-Meier as a Non-Parametric Technique

In non-parametric techniques, the models are adaptive and can exhibit a high degree of flexibility. The model structure is not specified a priori but is determined

Table 1. Table for Kaplan-Meier analysis.

SUBJECT	TIME (weeks)	STATUS (1 = event; 0 = censored)	TREATMENT GROUP (1 or 2)
A	4	0	1
B	6	1	1
C	10	1	1
D	12	1	1
E	13	1	1
F	15	0	1
G	18	1	2
H	19	1	2
I	20	1	2
J	22	0	2
K	25	0	2
L	30	0	2

from data, this does not necessarily mean that the model lacks parameters, but, instead, the number and nature of the parameters are flexible and not fixed in advance. In simple statistical terminology, the model does not conform to a normal distribution as they rely on continuous data and not discrete data as observed by Nisbet *et al.* [6]. The above then defines the characteristics of the Kaplan-Meier estimate (K.M). The Life Table Procedure also falls under the non-parametric model. As has been described earlier, Survival analysis deals with non-negative regression and density estimates for a single random variable in the presence of censoring.

The merit of using the non-parametric model is that: it is *flexible*, with the *model becoming more complicated as the number of observations increase*. The demerits are twofold: Firstly, **it is not easy to incorporate covariates**, this can be overcome by fitting a different model on two subpopulations and comparing them. Secondly, **the survival functions are not smooth**. That is to say, they are piecewise constant. This challenge could be smoothed out by adopting the kernel smoothing techniques.

1.3. Generating the Kaplan-Meier (K-M) Product-Limit Estimator

The product limit estimator is used to estimate the survival functions, using censored data, the intervals comprise the spaces between the data as they were observed, some of the data may be censored while others may be uncensored.

Suppose that T_1, T_2, \dots, T_n are right censored by constants t_1, t_2, \dots, t_n . We observe (Z_j, δ_j) , $j = 1, 2, \dots, n$, where $Z_j = \min(T_j, t_j)$ and

$$\delta = \begin{cases} 1, & \text{if } T_j \leq t_j \text{ uncensored} \\ 0, & \text{if } T_j > t_j \text{ censored} \end{cases}.$$

Since the data occur naturally as ordered pairs, we know exactly which points were censored and which points were not censored, assuming there were no tied observations, then

$$Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$$

Since the observed data are all different, we must always ensure that when the data are ranked from smallest to largest, the censoring indicators are also re-arranged to follow the new ranking. Let δ_j denote the value of the indicator associated with $Z_{(j)}$. To construct the survival estimator, Namboodiri and Suchundran [7] assets that we divide $(0, Z_n)$ into disjoint intervals $I_j = 1, 2, \dots, n$ such that $Z_{(0)} = 0$.

Definition: The risk set at time t denoted by $R(t)$, is the set of indices of subjects which are still alive and under observation at time t^- , that is, at time prior to t , then

N_j (The number at risk) is the number of elements in the set $R(Z_j)$.

D_j (The number of deaths) is the observed failure at Z_j and is equal to 0 or 1.

p_j (The conditional probability) = P (Surviving through I_j | Alive at the start of I_j).

$$\text{Estimate of } p_j = \frac{\text{number dying in } I_j}{\text{number with the potential to die in } I_j} = \begin{cases} 1 - \frac{1}{N}, & \text{if } \delta_{(j)} = 1 \\ 0 & \text{if } \delta_{(j)} = 0 \end{cases}$$

The product of all such estimates \hat{p}_j gives an estimate of the survival functions S , we note that as j increases, the risk sets diminish one at a time: $N_j = n - (j - 1) = n - j + 1$, putting these ideas together, for a fixed value of u

$$\begin{aligned} \hat{S}(t) &= \prod_{j: Z_{(j)} \leq t} \hat{p}_j \\ &= \prod_{j: Z_{(j)} \leq t, \delta_{(j)} = 1} \left(1 - \frac{1}{N}\right) \\ &= \prod_{j: Z_{(j)} \leq t} \left(1 - \frac{1}{N}\right)^{\delta_{(j)}} \\ &= \prod_{j: Z_{(j)} \leq t} \left(1 - \frac{1}{n - j + 1}\right)^{\delta_{(j)}} \\ &= \prod_{j: Z_{(j)} \leq t} \left(\frac{n - j}{n - j + 1}\right)^{\delta_{(j)}} \end{aligned} \quad (1)$$

This estimator (1) is the most common method of estimating the survival function:

$$S(t) = P(T > t) = \hat{S}(t_j) = \hat{S}(t_{j-1}) \times \widehat{Pr}(T \geq t_j | T \geq t_j) \quad (2)$$

Equation (2) gives the probability of surviving past the previous failure time t_{j-1} multiplied by the conditional probability of surviving past time t_j .

Suppose that we have n observations and that there are m unique event times arranged in ascending order $t_{(1)}, t_{(2)}, \dots, t_{(m)}$.

Between $t = 0$ and $t = t_{(1)}$ that is, the time of the first event, the estimate of the survival function $\hat{S}(t) = 1$.

n_j = # of subjects at risk of the event just before time t_j .

d_j = # of events observed (death) at time t_j .

$p_j = P(\text{surviving through } t_{(j)} \mid \text{alive at the beginning of } t_{(j)}) = P(T > t_j \mid T \geq t)$.

$q_j = 1 - p_j$ (That is, the complement of p_j).

Recalling the multiplication rule for joint events A_1 and A_2

$$P(A_1 \cap A_2) = P(A_2 \mid A_1)P(A_1)$$

We can write the survival probability as $S(t) = P(T > t) = \prod_{t_j \leq t} p_j$. The estimate of p_j and q_j are $\hat{q}_j = \frac{d_j}{n_j}$ and $\hat{p}_j = 1 - q_j = 1 - \frac{d_j}{n_j} = \frac{n_j - d_j}{n_j}$, thus the K-M estimator is given by

$$\begin{aligned} \hat{S}(t) &= \prod_{t_j \leq t} \hat{p}_j = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) \\ \hat{S}(t) &= \prod_{t_{(j)} \leq t} \frac{n_j - d_j}{n_j} \end{aligned} \quad (3)$$

Using the K-M definition the estimator of survivorship in Equation (3) can be written alternatively as

$$\begin{aligned} \hat{S}(t_j) &= \hat{S}(t_{(j-1)}) \times \hat{P}(T > t_{(j)} \mid T \geq t_{(j)}) \\ \text{where } \hat{S}(t_{(j-1)}) &= \prod_{i=1}^{j-1} \hat{P}(T > t_{(i)} \mid T \geq t_{(i)}) \end{aligned} \quad (4)$$

When there is no censoring, n_j is just the number of survivors prior to time t_j . With censoring, n_j is the number of survivors less the number of losses (censored cases). It is only those surviving cases that are still being observed and are “at risk” of death. The principle governing the use of the K-M estimate is illustrated as follows:

1.4. Kaplan-Meier Computations

We use part of the Malaria data collected to provide the computational mechanism of the Kaplan-Meier Estimator. The data is produced below.

Group	Length of stay at the hospital (in days)
Users of ITN	9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48
Non-Users of ITN	5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45

Note: + indicates a censored value.

In the computation of K-M estimates, we include incomplete information about patients. We first present the data layout needed to do the computation and proceed with the use of Equation (4) (**Table 2**).

Table 2. The basic data layout needed to understand survival data.

<i>Time in days</i> (t_i)	<i>Number Dead</i> (d_i)	<i>Number Censored</i>	<i>Risk set</i> ($R(t_i)$)
$t_{(0)} = 0$	0	0	10 persons survive ≥ 0 days
$t_{(1)} = 9$	1	0	10 persons survive ≥ 9 days
$t_{(2)} = 13$	1	1	9 persons survive ≥ 13 days
$t_{(3)} = 18$	1	0	7 persons survive ≥ 18 days
$t_{(4)} = 23$	1	1	6 persons survive ≥ 23 days
$t_{(5)} = 31$	1	0	4 persons survive ≥ 31 days
$t_{(6)} = 34$	1	1	3 persons survive ≥ 34 days
$t_{(7)} = 48$	1	0	1 person survive ≥ 48 days
Totals	7	3	

1.5. Computing the K-M Survival Function Using Data from Table 2

In computing the survival function of users of ITN the following K-M formula was used

$$\hat{S}(t_j) = \hat{S}(t_{(j-1)}) \times \hat{P}(T > t_{(j)} | T \geq t_{(j)}) \quad \text{where} \quad \hat{S}(t_{(j-1)}) = \prod_{i=1}^{j-1} \hat{P}(T > t_{(i)} | T \geq t_{(i)})$$

Observation: It is noted from **Table 3** that the censored figures were not included in the time column, that notwithstanding they were captured under the censored status. For instance, 28+ was captured between time 23 and 34 and recorded as censored in line with time $t_{(4)}$ and beyond. The censored figure 45+, was also captured between the time periods 34 and 48 days and recorded in line with $t_{(6)}$.

1.6. The Life Table Method

Selvin [8] has defined a life table as a highly organized description of age-specific mortality rates. Its application ranges well beyond mortality analysis, it has also attracted studies in the social sciences. There is a close connection between life tables and survival analysis in that both theories are expressed in terms of follow-up of a cohort; both rely on concepts like survival function and hazard functions. The difference is that the method of survival analysis is applied to data collected by following a relatively small cohort over a short period of time whereas life table methods are generally used to analyze cross-sectional data from large populations. A significant feature of the life table method according to Govindurajulu and Qadri [9] is that it does not require a standard population. Alison [10] has observed that life table is among the most common methods used to analyze survival data. Demographic and insurance studies also fall under its jurisdiction. The primary function of a life table is to summarize survival data grouped into intervals to provide estimates of the survivor function, the density

Table 3. Results of the K-M computations.

$\hat{S}(0)$	1.00
$\hat{S}(9) = \hat{S}(0) \times \frac{10-1}{10}$	0.90
$\hat{S}(13) = \hat{S}(9) \times \frac{9-1}{9}$	0.80
$\hat{S}(13+) = \hat{S}(13) \times \frac{8-0}{8} = \hat{S}(13)$	0.80
$\hat{S}(18) = \hat{S}(13) \times \frac{7-1}{7}$	0.69
$\hat{S}(23) = \hat{S}(18) \times \frac{6-1}{6}$	0.57
$\hat{S}(28+) = \hat{S}(23) \times \frac{5-0}{5} = \hat{S}(23)$	0.57
$\hat{S}(31) = \hat{S}(23) \times \frac{4-1}{4}$	0.43
$\hat{S}(34) = \hat{S}(31) \times \frac{3-1}{3}$	0.29
$\hat{S}(45+) = \hat{S}(34) \times \frac{2-0}{2} = \hat{S}(34)$	0.29
$\hat{S}(48) = \hat{S}(34) \times \frac{1-1}{1}$	0.00

function, and the hazard rate. Kalbfleisch and Prentice [11] and Allison [12] have indicated that the life table is designed for situations where only the interval in which failure or censoring occurred is known but the actual failure or censoring time is unknown. Teachman [13] has alluded that Life Tables are also useful for preliminary evaluation of data and evaluating the fit of regression models. It also allows assessment of exogenous variables in more complex analyses. Additionally, it can be used to assess the mortality level of a population and its age structure, project the population into the future, assess the survival rate and the number of cases at risk within a cohort.

There are basically three types of life tables; ordinary life table, multiple decrement life tables and cause deleted life tables. In ordinary life table membership in a well-defined study group can be terminated by a single attrition factor while in a multiple-decrement life table, there are multiple reasons for attrition. The life table methods can be used to study a variety of cohorts. A birth cohort where the attrition factors are causes of death. A marriage cohort where the attrition factors are causes of marital disruption and death due to a disease cohort where the attrition factor is death due to diseases. In this section we limit our discussions to ordinary life tables.

1.7. Ordinary Life Tables

According to Shryock and Siegel [14], the ordinary life table is a statistical technique that gives the summary of mortality experiences, survivorship and the life expectancy of a population of varying age. It gives a measure of period age-specific death rates, it also describes the hypothetical survival experiences of a synthetic or fictitious cohort subject to current death rates over an imagined lifetime. Another school of thought, views the table's numbers as a description of the stationary population model build up by a succession of birth cohorts of the same size all of which experience the same age-specific death rates. The Table also gives information about longevity and expectation of life. It does provide a technique of analyzing data for all the causes of the event of interest. This discussion is limited to the study of current or period life table, which is based on the mortality experiences of a hypothetical cohort of newborn babies, usually 100,000 newborns who are subject to the age-specific mortality rates on which the table is based. It traces the cohort of newborns throughout their lifetime under the assumption that they are subject to the age-specific mortality rates of a country. There are two types of current life tables: Unabridged for single years of life and abridged for five (5) year cohorts of life, again, current discussion is based on the abridged life table.

From **Table 4** and **Table 5**, we give brief explanations to some items in the nine columns. To calculate the age specific mortality rate for the first year of life, we proceeded as follows:

$$P_x = 90035, \quad D_x = 738$$

$${}_5q_0 = \frac{738}{90035 + 0.5 \times 738} = 0.00816$$

Table 4. Column notation for a life table.

Column #	Notes Explaining the Notation Used	Notation
1	Period of life between two age brackets(years)	$(x, x + n)$
2	Population figures which will be collected from source documents	P_x
3	Number of deaths within the age bracket	D_x
4	Proportion of persons surviving at the beginning of the age bracket who die during the age bracket	${}_nq_x$
5	Out of the number of newborns who started in the life table (l_0 is called the radix, and is set at 100,000) the number surviving at the beginning of the age bracket	l_x
6	The number of persons in the cohort who die in the age bracket $(x, x + n)$	${}_nd_x$
7	Number of years of life (person-years of life in the age interval) lived by the cohort within the indicated bracket $(x, x + n)$	${}_nL_x$
8	Total person-years of life contributed by the cohort after attaining age x	T_x
9	Average number of years of life remaining for a person alive at the beginning of age bracket $(x, x + n)$	e_x

Table 5. Example of an abridged life table for a population under study.

$(x, x+n)$	P_x	D_x	${}_nq_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	e_x
(1)	(2)	(3)	(4) = (6)/(5)	(5)	(6)	(7)	(8)	(9)
0 - 5	90035	738	0.00816	100000	816	99266	6190163	61.90
5 - 10	90056	703	0.00777	99184	771	98799	6090897	61.41
10 - 15	89067	601	0.00672	98413	662	98082	5992099	60.89
15 - 20	88906	569	0.00638	97751	624	97439	5894017	60.30
20 - 25	75342	300	0.00397	97127	386	96934	5796578	59.68
30 - 35	77562	246	0.00316	96741	306	96588	5699644	58.92
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
80 - 85	42567	234	0.00548	78832	432	78616	234159	2.97
85 - 90	45362	300	0.00659	78400	517	78142	155543	1.98
90+	45367	564	0.01236	77883	962	77402	77402	0.99

The other columns can be calculated using the formulae given in the text above. l_0 which is the radix of the Life Table is arbitrary taken as 100,000. For example, in column (6) ${}_5d_0$ is equal to $0.00816 \times 100000 = 816$. In column (5), $l_1 = 100000 - 816 = 99184$ again ${}_5L_0 = 100000 - 0.9 \times 816 = 99266$. This is because in early childhood, mortality declines rapidly with age. $10L_5 = 99187 - 0.5 \times 771 = 98799$, it is assumed that at most ages' deaths are evenly distributed. $T_x = \sum {}_nL_x$. For instance, $T_{80} = 77402 + 78142 + 78616 = 234160$.

As stated earlier on, the value of e_x gives a summary of the life table survival in terms of mean life left to be lived by each age group x . From **Table 5**, $e_x = \frac{T_x}{l_x}$, it

could be inferred that the mean years of life left for a new born child e_0 is 61.9 years. The crude mortality rate is just the reciprocal of the expectation of life e_x . For a new born child the expectation of life is equal to $1/61.9$ which is equal to 0.016, this translates to 160 deaths per 100,000 person years. The crude mortality rate measures the risk of death.

$\sum_0^n {}_nd_x = l_0$, that is to say, the sum of the number of deaths in all age intervals gives us the radix. Number of persons dying before age x is given by $l_0 - l_x$, in the Life Table (**Table 5**), the number of new born children dying before age twenty (20) is given by $l_0 - l_{20} = 100000 - 97127 = 2873$.

The probability of new born children dying before age x , where x is twenty (20) years is given by $\frac{l_0 - l_x}{l_0} = \frac{100000 - 97127}{100000} = \frac{2873}{100000} = 0.0287$.

1.8. Calculation of Survival Rates

We use the abridged life table (**Table 5**) to calculate survival rates. For population projections, 5-year survival rates are computed and for estimates of net migration, 10-year survival rates are calculated. Calculation of survival rates relies on two columns from the table, *i.e.* T_x and L_x . Measure evaluation [15].

$$\bullet \quad 5\text{-year survival rate} = S_{(x+5)} = \frac{{}_5L_{(x+5)}}{{}_5L_x} \quad (5)$$

To calculate the probability that a child in age bracket (15 - 20) will survive to the next age bracket (20 - 25), is obtained as follows:

$$S_{(20-25)} = \frac{L_{(20-25)}}{L_{(15-20)}} = \frac{96934}{97439} = 0.9948$$

- Survivorship of the youngest age cohort (0 - 5) and (5 - 10):

$$S_{(0-5,5-10)} = \frac{L_{(0-5)} + L_{(5-10)}}{l_0 \times (\text{age interval})} = \frac{99266 + 98799}{100000 \times 5} = 0.3961$$

- Survivorship of the oldest age cohort $S_{95^+} = \frac{T_{90-95}}{T_{85-90}} = \frac{77402}{155543} = 0.4976$

1.9. Computation of Ordinary Life Table

LaMorte [16] and Kleinbaum [17] have presented a Life Table used widely in biostatistics called a **cohort Life Table**. This table summarizes the experiences of participants over a pre-defined period in a cohort study until the event of interest is observed or the study is terminated. We organize the table into five (5) equally spaced intervals (that is, 0 - 4, 5 - 9, 10 - 14, 15 - 19 and so on) (**Table 6**).

Observation: From time 17 to 21 years the survival probabilities were all the same, the simple reason is that all the subjects within that period were censored. The same applies to subjects within the time 24 to 30 years (**Table 7**).

2. Methodology

Secondary data was collected to assess the survival experiences of residents in

Table 6. Life table construction using the actuarial method.

t	N_t	$N_\rho = N_t - \frac{C_t}{2}$	D_t	C_t	$q_t = \frac{D_t}{N_\rho}$	$p_t = 1 - q_t$	$S(t) = p_t \times p_{t-1}$
0 - 4	32	$32 - (2/2) = 31.0$	3	2	$3/31 = 0.097$	$1 - 0.097 = 0.903$	$0.903 \times 1.000 = \mathbf{0.903}$
5 - 9	27	$27 - (2/2) = 26.0$	1	2	$1/26 = 0.038$	$1 - 0.038 = 0.962$	$0.962 \times 0.903 = \mathbf{0.869}$
10 - 14	24	23.0	1	2	$1/23 = 0.043$	$1 - 0.043 = 0.957$	$0.957 \times 0.869 = \mathbf{0.831}$
15 - 19	21	20.5	1	1	$1/20.5 = 0.049$	$1 - 0.049 = 0.951$	$0.951 \times 0.831 = \mathbf{0.790}$
20 - 24	19	18.5	2	1	$2/18.5 = 0.108$	$1 - 0.108 = 0.892$	$0.892 \times 0.790 = \mathbf{0.705}$
24 - 29	16	15	1	2	$1/15 = 0.067$	$1 - 0.067 = 0.933$	$0.933 \times 0.705 = \mathbf{0.658}$
30 - 34	13	12	2	2	$2/12 = 0.167$	$1 - 0.167 = 0.833$	$0.833 \times 0.658 = \mathbf{0.548}$
35 - 39	9	8.5	2	1	$2/8.5 = 0.235$	$1 - 0.235 = 0.765$	$0.765 \times 0.548 = \mathbf{0.419}$
40 - 45	6	5.5	1	1	$1/5.5 = 0.182$	$1 - 0.182 = 0.818$	$0.818 \times 0.419 = \mathbf{0.343}$

Footnote: t = Time interval in years; C_t = Number censored; N_t = Number at risk during interval; q_t = Proportion dying during Interval; N_ρ = Average number at risk during interval; p_t = Proportion surviving among those at risk; D_t = Number of deaths during interval; $S(t)$ = Survival function.

Table 7. Life table construction using the Kaplan-Meier model.

Time in Years t_i	Number at Risk N_t	Number of Deaths D_t	Number Censored C_t	Survival Function (Probability) $S_{t+1} = S_t \times [(N_{t+1} - D_{t+1})/N_{t+1}]$
0	32			=1.00
1	32	1		$1.00 \times (32 - 1)/32 = 0.97$
2	31	1	1	$0.97 \times (31 - 1)/31 = 0.94$
3	29	1		$0.94 \times (29 - 1)/29 = 0.91$
5	28	1		$0.91 \times (28 - 1)/28 = 0.88$
6	27		1	$0.88 \times (27 - 0)/27 = 0.88$
9	26	1	1	$0.88 \times (26 - 1)/26 = 0.85$
10	24		1	$0.85 \times (24 - 0)/24 = 0.85$
11	23	1	1	$0.85 \times (23 - 1)/23 = 0.84$
12	21		1	$0.84 \times (21 - 0)/21 = 0.84$
13	20		1	$0.84 \times (20 - 0)/20 = 0.84$
14	19	1		$0.84 \times (19 - 1)/19 = 0.80$
17	18		1	$0.80 \times (18 - 0)/18 = 0.80$
18	17		1	$0.80 \times (17 - 0)/17 = 0.80$
19	16		1	$0.80 \times (16 - 0)/16 = 0.80$
21	15		1	$0.80 \times (15 - 0)/15 = 0.80$
23	14	1		$0.80 \times (14 - 1)/14 = 0.74$
24	13		1	$0.74 \times (13 - 0)/13 = 0.74$
25	12		1	$0.74 \times (12 - 0)/12 = 0.74$
26	11		1	$0.74 \times (11 - 0)/11 = 0.74$
27	10		1	$0.74 \times (10 - 0)/10 = 0.74$
28	9		1	$0.74 \times (09 - 0)/09 = 0.74$
29	8		1	$0.74 \times (08 - 0)/08 = 0.74$
30	7		1	$0.74 \times (07 - 0)/07 = 0.74$
31	6	1		$0.74 \times (06 - 1)/06 = 0.62$
32	5	1		$0.62 \times (05 - 1)/05 = 0.50$
33	4	1		$0.50 \times (04 - 1)/04 = 0.38$
34	3	1		$0.38 \times (03 - 1)/03 = 0.25$

the Sekondi-Takoradi district who use various means besides insecticide treated nets (ITNs) to prevent exposure to the malaria parasite. In doing so, the researcher assessed hospital in-patients who reported cases of malaria (origin state) to time until death or censored (recovered) as a result of causal factors (exposure levels to the malaria parasites). The event of interest was time until death. Predictor variables were age, gender, and preventive measures (level of exposure).

The extracted data for each person was coded as follows: Gender: Male = 1; Female = 2.

Type of preventive measure used: Insecticide treated net = 0; mosquito nets plus others = 1.

Censoring status: if patient died, it is coded as 1; If patient was discharged, died from a different sickness or was alive at the end of study, the code is equal to 0.

Difference between date of discharge/death and date of admission = survival time in days. Age was seen as a continuous variable. The entire data was analyzed with the help of the Kaplan-Meier and The Life Table functions in the SPSS application software.

The following command was followed: Select “Analyze”, “Survival”, “Kaplan-Meier”. “SURVT” was selected from the variable list and entered into the “Time box”. “Status” variable was selected and entered into the ‘Status box’. When a question mark in parentheses appeared after the status variable, “the define event button” was clicked and a value of 1 was inserted in the box since the variable status was coded 1 for events and 0 for censorships. The process was continued then okayed. To view the output, the “Paste” option was selected rather than OK. The following syntax was obtained:

$$\text{KM survt/STATUS} = \text{status/PRINT TABLE MEAN} \quad (1)$$

To obtain Life Table estimates these steps were followed: Analyze > Survival > Life Tables. The time-to-event and status variables were defined similarly as described above for K-M estimates. The time intervals were defined and used for the life table analysis. 0 through 100 by 5 was imputed, this produced 20-time intervals of equal length. The factor variable “preventive measure” was also keyed into the factor box and the range (being the number of categories) was defined, in this example the categories were two, thus (min = 1, max = 2). The prompt actions were followed till the end of process.

3. Results and Discussion

3.1. Data on Malaria Cases for Users and Non-Users of ITN

We see from (Table 8) that out of the 1793 patients sampled 405 representing 22.6% were using insecticide treated nets while the majority (77.4%) was using other types of nets like window netting and ordinary treated nets. It was also established that of those who were using ITN 16% died within the six months study period while 84% survived, again out of the non-users of ITN 23% died

Table 8. Summary of data on malaria cases for users and non-users of ITN.

Event	ITN Users	Non-ITN Users	Total
Obtained event	66	325	391
Censored	339	1063	1402
At Risk	405	1388	1793

while 77% survived within the same study period. The ratio of death of male to female was 1.2:1. This ratio indicates that death due to malaria for the period of observation was not gender related.

3.2. Survival Functions for Gender and Used Method of Protection

The study demonstrated the use of real-life data in the computation of both the K-M estimators and the life table method. From the graph of **Figure 1**, it could be inferred that the survival experiences of males and females were approximately the same, this implies that gender did not contribute significantly to death due to malaria. The plot of **Figure 2** gave a graphical picture of the survival curves of the two groups of users of ITN. It was noticed from the graph that the survival experiences at the first few days of the study appeared to be the same but thereafter the differences showed up clearly, it was also revealed that the curve for the users of ITN consistently laid above that of the non-users, this characteristic shows that users of ITN had a better survival prognosis than non-users. The difference further meant that ITN was effective at all points during the period. From the graph one could estimate the median survival times for the two classes of users. This was done by locating 0.5 on the y-axis and proceeding horizontally till it met the curves, once the horizontal line meets the curve, a vertical line is drawn from the point of intersection of the curve and the horizontal line to meet the x-axis. From the graph the median survival time of the non-users of ITN was approximately 10 days while that of the users of ITN was close to forty (40) days. The median value further confirms the claim that users of ITN had better survival prognosis than non-users.

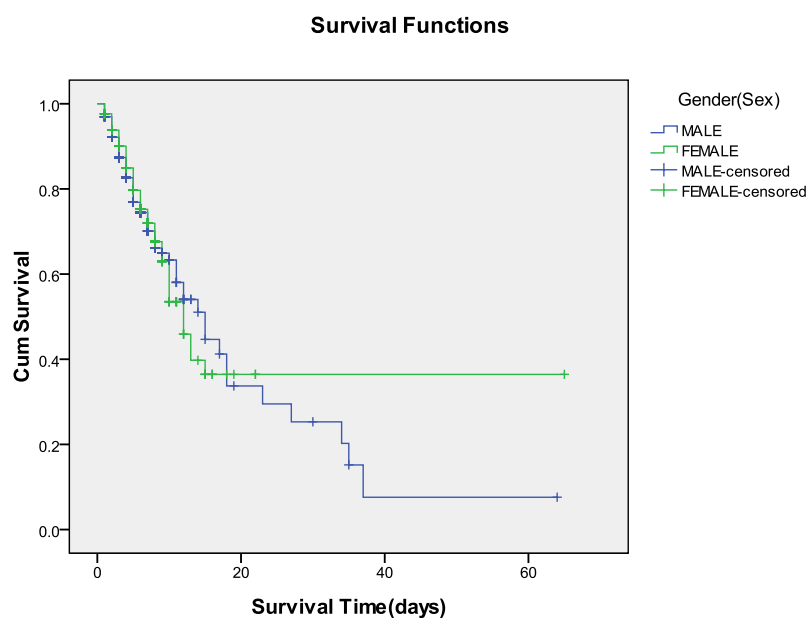


Figure 1. Survival curves of male and females and risk of malaria death. The curves for the male and the female cross each other at various points. Results show that a lot more of the females were censored than their male counterparts.

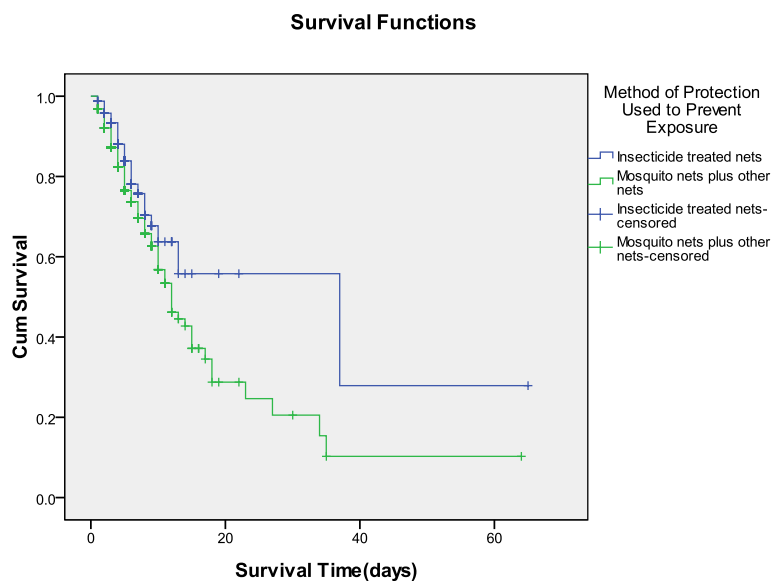


Figure 2. Survival curves of users and non-users of ITN. The curve for users of ITN lies above that of the non-users. For those who did not use ITN there were many steps within the curve with each step representing death.

The comparative analysis made on the Life Table results (**Table 9**) suggests that patients who subscribed to ITNs as means of preventing exposure to the mosquito parasites had better and longer survival experiences than patients who subscribed to other means of protection. From the life table results presented as **Table 9**, we compared the survival experiences of residents who subscribed to ITNs as means of reducing exposure to the mosquito parasites with those who subscribed to mosquito nets and other nets as means of protection. The results of the overall comparisons provided by the life table (Wilcoxin/Gehan statistics, $p = 0.002$) was the same as the one provided by the log rank test in the Kaplan-Meier method, they all showed that the method of protection adopted by the residents to prevent exposure to the mosquito parasite was highly significant. For the patients who subscribed to ITNs, the probability that they would survive for at least a day after admission at the hospital was about 0.9, for the non-users of ITNs, probability that they would survive for at least a day after admission at the hospital was 0.8, again, the probability that patients who use ITNs will survive for at least 30 days after admission was 0.6, while for patients who used other means of protection their probability was 0.2, finally, the probability that patients who used ITNs would survive for at least 60 days after admission was 0.3, but for those who subscribed to other means of protection, the probability of surviving for at least 60 days was 0.1.

3.3. Hypothesis Testing

A test of hypothesis was conducted to ascertain whether the survival curves for the two categories of users were the same or not.

Hypothesis: H_0 : The survival curves of the exposed and the less exposed

Table 9. Life table results showing the survival functions for the method of protection adopted by the dwellers against mosquito bites.

First-order Controls	Interval Start Time	Cumulative Proportion Surviving at End of Interval	
	0	0.86	
	5	0.69	
	10	0.57	
	15	0.57	
	20	0.57	
	25	0.57	
<i>Insecticide Treated Nets</i>	30	0.57	
	35	0.29	
	40	0.29	
	45	0.29	
	50	0.29	
	55	0.29	
	60	0.29	
	65	0.29	
	Measures used to Prevent Exposure to Mosquito Parasites	0	0.80
		5	0.62
10		0.43	
15		0.29	
20		0.25	
25		0.21	
<i>Mosquito Nets Plus Other Nets</i>		30	0.16
		35	0.10
		40	0.10
		45	0.10
		50	0.10
55	0.10		
60	0.10		

groups are equivalent.

Test statistic: Log-Rank test (Mantel-Cox)

Decision criteria; Reject the H_0 if the p-value is less than $\alpha = 0.05$.

The computer output (**Table 10**) provides useful information needed to test

Table 10. Summary of test results for testing the equality of Survival curves for exposed and less exposed groups. The log-rank test was used to test whether Kaplan-Meier curves for the two exposure groups in the entire population were statistically equivalent.

Test	Test	dof	Significant
Log Rank (Mantel-Cox)	9.323	1	0.002
Breslow (Generalized Wilcoxin)	10.310	1	0.001
Tarone-Ware	9.720	1	0.002

whether there was any difference in the population survival curves for the two classes of users of ITN, after adjusting for gender (since gender did not contribute significantly to the risk of malaria death). The null hypothesis for this test was that there was no difference in the survival curves of the users of ITN and non-users. The p-value of the log-rank test ($0.002 < 0.05$) was highly significant, implying that there was a statistically significant difference between the population survival curves. In my candid opinion, the purpose for which the researcher set out to carry out this study has been achieved

4. Conclusion

Survival analysis is an important field in the theory and practice of biostatistics. As alluded by Kosińska and Szwed [18] survival analysis is suitable for studies of the distribution in time of developmental events. It can be used to indicate the factors which significantly influence the course of development by modifying the duration of developmental stages. The techniques developed in survival analysis have penetrated many disciplines; various methods are available in the literature for analyzing survival data. Due to time and space, the study touched on the Kaplan-Meier Estimator and Life Tables. Lyu [19] used the Kaplan Meier approach to study the behavior of lung cancer patients. Through this discourse, it has been shown how survival probabilities could be obtained and graphed, how one could find the number at risk, conditional probabilities, number of deaths and probability of a person surviving to the next period given that he has survived the current period. It has also been demonstrated how the life table procedure could be generated and used to find the age-specific mortality rates, survival rates, longevity or expectation of life, mean years left, and crude mortality rates. Additionally, it has been shown how researchers could find the number of newborns dying before age twenty and for that matter, the probability of newborns dying before age twenty. Above all, the study has shown how one could include incomplete information into the computation of both the K-M estimators and the life table methods.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Hosmer, D.W., Lemeshow, S. and May, S. (2008) Descriptive Methods for Survival Data. In: *Applied Survival Analysis*, John Wiley, Hoboken.
<https://doi.org/10.1002/9780470258019>
- [2] Hosmer, D.W.Jr. and Lemeshow, S. (1999) *Applied Survival Analysis, Regression Modeling of Time to Event Data*. John Wiley and Sons, Inc., New York.
- [3] Kaplan, E.L. and Meier, P. (1959) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**, 457-481.
<https://doi.org/10.1080/01621459.1958.10501452>
- [4] Rich, J.I., Neely, J.G., Courtney, P., Voelker, C.J., Nussenbaum, B. and Wang, E.W. (2010) A Practical Guide to Understanding Kaplan Meir Curves. *Otolaryngology—Head and Neck Surgery*, **143**, 331-336.
<https://doi.org/10.1016/j.otohns.2010.05.007>
- [5] Turkson, A.J. (2010) Understanding the Basic Concepts of Survival Analysis. *Journal of Biostatistics*, **4**, 63-80.
- [6] Nisbet, R., Gray, M. and Yale, K. (2018) *Handbook of Statistical Analysis and Data Mining Application*. Elsevier Inc., Academic Press, Cambridge.
- [7] Namboodiri, K. and Suchindran, C.M. (1987) *Life Tables and Their Applications*. Academic Press, Orlando.
- [8] Selvin, S. (2008) *Survival Analysis for Epidemiology and Medical Research: A Practical Guide*. Cambridge University Press, New York.
- [9] Govindarajulu, U.S. and Qadri, M. (2019) Survival and Mediation Analysis with Correlated Frailty. *Current Research in Biostatistics*, **9**, 21-30.
<https://doi.org/10.3844/amjbsp.2019.21.30>
- [10] Allison, P.D. (1995) *Survival Analysis Using the SAS System: A Practical Guide*. SAS Institute, Inc., Cary.
- [11] Kalbfleisch, J.D. and Prentice, R.L. (1980) *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, New York.
- [12] Allison, P.D. (1984) *Event History Analysis*. Sage Publications, Beverly Hills.
<https://doi.org/10.4135/9781412984195>
- [13] Teachman, J.D. (1983) Analyzing Social Processes: Life Table and Proportional Hazards Model. *Social Science Research*, **12**, 263-301.
[https://doi.org/10.1016/0049-089X\(83\)90015-7](https://doi.org/10.1016/0049-089X(83)90015-7)
- [14] Shryock, H. and Siegel, S. (1973) *The Methods and Materials of Demography*. US Government Printing Office, Washington DC.
- [15] Measure Evaluation (n.d.) Overview of Life Tables and Survival Rates.
<https://www.measureevaluation.org/resources/training/online-courses-and-resources/non-certificate-courses-and-mini-tutorials/population-analysis-for-planners/lesson-7.html#:~:text=Survival%20rates%20are%20derived%20from%20life%20table%20or%20census%20data,be%20alive%20in%20the%20future.&text=Life%20tables%20are%20used%20to,several%20types%20of%20life%20tables>
- [16] LaMorte, W.W. (2016) *Survival Analysis*.
https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Survival/index.html
- [17] Kleinbaum, D.G. and Klein, M. (2005) *Survival Analysis a Self-Learning Test*. Springer Science Business Media, New York. <https://doi.org/10.1007/0-387-29150-4>
- [18] Kosińska, M. and Szwed, A. (2014) Application of Survival Analysis in Studies of Human Ontogeny. *Applied Mathematics*, **5**, 1697-1704.

<https://doi.org/10.4236/am.2014.511162>

- [19] Lyu, R. (2020) Survival Analysis of Lung Cancer Patients from TCGA Cohort. *Advances in Lung Cancer*, **9**, 1-15. <https://doi.org/10.4236/alc.2020.91001>.