A Webinar Paper on Survival Analysis using Modified Kaplan Meier Model

by

Kazeem Eyitayo Lasisi (PhD)
Professor of Biostatistics
Mathematical Sciences Department
Abubakar Tafawa Balewa University, Bauchi

INTRODUCTION

- Survival analysis, or more generally, time-to-event analysis, refers to a set of methods for analyzing the length of time until the occurrence of a well-defined end point of an interest. A unique feature of survival data is that typically not all patients experience the event (e.g, death) by the end of the observation period, so the actual survival times for some patients are unknown. This phenomenon, referred to as censoring, must be accounted for in the analysis to allow for valid inferences.

- Moreover, survival times are usually skewed, limiting the usefulness of analysis methods that assume a normal data distribution. Analyzing survival data is unique, in that the research interest is typically a combination of whether the event has occurred (binary outcome) and when it has occurred (continuous outcome).

- Appropriate analysis of survival data requires specific statistical methods that can deal with censored data. As the assessed outcome is frequently mortality, these techniques are subsumed under the term survival analysis.
- Generally, these techniques can be used for the analysis of the time until any event of interest occurs (e.g, recurrence of a disease; initial breakthrough, postoperative pain; or failure of an implanted medical device), and such data can thus also be called time-to-event or failure time data (Schober, 2018). In medicine, time duration from the patient had been diagnosed up to the time of his death.

- The objective of a survival data analysis may be just to describe a single sample of data to describe the lifetimes of a single population or to compare the lifetimes of two or more groups of subjects; for example, the two groups may have received different medical treatments and the lengths of survival time measure how effective the treatments are.
- The two main characteristics of data on times until failure are: (i) The times are non-negative and have skewed distributions with long tails. (ii) A distinctive feature of survival data is that some observations may be censored: often the event of interest (e.g. death of patient, failure of component, recovery of patient) has not occurred by the time of recording so that all is known is that the lifetime for that subject is at least some value (and may well be greater than this value).

- Such censoring cannot just be ignored since they carry important information about the effectiveness of the treatment. This introduces a complication in the statistical description and analysis of the data. (Fieller, 2011).
- There are five different types of censoring, namely: right, left, interval, double and middle censoring (Jammalamadaka and Mangalam, 2003; Collett, 2003; Turnbull, 1974).

Kaplan Meier Model

- Kaplan Meier is derived from the names of two statisticians; Edward L. Kaplan and Paul Meier, in 1958 when they made a collaborative effort and published a paper on how to deal with time to event data. Therefore, they introduced the Kaplan-Meier estimator which serves as a tool for measuring the frequency or the number of patients surviving medical treatment. Later on, the Kaplan-Meier curves and estimates of survival data have become a better way of analyzing data in cohort study. Kaplan- Meier (KM) is non-parametric estimates of survival function that is commonly used to describe survivorship of a study population and to compare two study populations.

- KM estimate is one of the best statistical methods used to measure the survival probability of patients living for a certain period of time after treatment. It is an intuitive graphical presentation approach. In clinical trials or community trials, the intervention effect is assessed by measuring the number of participants saved or survived after that intervention over a period of time. KM estimate is the simplest procedure of determining the survival over time in spite of all the difficulties associated with subjects or situations. Curves are used in Kaplan Meier estimate to determine the events, censoring and the survival probability.

- KM estimate is one of the best statistical methods used to measure the survival probability of patients living for a certain period of time after treatment. It is an intuitive graphical presentation approach. In clinical trials or community trials, the intervention effect is assessed by measuring the number of participants saved or survived after that intervention over a period of time. KM estimate is the simplest procedure of determining the survival over time in spite of all the difficulties associated with subjects or situations. Curves are used in Kaplan Meier estimate to determine the events, censoring and the survival probability.

- Kaplan-Meier survival curve is used in epidemiology to analyze time to event data and to compare two groups of subjects. The survival curve is used to determine a fraction of patients surviving a specified event, like death during a given period of time. This can be calculated for two groups of patients or subjects and also their statistical difference in the survivals.

- Product Limit estimate (PLI) is another name of Kaplan Meier estimate. The product-limit formula estimates the fraction of organisms or physical devices surviving beyond any age *t*, even when some of the items are not observed to die or fail, and the sample is rather small. It involves computing the probabilities of occurrence of event at a certain point of time. These successive probabilities will be multiplied by any earlier computed probabilities to determine the final estimate.

- For example, the probability of a sub-fertile woman surviving the pregnancy three months after laparoscopy and hydrotubation can be considered to be the probability of surviving the first month multiplied by the probabilities surviving the second and third months respectively given that the woman survived the first two months. The third probability is known as a conditional probability (Etikan, 2017).

OBJECTIVES OF THE PAPER

The objectives of this seminar paper are to;

- introduce Kaplan Meier as a statistical model for survival analysis,
- identify various types of censoring in survival analysis,
- describe survival and hazard function in survival analysis and
- finally, introduce the modified model

LITERATURE REVIEW

- This section presents the brief literature about the Kaplan Meier and various types of censoring in survival analysis.
- Kaplan Meier is a non-parametric statistic that deals with time-to-event data, which analyze the patients or participants that will be lost to follow-up or dropped out of the study; those who will develop the disease of interest or those that will survive it.

Time -to- Event Data and Censoring

- Time-to-Event Data is the data that have the time as a principal end point as an event occurs. Some examples for time-to-event data is time until an electrical component fails, time needed to recover from illness and then the subjects in these studies may provide a survival time or a censoring time. If not all the subjects in data will have experienced the event at the end of the follow-up period, this is called censoring, meaning that the observation period ended without observing the event of interest.

Censoring may occur due to various reasons (Klein and Kleinbaum, 2005) listed three main reasons of censoring:

- 1- An object does not experience the event before the study ends.
- 2- An object is lost to follow-up during the study period.
- 3- An object withdraws from the study because of other failure reasons.

- It is not possible in these cases to know if and when the subject would have experienced the event. As an example, consider a study of a group of patients who suffer from a certain disease. Let T be the time to death after diagnosing the disease. Subjects who are still alive at the end of the study or who were dropped from the study and the current status is unknown, will be the censored observations. Consider another example as a study of a group of patients who received treatments for a certain illness. Let T be the time to cure after the treatment. In this case, censored objects are those who still need treatments at the end of the study.

Thus censoring is usually coded as follows;

$$\delta = \begin{cases} 1: \text{if not censored} \\ 0: \text{if censored} \end{cases} \qquad (2.1)$$

There are different censoring mechanisms (Abeysundara and Hemalika, 2010).

2.2 Censoring Mechanism:

- There are several mechanisms that can lead to censored data. We may distinguish the following types of censoring:

Type I censoring:

It refers to studies where all subjects enter the study at the same time and experiment is terminated at a specific time t_c

Hence;

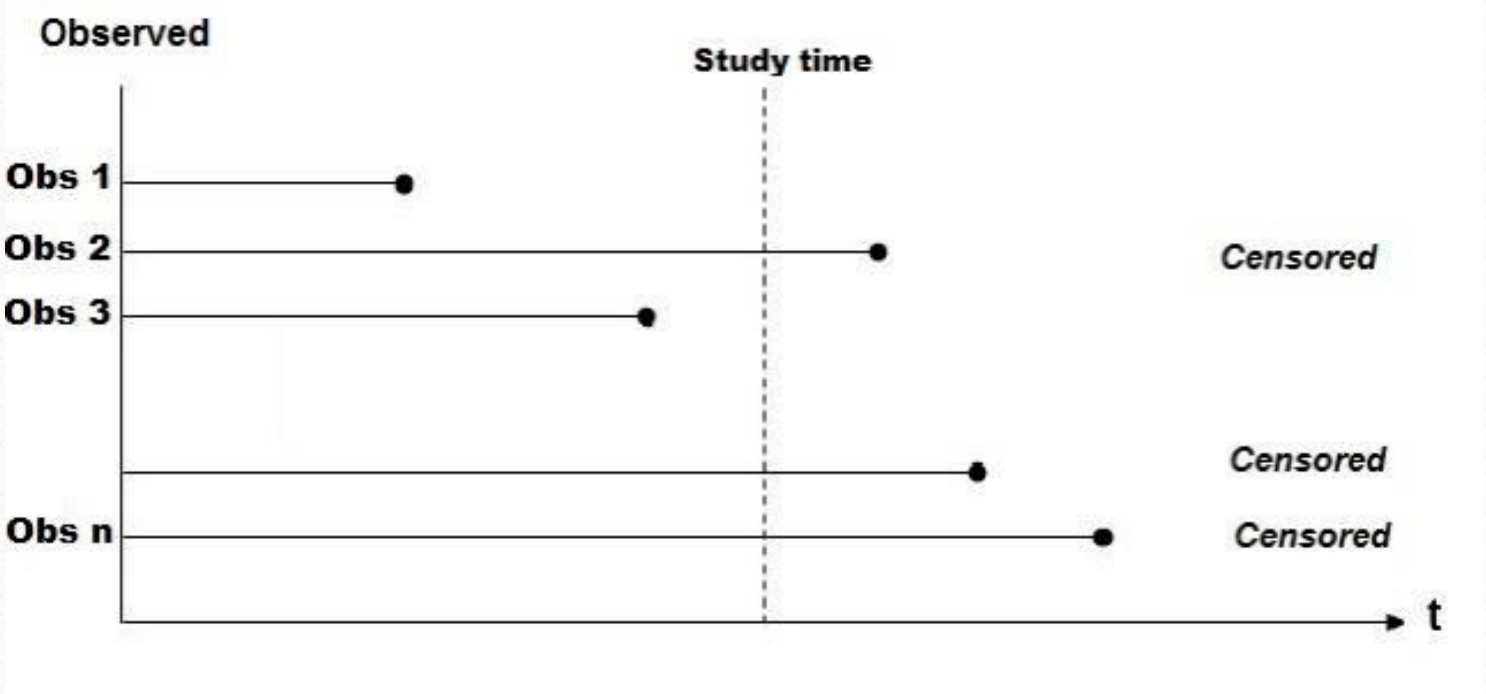
$$\delta = \begin{cases} 1: \text{if } T \leq t & \textit{event occurred} \\ 0: \text{if } t_c < T & \textit{censored} \end{cases} \qquad (2.2)$$

In type I censoring, if there are no accidental losses, all censored observations equal the length of the study period (Lee, 2003).

- In this type, the total duration of the study is fixed while the number of events (i.e, the number of individuals who have been exposed to the event) is a random variable. This kind of censoring is called (fixed censoring), within which the time of stopping the study is determined after a specified period of time (Fieller, 2011).

- Figure below shows the Type I censoring mechanism where the subjects who observed the event of interest after time t_c are the censored data (Abeysundara and Hemalika, 2010).

Figure 2.1: Type I censoring



- Up to date, there are five known censoring mechanism of Type I censoring occur in survival data, which are mainly classified according to the status of objects and time of checking up the occurrence of the desired event.

---

1. **Right censoring:**

- It is the most common situation in survival data, and this case is linked to the individuals that the event did not happen to it. Some individuals stay alive at the end of the study which means that the residence time of the individuals exceeds a point of the end of the study t_c and the individuals are called at the right censoring.

Right censoring may occur because of:

- 1. Researcher's decision to end the study before the event occurs.
- 2. The inability to reach the individuals for any reason.
- 3. Some individuals did not get the event.

2. Left censoring:

- Suppose that we have a single individual in the study, but the exposure time of the risk is unknown, for example, cancer patients and AIDS patients. The beginning of the infection time is unknown but death time is well known because of this disease. This individual is called left censoring.

- **3. Interval censoring:**
- In this case, the time of occurrence of the event is not exactly known for some individuals, but the known is the period of time in which the event occurred, and it is said about this individual's interval censoring (Fieller, 2011).

4. Double Censoring:

- When some of the data are censored on the left and some on the right, in other words if both left and right censoring occurs simultaneously. (Turnbull, 1976) introduced a special type so-called doubly censored, for example analyzed a real sample consists of 65 children from Nigeria are tested monthly, if they had learned certain tasks, double censoring occurred due to late arrivals (those who already learned the skills before entering the study) and losses (those who had not acquired the skill by the end of study time).

- **5. Middle Censoring:**
- Jammalamadaka and Mangalam (2003) introduced the middle censoring' scheme in non-parametric set up is called a general censoring scheme. If a data point is not observable when it falls inside a random interval here middle censoring is occur. The middle censoring scheme can be described as follows:

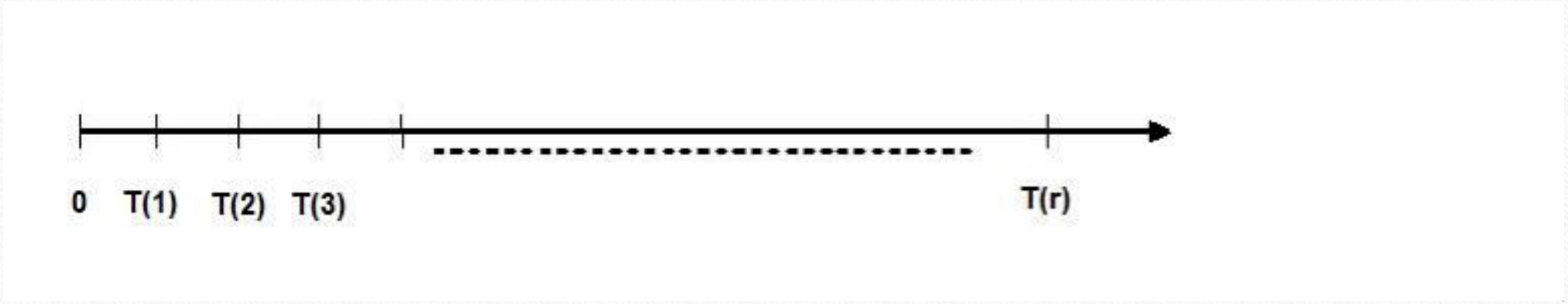
- Suppose n identical items are set in test and the lifetimes of these items are $T_1 \ldots T_n$. For the ith item, there is a random censoring interval $[L_i, R_i]$ which follows some unknown bivariate distribution. T_i is observable only if $Ti \in [L_i, R_i]$ otherwise it is not observable.

- Iyer, *et al.* (2008) presented the analysis of middle censored data with exponential lifetime distribution, and recently, Bennett (2011) explored middle censoring for further parametric models like the Wei-bull and Gamma families and for parametric models with covariates, recently Abuzaid, *et al.* (2015) studied the robustness of middle censoring.

2.2.2. Type II censoring

- In this case, the number of individuals that have known event occurs (fixed) in advance, while the total study period is a random variable which cannot be known in advance and the time of the end of the study is to be determined after certain number of cases occurrence of the event.
- Also we refer to the censoring as the type II censoring, when it is possible to terminate the experiment after r^{th} failure occurs out of n items.
- The first n observations will obtain survival times and the rest of the $n - r$ items will obtain failure time = T_r as explained in the Figure below (Abeysundara and Hemalika, 2010)

Figure 2.2: Type II censoring



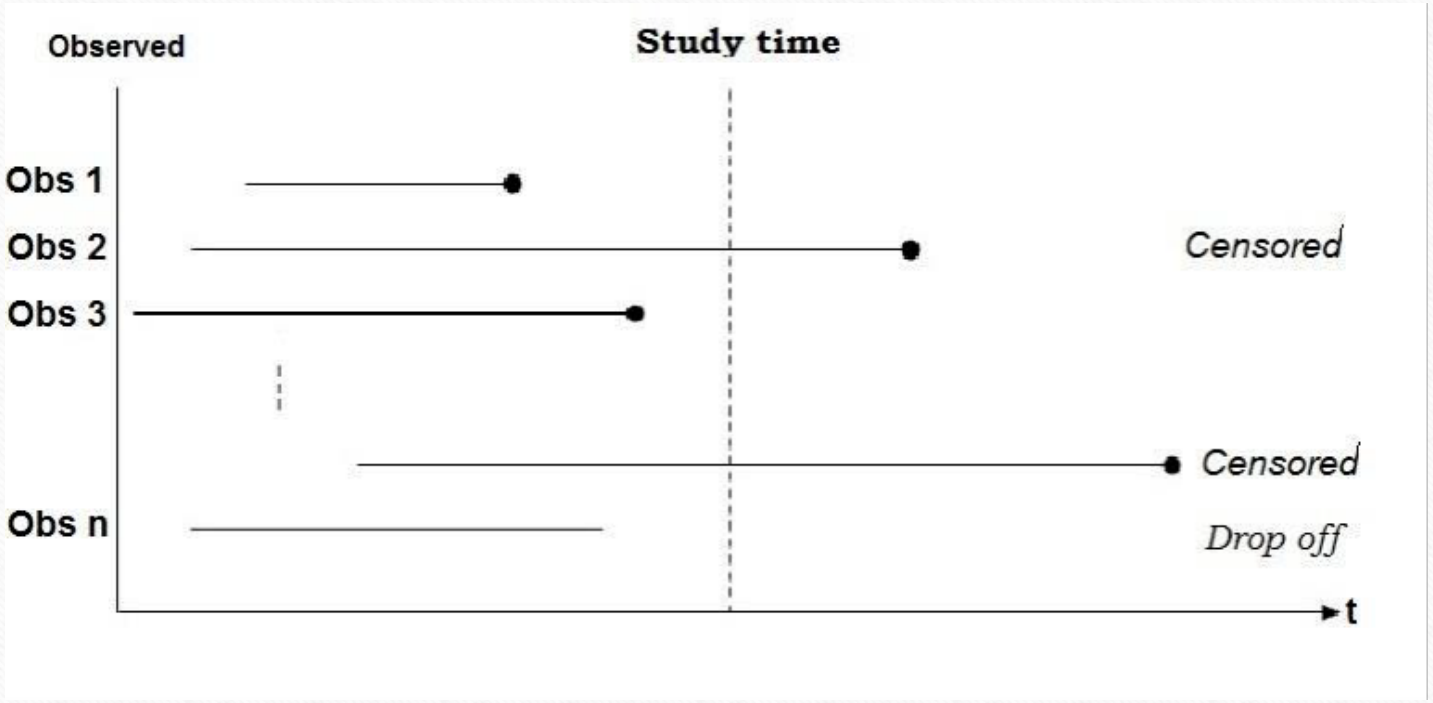
2.2.3 Type III censoring (Random Censoring)

- In this type, every individual has expected time censoring C_i and expected survival time T_i assumed that the time censoring and survival time of two random variables are independent. We note that $Y_i = \min(C_i, T_i)$ Y_i is a time of survival or censoring time whichever is less, and the cursor variable named d_i and tells us that the viewing ended in death or censoring, and this type is a combination of two previous types (Fieller, 2011).
- In real life practice, not all the subjects are enrolled at the same time and thus the follow-up period can vary from one subject to another. Random censoring allows subjects to enter the experiment at any time. Right censoring is presented here and left censoring is analogous.

They may;

1. Loss to follow up - we do not see them but only know that they are still alive - Censored.
2. Drop out/died - Uncensored
3. Survive at the end of the experiment – Censored

Figure 2.3: Random censoring



- A life time associated with a specific individual in a study is considered to be left censored if it is less than the censoring time. That is, the event of interest has already occurred before the beginning of the study. A study may contain right censored data, left censored data or both.
- Both means that the random censoring data that is a combination of right censoring and left censoring (Abeysundara and Hemalika, 2010).

2.2.4 Informative and Non-informative censored

- In informative censoring, the event and censoring rates are assumed to be the same conditional on the level of the covariates. Essentially means that within any subgroup of interest, the subjects who are censored at time *t* should be representative of all the subjects in that subgroup who remained at risk at time *t* with respect to their survival experience. In other words, censoring is independent provided that it is random within any subgroup of interest. So independent censoring is a less restrictive form of random censoring (where we would not be taking into account the survival profile by covariates) (Campigotto and Weller, 2014).

- Non-informative censoring is when time to event and time to censoring are independent conditional on the level of covariates. It occurs if the distribution of survival times (*T*) provides no information about the distribution of censorship times) (*C),* and vice versa. The assumption of non-informative censoring is often justifiable when censoring is independent and/or random; nevertheless, these assumptions are not equivalent. In non-informative censoring we assume that the time to censorship distribution is not related to the time-to-event distribution (e.g. if a patient in a study received the event, then another patient in the study is selected randomly to leave the study). (Kleinbaum and Klein, 2011)

2.3.1 The Survival Function

- Survival function is the probability that the random survival time variable T is greater than or equal to a specific t. Let F (t) be the cumulative distribution function of t and the survivor function is the right tail probability, where $T>0$ have a pdf $f(t)$ *and cdf* F (t). (Smith, *et al.*, 2003).

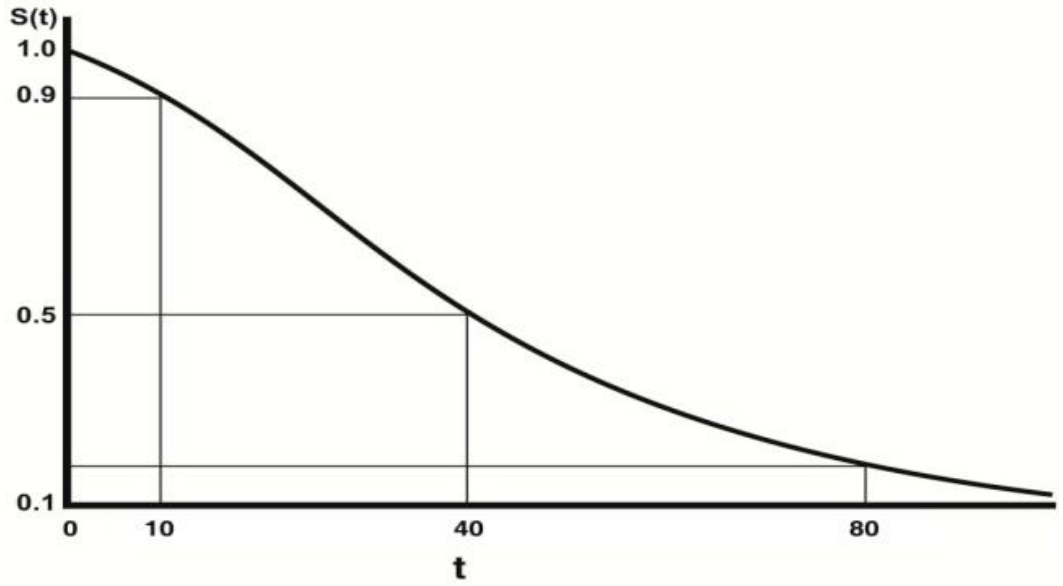
The purpose of survival analysis is to compare and estimate survival experiences of different groups and it can describe survival experience by the cumulative survival function, then the survival function takes on the following form:

$$S(t) = \mathrm{P}\{T>t\} = 1 - F(t) \qquad (2.3)$$

Where, $F(t) = \mathrm{P}(T<t) = \int_0^t f(u)\,du$

Survival function gives the probability of surviving or being event-free beyond time t . Because $S(t)$ is a probability, it is positive and ranges from 0 to 1, it is defined as $S(0) = 1$, and as t approaches ∞ , $S(t)$ approaches 0 . Survival curve describes the relationship between the probability of survival and time.

F ıe



- They are non-increasing; that is, they head downward as t increases. At time as $t \longrightarrow \infty$, $t = \infty$, $S(t) = S(\infty) = 0$; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually fall to zero.

- The time when $S(t) = 0.5$ is termed the life expectancy in the population. Usually in practice, we don't reach the median survival at exactly one of the failure times. In this case, the estimated median survival is the smallest time T such that $\hat{S}(t) \leq 0.5$ (Sameer, 2009).

- Also, the median survival time could be simply estimated by the sample median of survival times with every survival time observed exactly in the absence of censoring, while in the presence of censoring, we need to use the Kaplan–Meier estimate $\hat{S}(t)$ to estimate the median (Vittinghoff, *et al.*, 2004).

2.3.2 The Hazard Function

- We often modeling the lifetime through the hazard function $h(t)$ which measures the 'risk' or 'proneness' to death at time t, given survival up to time t. It is the probability that an individual die at time t, conditional on him having survived to that time. Hazard function represents the instantaneous death rate for an individual surviving to time t (Fieller, 2011).

- The hazard function describes the concept of the risk of an outcome (e.g., death, failure and hospitalization) in an interval after time t, conditional on the subject having survived to time t. It is the probability that an individual die somewhere between t and $t+\Delta$, divided by the probability that the individual survived beyond time t.
- It seems that the hazard function be more intuitive to use in survival analysis than the *pdf* because it aims to determine the size of the instantaneous risk that an event will take place at time t given that the subject survived to time t (Smith, *et al.*, 2003).

- The hazard function species the instantaneous rate of failure at $T = t$ given that the individual survived up to time t and is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t} = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \qquad (2.4)$$

- The hazard function is also referred to as the risk or mortality rate. The hazard is a rate rather than a probability. It can assume values in $(\infty, 0)$. It is easily verified that $h(t)$ specifies the distribution of T, since

$$h(t) = lim \frac{-dS(t)/dt}{S(t)} = -\frac{dlog(S(t))}{dt} \qquad (2.5)$$

- The cumulative hazard function is the integral of the hazard function. It can be interpreted as the probability of failure at time x given survival until time x:

  - $$H(t) = \int_{-\infty}^{x} h(t)dt \qquad (2.6)$$

- This function is supported for continuous distributions only. It can be expressed about the cumulative sum of the hazard probability function as:

  - $$H(t) = -\log S(t) \qquad (2.7)$$

- The relationship between survivor function and hazard function can be summarized as follow:

1. Survivor function, $S(t)$ defines the probability of surviving longer than time t.

2. Hazard function is the first derivative of the survivor function over time

   $h(t) = dS(t)/\ d(t)$.

3. Instantaneous risk of event at time t (conditional failure rate).

4. Survivor and hazard functions can be converted into each other.

(Gage, 2004),.

2.4 Estimation of the Survival Function

- The estimation of survival function can be obtained either parametrically or non-parametrically.
- There are three main assumptions that are necessary to ensure the correct estimate of the survival probability in presence of censored data. These assumptions are as follows: (Abeysundara and Hemalika, 2010);

1. The subjects who are censored have the same survival distribution as those who continue the observation during the study. That is the censoring process which is independent of the primary endpoint.

2. The subjects who first joined the study have a longer observation period and then have more chances to experience the event than subjects enrolled at the end of recruitment period.

3. Survival probability is the same for subjects enrolled at the beginning and at the end of recruitment period. And then for some events such as diagnosing a cancer, it is not possible to identify the exact date when the event happens.

- We know that the survival function $S(t)$ gives the probability that a person survives longer than some specified time t : that is, $S(t)$ gives the probability that the random variable T exceeds the specified time t and survivor function in practice is estimated as:

$$S(t) = \frac{Number\ of\ individual\ in\ the\ sample\ who\ survived\ longer\ than\ t}{Total\ number\ of\ individuals\ in\ the\ sample} \qquad (2.8)$$

- We have previously mentioned that the hazard function denoted by $h(t)$, also called the instantaneous failure rate, conditional mortality rate, force of mortality and age specific failure rate.

- In practice, the hazard function is estimated as the proportion of patients dying in an interval per unit time, given that they have survived to the beginning of the interval:
- $$\hat{h}(t) = \frac{Number\ of\ patients\ dying\ per\ unit\ time\ in\ the\ interval}{Number\ of\ patients\ surviving\ at\ time\ t} \qquad (2.9)$$
- A non-parametric estimation of survival and hazard function are obtained by Kaplan-Meier. Besides the description of a variety in two or more groups of the estimated survival time distributions and the plots of the survival rates are simply start of the survival analysis, researchers require a statistical test to conclude that, these differences are statistically significant or caused by "chance variation".

Kaplan-Meier Survival Curves

- Kaplan and Meier (1958) put forth a new, efficient method for estimating patient survival rates, taking into account the fact that some patients may have died during a research trial while others will survive beyond the end of the trial. The method is derived based on information from those who have died and those who have survived to estimate the proportion of patients alive at any point during the trial is called the Kaplan-Meier estimator (also known as the product limit estimator).

- The estimator is plotted over time and the resulting curve, which is a series of horizontal steps of declining magnitude that, when a large enough sample is taken, approaches the true survival function for that population, this is called the Kaplan-Meier curve. In medical research survival curves are almost universally generated by the Kaplan–Meier method (Costella, 2010).

- The Kaplan-Meier estimate is the simplest way of computing the survival over time despite all these difficulties associated with subjects or situations. An important advantage of the Kaplan–Meier curve is that the method can take into account some types of censored data, particularly right-censoring, which occurs if a patient withdraws from a study. On the plot, small vertical tick-marks indicate loss, where a patient's survival time has been right-censored. When no truncation or censoring occurs, the Kaplan–Meier curve is the complement of the empirical distribution function.

- Let $S(t)$ be the probability that a member from a given population will have a lifetime exceeding time, t. For a sample of size N from this population, let the observed times until death of the sample members be N $t_1 \leq t_2 \leq t_3 \leq \ldots \leq t_N$.

Corresponding to each t_i is n_i, the number "at risk" just prior to time t_i, and d_i, the number of deaths at time t_i.

- The Kaplan–Meier estimator is the nonparametric maximum likelihood estimate of $S(t)$, where the maximum is taken over the set of all piece-wise constant survival curves with breakpoints at the event time t_i. It is a product of the form

- $$\hat{S}(t) = \prod_{t_{i<t}} \frac{n_i - d_i}{n_i} \qquad (2.10)$$

- When there is no censoring, n_i is the number of survivors just prior to time t_i. With censoring, n_i is the number of survivors minus the number of losses (censored cases). It is only those surviving cases that are still being observed (have not yet been censored) that are "at risk" of an (observed) death (Collett , 2003).

Non-parametric estimators of the hazard function – the Kaplan-Meier estimator

- Let $\{T_i\}$, i = 1,…, n are independent and identically random variables with distribution function F and survival function $S(t)$. However, we do not know the class of functions from which F or $S(t)$ may come from. Instead, we want to estimate $S(t)$ non-parametrically, in order to obtain a good idea of the 'shape' of the survival function. Once we have some idea of its shape, we can conjecture the parametric family which may best fit its shape. If the survival has not been censored the 'best' non-parametric estimator of the cumulative distribution function F is the empirical likelihood

$$\hat{F}n(t) = \sum_{i=1}^{n} I_{(t \le Ti)} \qquad (2.11)$$

- Since $\hat{F}n\,(t)$ is an estimator of the distribution function F where I is an indicator function. It is clear that an estimator of the survival function S (t) is $\hat{S}n\,(t) = 1 - \hat{F}\,n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{(t \le Ti)}$ and the maximum likelihood estimator of the hazard

$\hat{h}_s = P(T = t_s)/P(T \ge t_{s-1})$ is $\hat{h}_s = \frac{d_s}{N_s}$

(2.12)

where d_s are the number of failures at time t_s and N_s are the number of survivors just before time t_s.

- In many respects, this is a rather intuitive estimator of the hazard function. For example, if there is no censoring then it can be shown that maximum likelihood estimator of the hazard is

$$\hat{h} = \frac{d_s}{\sum_{i=s}^{\infty} d_s} = \frac{number\ of\ failures\ at\ times}{number\ who\ survive\ just\ before\ times} \qquad (2.13)$$

which is a very natural estimator.

In this section, methodology used on survival analysis is basically described below;

$\mathbf{S}_t$ = *Number of subjects living at the start – Number of subjects that died* (3.1)

Number of subjects living at the start

$$\hat{S} = \prod_{i:t_i \leq t} . \ (1 - \frac{d_j}{n_j}) \qquad (3.2)$$

t_i: A time when at least one event happened.

d_i: Number of deaths that happened at time t_i

n_i: Individual known to survive (censored) at time t or number of subjects at-risk at i^{th}: ordered time

$\boldsymbol{d_i}$: Number of events, $\boldsymbol{n_i}$: total individuals at risk at time t_i.,

$\boldsymbol{h_i}$: discreet hazard rate (probability of an individual with an event time t_i.

$$\mathbf{S}(t) = \prod (1 - h_i)$$

$$L\left[h_{j:j\le i}/d_{j:j\le i}, n_{j:j\le i}\right] = \prod_{j=1}^{i} h^{d_j}. (1 - h_j)^{n_j - d_j} \qquad (3.3)$$

Therefore, the log likelihood will be;

$$\log(L) = \sum_{j=1}^{i} (d_j \log(h_i) + (n_j - d_j)\log(1 - h_i)) \qquad (3.4)$$

Finding the maximum of log likelihood with respect to $\boldsymbol{h_i}$ yield;

$$\frac{\delta \log(L)}{\delta h_i} = \frac{d_i}{\hat{h}_i} - \frac{n_i - d_i}{1 - \hat{h}_i} = 0 \rightarrow \hat{h}_i = \frac{d_i}{n_i} \qquad (3.5)$$

where $\hat{h}_i$ is used to denoted maximum likelihood estimation given this result, we can write;

$$\hat{S} = \prod_{i:t_i \le t} (1 - \hat{h}_i) = \prod_{i:t_i \le t} (1 - \frac{d_i}{n_i}) \qquad (3.6)$$

- **3.1 Survival Model:**

$$\hat{S}(t) = \prod (1 - \frac{d_i}{n_i}) \qquad (3.7)$$

$$= (1 - \frac{d_1}{n_1}) \times (1 - \frac{d_2}{n_2}) \times \ldots \times (1 - \frac{d_i}{n_i})$$

t_i: i^{th} ordered follow-up time

d_i: Number of deaths at i^{th} ordered time

n_i: Number of subjects at-risk at i^{th} ordered time

3.2 Modified Model

$$\hat{S}(t) = \prod \left(1 - \frac{d_j}{n_j}\right)\theta_i \qquad (3.8)$$

Where $\theta = \alpha_{t_i} * \beta_{t_i} * \tau_{t_i}$

$$= (1 - \frac{d_1}{n_1})\,\theta_1 \text{ x } (1 - \frac{d_2}{n_2})\,\theta_2 \text{ x} \ldots \text{x } (1 - \frac{d_j}{n_j})\,\theta_n \qquad (3.9)$$

t_i: i^{th} ordered follow-up time

d_i: Number of deaths at i^{th} ordered time

n_i: Number of subjects at-risk at i^{th} ordered time

α_{t_i}= Initiation of evaporated milk/milk extract at time t

β_{t_i}=Initiation of natural food supplement (water, cooked grains, etc.) at time t

REFERENCES

- Abeysundara, H. and Hemalika, T. K. (2010). Non-Parametric Estimation of Bivariate Survival Function,Master Thesis. Texas Tech University.
- Abuzaid, A., Abu El-Qumsan,M. and El-Habil, A. (2015). On the Robustness of Right and Middle Censoring Schemes in Parametric Survival Models. *Communications in Statistics-Simulation and Computation* (DOI: 10.1080/03610918.2015.1011337)
- Bennett, N. A. (2011). Some contributions to middle censoring, Phd. Dissertation, Department of Statistics and Applied Probability, University of California Santa Barbara.
- Campigotto, F. and Weller, E. (2014). Impact of informative censoring on the Kaplan-Meier Estimate of Progression-Free Survival in phase II Clinical Trials. *American Society of Clinical Oncology, Boston.*
- Collett, D. (2003). Modeling survival data for medical research, London, UK: Chapman-Hall. 2nd Edition.
- Costella, J. P. (2010). A Simple Alternative to Kaplan–Meier for Survival Curves. Peter MacCallum Cancer Centre.
- Fieller, Dr N. (2011). *Medical Statistics*: *Survival Data.* University of Sheffield.
- Gage, B. F. (2004). *Introduction to Survival Analysis*. Division of General Medical Sciences, Washington University.

- Ilker Etikan, Suleiman Abubakar, and Rukyya Alkassim (2017) **The Kaplan Meier Estimate in Survival Analysis: Biometrics & Biostatistics International Journal** Volume 5 Issue 2 - 2017 *Department of Biostatistics, Near East University Faculty of Medicine, Cyprus (Article)*
- Iyer S.V., Jammalamadaka S.R. and Kundu, D.(2008). Analysis of Middle-Censored Data with Exponential Lifetime Distribution. *Journal of Statistical Planning and Inference*, **138,** 3550-3560.
- Jammalamadaka, S. R. and Mangalam, V. (2003). Nonparametric Estimation for Middle-Censored Data, *Nonparametric Statistics*, **15** (2), 253–265.
- Kaplan, E. L. and Meier P. (1958). Non Parametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**, 457-481.
- Kleinbaum, D. G. and Klein, M. (2005). Survival Analysis. A Self-Learning Text, Wiley, New York, *2nd edition*.
- Kleinbaum, D.G. and Klein, M. (2011). Survival Analysis. A Self-Learning Text. Springer.
- Lee E.T., (2003). *Statistical Methods for Survival Data Analysis*. Wiley New York, third edition.
- Patrick Schober, MD, PhD, MMedStat, and Thomas R. Vetter, MD, MPH(2018): Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare(Article).

- Sameer, F. O., (2009). Cox Proportion Hazard Model for Patients with Hepatitis Disease in Iraq. *Baghdad Science Journal*, **3,** 612-617.
- Shostak, J. and Walker, G. A. (2010). *Common Statistical Methods for Clinical Research with SAS Examples,* Third Edition. Cary North Carolina, USA. SAS Institute Inc.
- Siddiqa,M. (2012). Survival Analysis of Dialysis Patients under Parametric and Non parametric approaches. *Electronic Journal of Applied Statistical Analysis*,**2**, 271 – 288.
- Smith, T., Smith, B. and Ryan, M. AK. (2003). Survival Analysis Using Cox Proportional Hazards Modeling For Single and Multiple Event Time Data Survival Analysis Using Cox Proportional Hazards Modeling For Single and Multiple Event Time Data, 28 – 254.
- Turnbull, B.W.(1976). The Empirical Distribution Function with Arbitrarily Grouped censored and Truncated Data, *Journal for the Royal Statistical Society*, **38**, 290-295.
- Turnbull, B.W., (1974). Nonparametric Estimation of a Survivorship Function with Doubly Censored Data. *J. Amer. Statist. Assoc.***69**, 345,169-173.
- Vittinghoff, E., Shiboski, S. C., Glidden, D. V. and McCulloch, C. E. (2004). Regression Methodsin Biostatistics Linear, Logistic, Survival and Repeated Measures Models.