

Review of prostatic tumor using kaplan meier and cox regression

Abstract

The research look into a prostatic tumor study using Kaplan Meier and Cox regression from Survival Analysis to evaluate the effectiveness of a treatment method. A total of thirty eight (38) patients were employed for the study, from the Second Edition of David Collett 2003; Modeling Survival Analysis Data in Medical Research. The data obtained was analyzed using Statistical Packages for the Social Sciences (SPSS) version 18. The analysis covers treatment, survival time, event, age, serum hemoglobin, tumor size and Gleason index of the prostatic tumor patients. The descriptive statistics showed that 47.4% of the patients were administered placebo while 52.6% were subjected to the actual treatment method. According to the Patient status, 84.2% of the patients were still alive responding to treatment while 15.8% were reported dead. In terms of age category, 55.3 % of the patients were between ages 71-80 years old, 28.9% were between ages 61-70 years old while 15.8% were between ages 51-60 years old respectively. The log rank test of the Kaplan Meier analysis revealed that patients subjected to an actual treatment stand a good chance of longer survival than the placebo patients' cohort. While the cox regression output shows that hazard occurrence between placebo and treatment group is not statistically significant. That is, there is no evidence of a greater risk of death in either of the patient groups.

Keywords: survival analysis, censoring, kaplan meier method

Volume 6 Issue 5 - 2017

İlker Etikan, Kabiru Bala, Ogunjesa Babatope, İsmail Bakır and Meliz Yuvalı

Department of Biostatistics, Near East University, Cyprus

Correspondence: İlker Etikan, Near East University, Faculty of Medicine, Department of Biostatistics, Nicosia-TRNC, Cyprus, Email ietikan@gmail.com; ilker.etikan@neu.edu.tr

Received: August 20, 2017 | **Published:** December 06, 2017

Introduction

The branch of statistics relative to analysis of predictable period of time until one or more event occurs is termed survival analysis. Survival analysis is a system of data collection and analysis where the outcome variable is the time until the event of an interest occurs.¹ The technique deals with the creation of timing data that goes with event of either failure or death. However the pattern of complete with incomplete data is a major characteristic of this technique because some people will experience the event of interest while others will not experience such event. This method of analysis is widely applied in different fields of studies. In the biological field, it could be considered as the death of organism prior a specific time and this is generally regarded as survival time analysis. In the engineering field, it is called reliability theory which analyses the failure of systems prior a stipulated time or period. In finance and economics, such analysis is called duration analysis while in the field of sociology, it is the event of history analysis.^{2,3}

The observations that failed to experience the event of interest in survival analysis are generally refers to as censored observations and this occurrence is called **censoring**.⁴ Censoring generally ensued when we have limited survival time knowledge about a person or object but the survival time is not exactly known. The incomplete information about such observations is treated as missing data which

likewise form an integral part of survival analysis. There are three major reasons why censoring happen;⁵ The person will not experience the event before the study ends; The person will be lost to follow-up during the study period and finally; The person may possibly withdraw from the study because of death or other reasons such as poor drugs reaction or other competitive risk (diseases).

Significance of the study

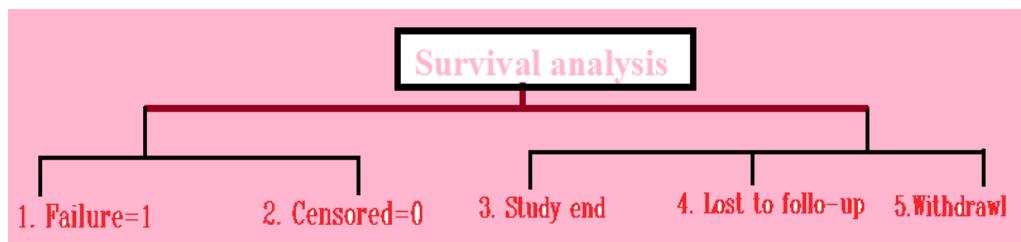
The research study will help to ascertain how Kaplan Meier and Cox regression from survival analysis assist in determining the level of individuals in terms of treatment subjected to.

Aims and objectives of the study

1. To apply survival techniques to evaluate factor (forms of treatment) efficiency
2. To determine which group stands a better chance of survival
3. To make a general conclusion base on discovered findings

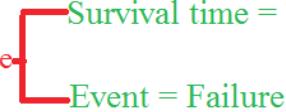
Scope and limitation of the study

The study made use of data of 38 patients with prostatic tumor sourced from the Second Edition book of David Collett⁶ on Modeling Survival Analysis Data in Medical Research.



Time in survival analysis: This is the beginning of any level of observation or follow-up of individuals pending an event occurring, which could be hourly, daily, weekly, monthly or yearly.⁷ Also, it could be the age of an individual when age is considered an event occurrence that signifies the time. Time in survival analysis is continuous in nature.

Event in survival analysis: Event of interest can be seen as diseases incident, death, re-lapse from reduction, recovery, returning to work, beginning of any treatment or surgery, losing of contact, divorced, marriage, withdrawal from the study, end of the study or any kind of designated experience of interest that may happen to an individual.⁸

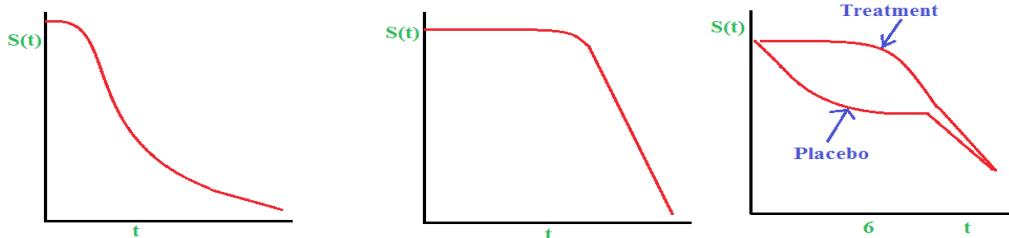
SURVIVAL ANALYSIS	OUT COME VARIABLE	EVENT
Statiscal method of data analysis procedures which the outcome variable of interest is time until the occurrence of an event	<p>Start follow-up</p>  <p>Time</p>  <p>Event</p> <p>Survival time = Time</p> <p>Time</p>  <p>Event = Failure</p>	<p>Death, re-lapse, disease, recovery etc</p> <p>Recurrent event</p> <p>> 1 Event</p> <p>Competing risk</p>

Recurrent event: Event of interest usually occurs more than once in biomedical investigation. However greater part of analyses focuses only on time to the first event by overseeing the succeeding events (i.e. the relationship between those events arising in the same subject when the association of the data is ignored is a common feature among these events). Recurrent event data has two main features; the events are systematic and the study can only be at risk for one event at a particular time. There are so many factors used in determining the idea of analysis of recurrent event, such as number of the events, relationship linking the following of the events, effects of changing, biological process and so on.⁹ Furthermore we have other models used in describing the link between recurrent events such as covariance matrix, frailties or time varying covariates (marginal means). Frailties

models are specifically used for repeated events with a constant hazard between recurrences and such events model assist in acquiring the understandings of development of disease. Examples of recurrent events are; admissions to hospitals, migraines, cancer recurrences, upper respiratory and ear infections etc.

Aims and objectives of survival analysis

- To assess and explain survivor hazard functions from survival numbers.
- To measure up survivor function or the hazard function.
- To evaluate the association of growing variables to the survival time.



The three graphs above show different survivor functions. The first graph shows a fast fall in survival probabilities at the early stage during the study period but leveling at the later stage with a very slow decrease in survival probabilities at earlier point of follow-up. There was a sharp decrease from the second graph.

The third graph is the combination of plots of survivor functions for treatment group with the group of placebo by graphing the functions on the equal position. As it was observed getting to the six weeks, the survivor function for treatment group lies over the placebo group while at a point the two functions were about to be on the same level. This indicates that up to six weeks of follow-up, the treatment group survives more actively than the placebo then later on has the same effect.

Uses of survival analysis in different viewpoint

- It is used in defining the survival participants of a group, such as Kaplan Meier, life table, survival function and hazard function.
- Used in comparing the survival time of two or more units by making use of Log-rank test.
- It is used in describing the outcome of categorical or quantitative

variables on survival using techniques such as survival trees, Cox proportional regression model, survival random forests, and parametric survival models etc.

Censoring

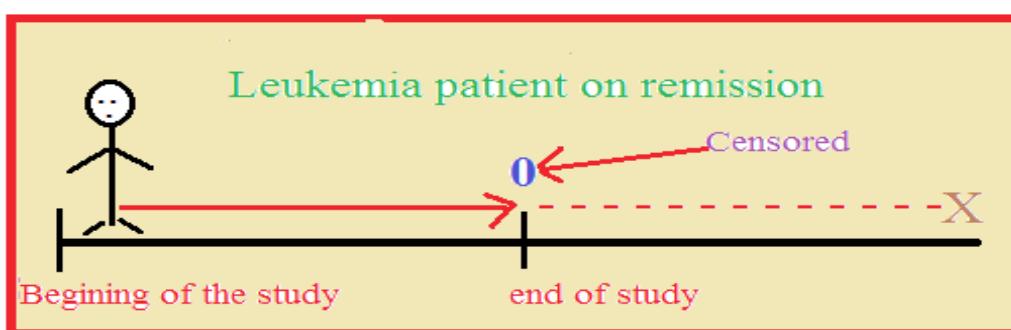
As earlier stated, the term Censoring usually surfaced when we have limited available survival time knowledge of information about a person but the survival time is not correctly proven. Furthermore, it can be referred to a kind of problematic occurrence of missing data that is universally applied in survival analysis.

Reasons for the censoring

Generally we have three reasons for censoring to take place which are;

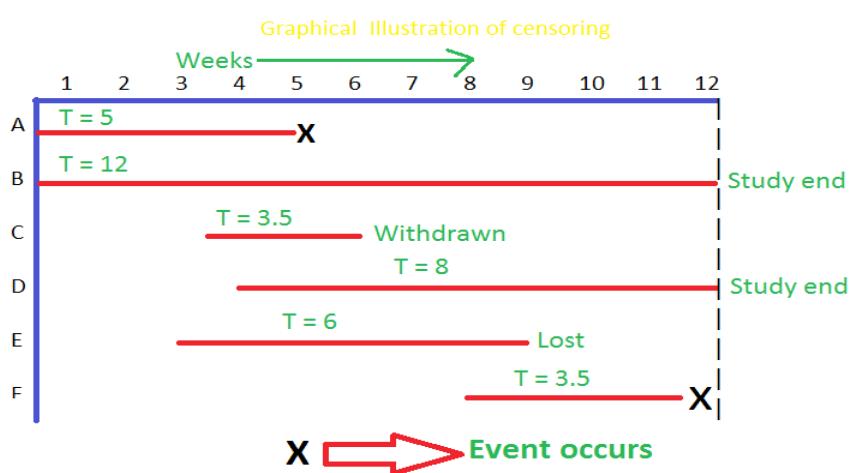
- Any individual that experience the event sooner than the end of the study
- A person that is nowhere to be found during the study period

A person missing from the study due to either event of death or withdrawal



The example above depicts a censored occurrence for a patient with leukemia. Followed until the time of reaching remission as indicated by X. This graphical image shows that the event of interest

does not happen when the patient is still in remission and the patient survival time is believed to be censored.



The above graphical presentation gave a clearer understanding of several persons followed up over a period of time. The person getting the event at X shows such observation is not censored. Assuming for example, participant A was followed from the beginning of the

study and the event occur at week five, participant A is not censored and his survival time is 5 weeks. Participant B also was followed up from the beginning of the study up to the end of week 12 without experiencing the event of interest. The non-occurrence of the event

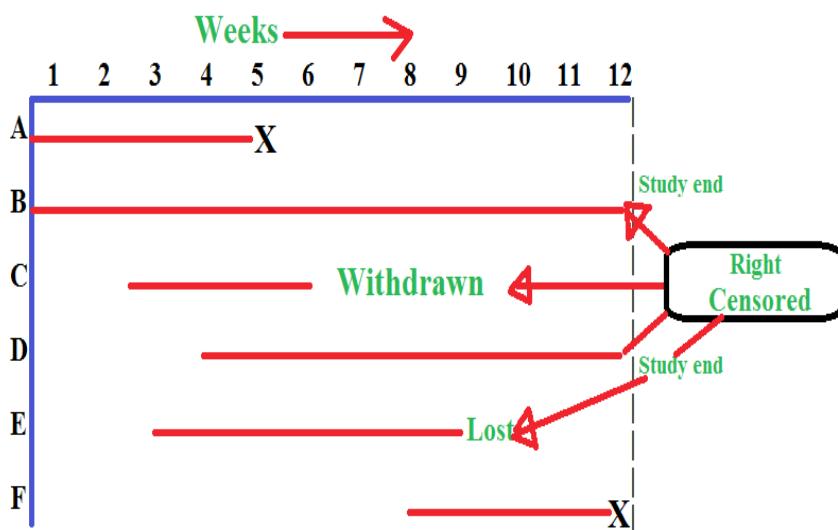
of interest within the period of the study made Participant B to be a censored participant. Participant C joined the study in between week 3 and 4. He was followed up until week 6 when he withdrew from the study. The withdrawal makes participant C a censored participant with a survival time of 3.5 weeks. Participant D joined the study at the beginning of week 4 and followed up until the end of the study. He became a censored participant at week 8 without experiencing the event of interest. His total time enrolment in the study is 8 weeks. Participant E was followed up from week 3 of the study until week 9

before missing out in the study. Missing out of the study makes him a censored participant with a total enrolment time of 6 weeks. Finally, participant F enrolled in the study at week 8 and experienced the event of interest at week 11.5. Hence, participant F is not a censored participant due to the occurrence of the event of interest. His total enrolment time before the occurrence of event of interest is 3.5 weeks. In summary participants A and F are uncensored observations while participants B, C, D, and E were censored participants.

Person	Survival time	Failed (1) & Censored (0)
A	5	1
B	12	0
C	3.5	0
D	8	0
E	6	0
F	3.5	1

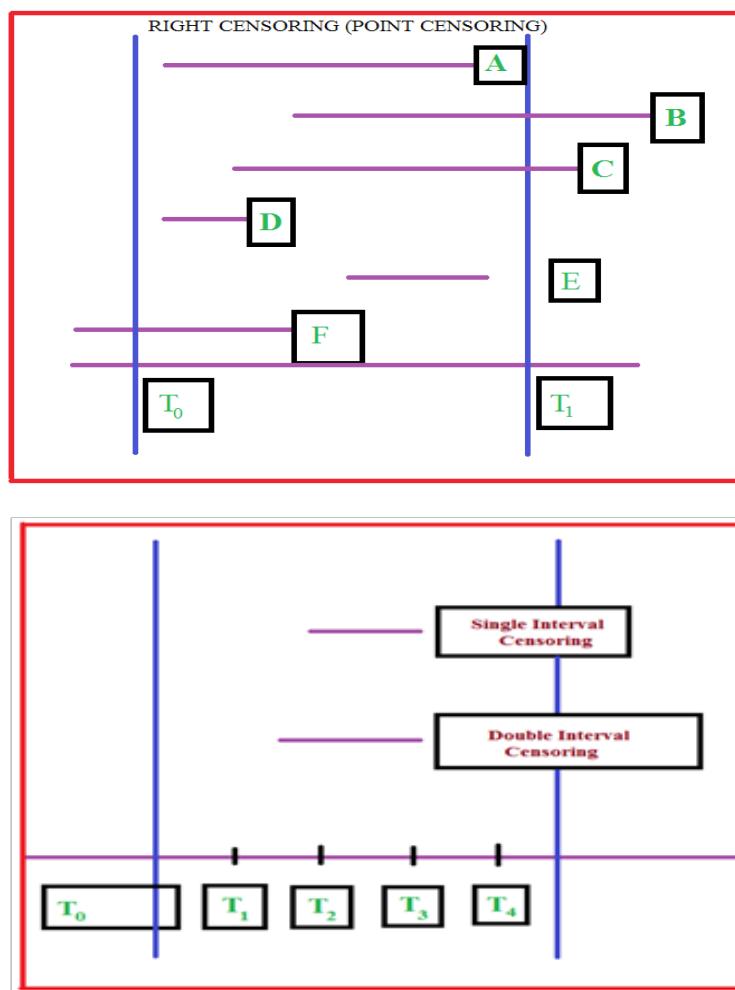
The presentation above gave the survival data for the six participants with each assigned with their corresponding survival time up to the occurrence of the event or to a censorship. The status function for censoring or not censoring is reveal in the last column with one (1)

signifying failed event and zero (0) denoting censored respectively. Since the censoring indicator is zero for participant C, the observed time 3.5 is an **event time**. While for uncensored participant F, the **survival time** was 3.5.



The diagram above displayed each of the four participants that were censored. When the participant survival time is incomplete by the right side during the study, ending of experiment, when the person is lost or withdrawn during the follow up, it can be regarded as a right censoring technique. The demonstration above is a right censored

illustration because complete survival time interval for the data is not seen, implying censored at the right side of the surveyed survival time interval. Although survival time data could be left censored but most survival data appears to be right censored.



Right censoring: Giving that the event of interest is death in a study, Right Censoring of data often occur when participants are alive even when the study ended, or are lost to follow up or when the study ends abruptly without the participants experiencing the expected even in the initial specified duration. Or perhaps, participants die from other causes independently from the cause of interest and or lost to the search, by dropping out, or getting transferred to another different area.

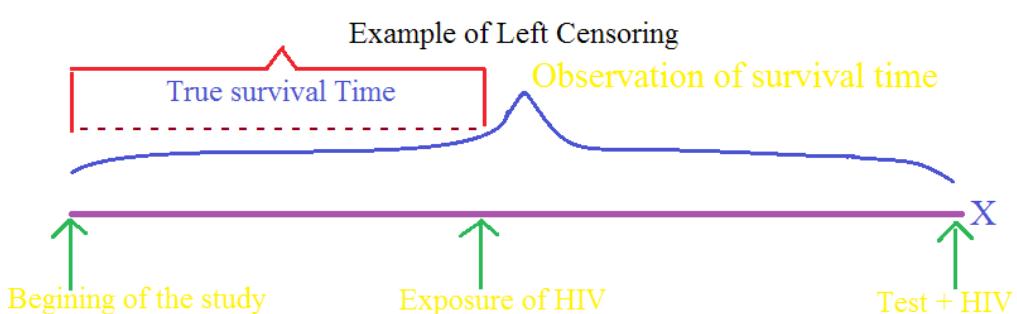
Types of independent right censoring

Type I: This is a complete and deliberate drop- out from a study. It could happen as a result of relocation or fixed time allocated to end the

study without the occurrence of any event of interest.

Type II: Study ends when a fixed number of events amongst the subjects have occurred or a study ends when there is specific number of events expected.

Left censoring: This occur when the subjects of a study already manifested the event of interest at the beginning of the study period while the evidence about when the first realization of the event is not sure. This can occurred usually when a person's survival time turn out to be incomplete on the left side during the study period. More so, the left censored survival data occurred when the subject's real survival time becomes incomplete on the left side of the follow-up period.¹⁰



Interval censoring: This type of censoring happen when time of event may be known only for a particular time interval. It may occur when a subject is taking periodic follow-up during an experiment, frequently inspection of equipment and so on. In this case the exact time of the occurrence of the event will not be identified, but interval timing can be detected. Generally, any condition of measurement during an observation which may be perform at a particular period especially when life time is known to fall within an interval is referred to as interval censoring.⁸

Truncation: This occurs when there is an experience of delay in some unique study types which makes it difficult to experience the occurrence of an event of interest. In this concept, subjects with truncation might need to reach a certain level of age before they will be observed or when a subject life time is less than some certain level that could be observed. These types of data are mostly found in actuarial work for pension and life insurance.

Fitting of parameter data for survival analysis

Survival data cannot be analyzed by ordinary least square regression methods especially as a result of censored data it contains. The likelihood function of survival model in the presence of censored data is explainable by means of conditional probability of a data given the parameter of the model, and taking into consideration that the data are given independent parameters while the likelihood function is the product of the likelihood of every datum.¹¹ This can be partition into four distinct categories which are as follows; the right censored, the left censored, the interval censored and the uncensored.

Types of censored base on definition

$$R(\theta) = \prod_{i \in r.c.} \Pr(T > T_i | \theta)$$

The right censored: The right censored data, were age at death is known to be greater than T_i will be

$$\Pr(T > T_i | \theta) = 1 - F(T_i | \theta) = S(T_i | \theta).$$

$$L(\theta) = \prod_{i \in l.c.} \Pr(T < T_i | \theta)$$

The left censored: A left censored data, were age at death is known to be less than T_i will be $\Pr(T < T_i | \theta) = F(T_i | \theta) = 1 - S(T_i | \theta)$.

$$L(\theta) = \prod_{i \in l.c.} \Pr(T_{i,l} < T_i < T_{i,r} | \theta)$$

The interval censored: The interval censored of data, were age at death is known to be less than $T_{i,r}$ and greater than $T_{i,l}$ will be

$$\Pr(T_{i,l} < T_i < T_{i,r} | \theta) = S(T_{i,l} | \theta) - S(T_{i,r} | \theta).$$

$$L(\theta) = \prod_{i \in unc.} \Pr(T = T_i | \theta)$$

The uncensored: Uncensored data with T_i equals to death, will be $\Pr(T = T_i | \theta) = f(T_i | \theta)$.

The model of survival analysis are as follows;

- Parametric approach
- Non-parametric approach

C. Semi-parametric approach

Distributions of parametric survival models

Parametric survival model are predictive modeling techniques for survival data with outcomes that are known to follow some probabilities distribution properties.¹² Linear regression, logistic regression, and Poisson regression are examples of parametric models that are commonly used in the field of health science were the outcomes of the model assumed to follow some distribution such as the Normal, Binomial, or Poisson distribution. It entails that the outcome follows some family of distributions of similar form with unknown parameters. In survival data modeling, these models are used by specifying the shape of the base line hazard function and covariate effects on hazard function in advance. Examples of distributions commonly used for survival time data are the Weibull model, the log-logistic model, lognormal distribution, the exponential model, the generalized gamma.

The parametric likelihood facilitates in describing the right, left, and interval-censored data. Parametric survival models distribution is basically quantified on time because the approach has a fixed number of parameters.¹³ Hence, survival analysis is used to analyze the data in which the time until the event of interest occurred. It is only when the value of the parameters is recognized that the exact distribution is fully specified. The response is systematically referred to as a failure time, survival time, or event time.

Non-parametric model: The models that do not follow the assumption of any probability distribution properties are called non-parametric models. In survival data, the hazard function is estimated base on empirical facts showing change over time. The non-parametric model used widely is the Kaplan Meier Survival analysis.¹⁴ The Kaplan Meier or Life table (actuarial estimator) considered only one group of patient for estimate of the survival function from the study population. The method is used for calculating the graph of survival probabilities as a function of time. It also applied in estimating the median survival time. The survival curve is distinguished by the probability of surviving in a given length of time and at the same time, it consider time in small intervals. The successive probability will be multiplied by the earlier computed probabilities in getting the final estimate using application law of multiplication of probability to compute the cumulative probability. In every time interval the survival probabilities is computed by the number of subjects surviving divided by the number of patients at risk. The total probability of survival till that time interval is calculated by multiplying all the probabilities of survival at all-time intervals preceding that time. Kaplan Meier curve and the log rank test go along with categorical predictor of variable such as placebo versus drug.¹⁵

Kaplan Meier capitalizes on the event rather than defining an interval establish on time. The interval in Kaplan Meier is defined base on the occurrences of death or termination. Each and every termination marks the end of one interval and the beginning of the subsequent interval. Kaplan Meier uses descriptive procedure for studying the distribution of time to event variable. The method involves the comparison of distribution by level of factor variable and stratification of variables. The assumption for the probability in the event of interest in Kaplan Meier should depend only on time after the initial event without covariate effect. The time variable must be a continuous variable while the status variable should be categorical or continuous in nature.¹⁶ The factor variable should be categorical

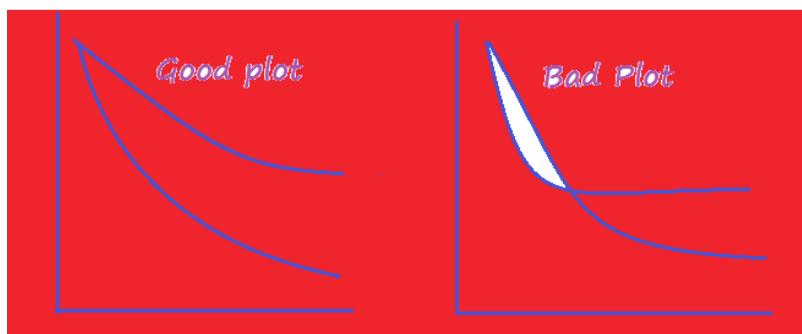
which represent a causal effect (for example treatment types) and then also the stratification variables have to be categorical variable.

Life table analysis: Life table is a descriptive procedure for examining the distribution of time to event variable. It also used to compare the distribution by level factor variable. The main aim is to subdivide the period of observation into smaller time interval, and then the probability from each of the interval will be estimated. The time variable must be continuous. The status variable should be binary or categorical variable which represent the event of interest and the factor variable should be categorical. The life table is very useful for analyzing one group or for comparison few groups defined by level of single categorical factors.

Semi-parametric approach: Under the semi-parametric approach, there is no assumption about the shape of the hazard function like the non-parametric modeling. Example of this model is Cox regression.¹⁷ Survival Inference is made by a specific mathematical model of survival that is established by cox estimates of hazard probabilities for the whole sample. The main reason why the Cox model is widely accepted is because it does not rely on distributional assumptions

for the outcome. Even though the regression parameters (β 's) are known, the distribution of the result remains unfamiliar. The cox proportional hazard model (Cox-regression model) operates by the use of explanatory variables. It has a better flexibility compared to parametric model, especially when it does not require any direct estimation of the baseline hazard function (No underlying probabilities assumption).

The status variable and the dependent variable in cox regression are binary in nature. The time variable computes the event time and it could be continuous or discrete in nature. However the covariate independent or predictor variable can be categorical or continuous variable. Specific variable may perhaps have different values at different periods of time but are not systematically related to time. In this situation, there is need for defining a segment time dependent covariate which can be done by using logical expression. The survival plot for each group using a cox regression approach must be very rightly separated without crossing individual as shown in the graph below.



Data presentation (Table 1)

Table 1 Data presentation

Patient victims on tumor treatment								
S/No.	Treatment	S/time	Status	Age	S/hemoglobin	S/tumor	G/index	
1	1	65	0	67	13.4	34	8	
2	2	61	0	60	14.6	4	10	
3	2	60	0	77	15.6	3	8	
4	1	58	0	64	16.2	6	9	
5	2	51	0	65	14.1	21	9	
6	1	51	0	61	13.5	8	8	
7	1	14	1	73	12.4	18	11	
8	1	43	0	60	13.6	7	9	
9	2	16	0	73	13.8	8	9	
10	1	52	0	73	11.7	5	9	
11	1	59	0	77	12.0	7	10	
12	2	55	0	74	14.3	7	10	
13	2	68	0	71	14.5	19	9	

Table Continued

Patient victims on tumor treatment								
S/No.	Treatment	S/time	Status	Age	S/hemoglobin	S/tumor	G/index	
14	2	51	0	65	14.4	10	9	
15	1	2	0	76	10.7	8	9	
16	1	67	0	70	14.7	7	9	
17	2	66	0	70	16.0	8	9	
18	2	66	0	70	14.5	15	11	
19	2	28	0	75	13.7	19	10	
20	2	50	1	68	12.0	20	11	
21	1	69	1	60	16.1	26	9	
22	1	67	0	71	15.6	8	8	
23	2	65	0	51	11.8	2	6	
24	1	24	0	71	13.7	10	9	
25	2	45	0	72	11.0	4	8	
26	2	64	0	74	14.2	4	6	
27	1	61	0	75	13.7	10	12	
28	1	26	1	72	15.3	37	11	
29	1	42	1	57	13.9	24	12	
30	2	57	0	72	14.6	8	10	
31	2	70	0	72	13.8	3	9	
32	2	5	0	74	15.1	3	9	
33	2	54	0	51	15.8	7	8	
34	1	36	1	72	16.4	4	9	
35	2	70	0	71	13.6	2	10	
36	2	67	0	73	13.8	7	8	
37	1	23	0	68	12.5	2	8	
38	1	62	0	63	13.2	3	8	

Source David collett 2003 modeling survival analysis of data in medical research second edition

Result presentation (Table 2)

Table 2 Descriptive characteristics of patient victims undergoing tumor treatment

Data of respondent		Frequency	Percentage	Cumulative frequency
Treatment	Placebo	18	47.4	47.4
	Treatment	20	52.6	100.0
	Total	38	100.0	

Table Continued

Data of respondent

		Frequency	Percentage	Cumulative frequency
Event	Alive	32	84.2	84.2
	Death	6	15.8	100.0
	Total	38	100.0	
Age category	51-60	6	15.8	15.8
	61-70	11	28.9	44.7
Age category	71-80	21	55.3	100.0
	Total	38	100.0	

The above table showed that 18 (47.4%) of the patients belong to placebo cohort while 18 (47.4%) belong to the treatment cohort. Status of the patient revealed that 32 (84.2%) were alive while 6 (15.8%) died in the course of treatment. Patients between the ages 71-80 have the highest percentage of 55.3% followed by those in the between the ages of 61-70 years with 28.9% and 15.8% for those between ages 51 - 60 category respectively.

Table 3 Case processing summary

Treatment	Total N	N of Events	Censored	
			N	Percent
Placebo	18	5	13	72.20%
Treatment	20	1	19	95.00%
Overall	38	6	32	84.20%

Table 4 Means and medians for survival time

Treatment	Mean ^a		Median					
	Estimate	Std. error	95% Confidence interval		Estimate	Std. error	95% Confidence interval	
			Lower bound	Upper bound			Lower bound	Upper bound
Placebo	58.84	5.067	48.909	68.772	69	0	.	.
Treatment	68.75	1.21	66.378	71.122
Overall	64.203	2.375	59.549	68.858

a. Estimation is limited to the largest survival time if it is censored.

Table 5 Overall comparisons

	Chi-square	df	Sig.
Log rank (Mantel-cox)	4.421	1	0.035
Breslow (Generalized wilcoxon)	3.416	1	0.065
Tarone-ware	3.659	1	0.056

Test of equality of survival distributions for the different levels of treatment.

The Chi-Square value for the Log Rank (Mantel-Cox) which considers the later difference in the factor group is 4.421 and the p-value is 0.035. Since the p-value (0.035) is lesser to the alpha value of 0.05, it indicates that the two groups are statistically significantly different. Hence, it can be concluded that that patients undergoing treatment have a greater survival time compared to placebo administered patients especially at the later course of time of the study.

However, it can be noticed that the Breslow and Tarone-Ware tests have p-values of 0.065 and 0.056 respectively which are all greater than alpha value of $p=0.05$. This indicates that there is no significant difference in the survival time between the treatment and placebo administered patients. The inferences that can be made from this is that the Breslow test concluded that there is no significant difference

Table 6 Case processing summary

		N	Percent
Cases available in analysis	Event ^a	6	15.80%
	Censored	30	78.90%
	Total	36	94.70%
Cases dropped	Cases with missing values	0	0.00%
	Cases with negative time	0	0.00%
	Censored cases before the earliest event in a stratum	2	5.30%
Total	Total	2	5.30%
	Total	38	100.00%

a. Dependent variable: survival time

Hazard function interpretation: The Hazard function is given as EXP (B) in the output analysis and interpreted as follows for each covariate variables. For the treatment covariate, the hazard value of 3.261 shows that the risk of death in placebo cohort patients is 3.261 higher than the treatment patients corresponding to a lower survival time for the placebo patients and a higher survival time for the treatment cohort. For the age, each additional unit of year increases the risk of hazard by 1.014. For the serum haemoglobin, each unit increment reduces the hazard risk by 0.978. For the size of tumor, each unit increment in the size of the tumor increases the hazard risk by 1.099. For the Gleason index, each unit increment in the index increases the chance of the hazard occurrence by 2.061.

P-value with alpha value: Considering the table result, the p value of the treatment covariate ($p=0.329$) is greater than the alpha value

in the survival time of the patients subjected to treatment and those administered placebo at the early time course of the study. While Tarone-Ware test concluded that there is no significantly difference in the survival time of the patients subjected to treatment and those administered placebo at around the middle time course of the study (Figure 1).

The survivor curve above shows that patients in the placebo cohort experience death (event of interest) more quickly in the course of time as indicated by the quick successive drops of the placebo line graph (Table 6).

The overall score of the Omnibus Test shows that our five covariates variables contributed significantly to explain the variability in the hazards of patients undergoing tumor treatment since p-value of 0.01 is less than the alpha value of 0.05 (Table 7&8).

of 0.05. This entails that the hazard occurrence between placebo and treatment group is not statistically significant. That is, there is no evidence of a greater risk of death following the tumor diseases either in the treatment cohort or placebo cohort. The p-value of the age ($p=0.541$) is greater than the alpha value of 0.05. This entails that age has no significant effect on hazard occurrence in the patients. The p-value of the serum haemoglobin ($p=0.961$) is greater than the alpha value of 0.05. This entails that serum haemoglobin has no significant effect on hazard occurrence in the patients. The p-value of the size of the tumor ($p=0.071$) is greater than the alpha value of 0.05. This entails that the size of tumor has no significant effect on hazard occurrence in the patients. However, the p-value of Gleason index ($p=0.039$) is lesser than the alpha value of 0.05. This entails that Gleason index has a significant effect on hazard occurrence in the patients (Table 9).

Table 7 Omnibus Tests of Model Coefficients^a

-2 Log likelihood	Overall (score)			Change from previous step			Change from previous block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
22.173	14.992	5	0.01	14.176	5	0.015	14.176	5	0.015

a. Beginning block number 1. Method = enter

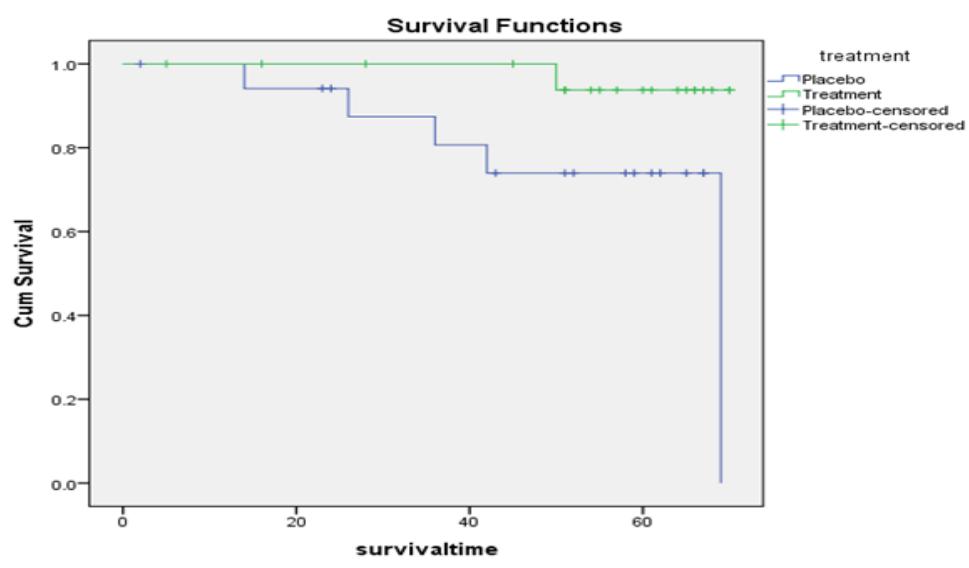
Table 8 Variables in the equation

	B	SE	Wald	df	Sig.	Exp(B)
Treatment	1.182	1.21	0.954	1	0.329	3.261
Age	0.044	0.072	0.373	1	0.541	1.045
Serum hemoglobin	-0.022	0.453	0.002	1	0.961	0.978
Size of tumor	0.094	0.052	3.254	1	0.071	1.099
Gleason index	0.723	0.35	4.273	1	0.039	2.061

Table 9 Covariate means and pattern values

	Mean	Pattern	
		1	2
Treatment	0.472	1	0
Age	68.278	68.278	68.278
Serum hemoglobin	14	14	14
Size of tumor	10.75	10.75	10.75
Gleason index	9.139	9.139	9.139

From the above table the mean shows as follows; treatment (0.472), age (68.278), serum hemoglobin (14.000), size of tumor (10.750) Gleason index (9.139) and age has (2.361) (Figures 2&3).

**Figure 1** Survival functions plot.

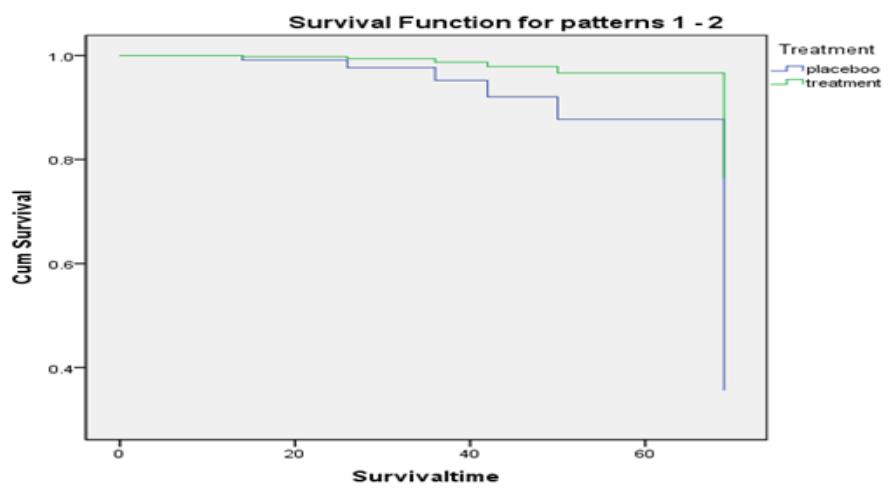


Figure 2 Survival function for treatment and placebo patient groups.

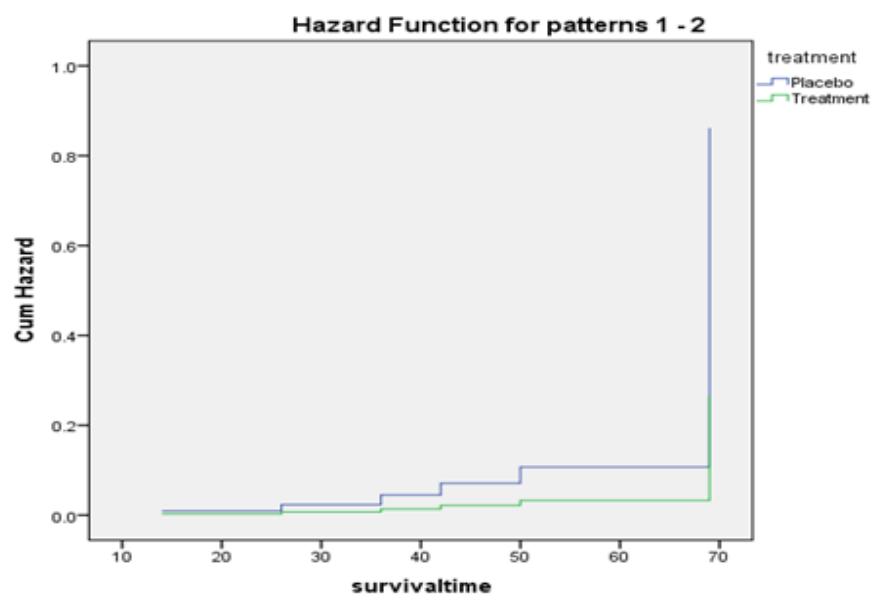


Figure 3 Hazard function for treatment and placebo patient groups.

The graph indicates that treatment cohort has lower tendency to risk of hazard compared to the placebo patient group. However, the statistic test shows the difference is not significant.

Conclusion

The Kaplan-Meier result output shows that the estimate mean survival time for the placebo and treatment cohort patients are 58.840 and 68.750 respectively. It was found that the Log Rank (Mantel-Cox) p – value of 0.035 is lesser than the alpha value (0.05) resulting into a significant difference in survival time between the two factors level. It can therefore be inferred that at the later course of time in the treatment programme, there is significant evidence that the survival time for the treatment cohort is higher than that of the placebo cohort as indicated by their respective mean values. However, the Breslow and Tarone-ware tests with p-values 0.065 and 0.056 respectively are higher than alpha value of 0.05 which indicates that there is no significant

difference in the survival time for the treatment and placebo group cohorts especially at the early and mid- period of the treatment plan.

From Cox Regression analysis report, the p-value of the treatment covariate is 0.329 which is greater than the alpha value of 0.05 shows that the risk of hazard is not statistically significant for the treatment and placebo patient cohorts. However, the hazard function indicates that placebo patients are three times likely to experience death compared to the treatment patients' cohort. The age, serum haemoglobin and size of tumor covariates are not statistically significant to influence the risk of the occurrence of the hazard while the Gleason index indicates a statistically significant relationship that could influence the occurrence of the hazard in the patients.

Acknowledgements

None.

Conflicts of interest

None.

References

1. Wang P, Li Y, Reddy AACK. Machine Learning for Survival Analysis: A Survey. XXX, Vol. X, No. X, Article 1. 2017.
2. Le CT. Introductory Biostatistics. John Wiley & Sons Inc Hoboken New Jersey. 2003.
3. van belle G, Fisher LD, Heagerty PJ, et al. Biostatistics: A Methodology for the Health Sciences. 2nd ed. USA: John Wiley & Sons Inc; 2004.
4. Prinjha S, Gupta N, Verma R. Censoring in Clinical Trials: Review of Survival Analysis Techniques. *Indian J Community Med*. 2010;35(2):217–221.
5. Jager KJ, van Dijk PC, Zoccali C, et al. The analysis of survival data: the Kaplan–Meier method. *Kidney International*. 2008;74(5):560–565.
6. Collett D. Text in Statistics Science Modelling Survival Analysis Data In Medical Research. 2nd ed. 2003.
7. Simona Despa. What is Survival Analysis?. A Stat News # 78 article presentation by the Cornell Statistical Consulting Unit, Cornell University, UK.
8. Clark TG, Bradburn MJ, Love SB, et al. Survival Analysis Part I: Basic concepts and first analyses. *Br J Cancer*. 2003;89(2):232–238.
9. Rondeau V. Statistical models for recurrent events and death: Application to cancer events. *Mathematical and Computer Modelling*. 2010;52(7–8):949–955.
10. Kleinbaum DG, Klein M. Introduction to Survival Analysis. In: Survival Analysis. Statistics for Biology and Health. New York, NY: Springer; 2012.
11. Held L, D Sabanes Bove. Applied Statistical Inference. Springer-Verlag Berlin Heidelberg. 2014.
12. Sullivan Lisa Survival Analysis.
13. Sparling YH, Younes N, Lachin JM, et al. Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*. 2006;7(4):599–614.
14. George B, Seals S, Aban I. Survival analysis and regression models. *J Nucl Cardiol*. 2014;21(4):686–694.
15. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan–Meier estimate. *Int J Ayurveda Res*. 2010;1(4):274–278.
16. Neels K. Event-history analysis: discrete- & continuous-time methods. Society for Longitudinal & Life-course Studies Summer School, University of Amsterdam. 2014.
17. Christiana K. Survival Analysis. *Diagnostic Histopathology* 2016;22(7):263–270.