

## Survival Analysis of Prostate Cancer In Ilorin, Kwara State

\*O.S. Balogun, \*\*M.R. Role and \*\*\*O.O. Dawodu

\*Department of Statistics and Operations Research, Modibbo Adama University of Technology, P.M.B. 2076, Yola, Adamawa State, Nigeria.

\*\*Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria.

\*\*\* Department of Physical and Computer Science(Statistics Unit), McPherson University, Seriki Sotayo, Ogun State.

**Abstract:** -This research work focuses on the analysis of survival times of prostate cancer patients. The data used are from clinical studies on prostate cancer in-patients of the University of Ilorin Teaching Hospital. The aim of the project is to obtain and compare the survival function, median survival time, hazard rates and to fit the semi-parametric model to the survival times.

The Kaplan-Meier Product Limit method was used to obtain the estimates of the median survival time and the survival function, while the test of equality of survival curves was carried out using the log-rank test.

The result from the analysis shows that the surgery and chemotherapy treatment group have median survival times of 19 and 40days respectively with corresponding hazard rates of 0.019 and 0.016 death per day. Result also shows that the difference in survivals for patients that underwent surgery and chemotherapy is not significant with p-value of 0.74, while the Cox's regression model reports the hazards ratio (HR) of 0.84 indicating that the risk of death in the chemotherapy group is less than that of surgery group.

**Keywords:** - Censoring, Survival function, Hazard function.

### I. INTRODUCTION

Cancer develops almost in any organ or tissue of the body, but certain types of cancer are more life-threatening than others. Studies have shown that cancer ranks as the second leading cause of death, exceeded only by heart disease. Each year, about 1.7 million Americans and more than 150,000 Canadians are diagnosed with cancer, and more than half a million Americans and about 70,000 Canadians die of the disease. For reasons not well understood, cancer rates vary by gender, race, and geographic region.

Although people of all ages develop cancer, most types of cancer are more common in people over the age of 50. Cancer usually develops gradually over many years, as a result of a complex mix of environmental, nutritional, behavioral, and hereditary factors. Scientists do not completely understand the causes of cancer, but they know that certain lifestyle choices can reduce the risk of developing many types of cancer; not smoking, eating a healthy diet, and exercising moderately for at least 30 minutes each day can lower the likelihood of developing cancer.

Prostate cancer is one of the cancer diseases in which malignant (cancer) cells form in the tissues of the prostate. The prostate is a gland in the male reproductive system located just below the bladder and in front of the rectum. It is about the size of a walnut and surrounds the urethra (the tube that empties urine from the bladder). Prostate gland produces fluid that is one of the components of semen. Prostate cancer is the most common non-skin malignancy in men and is responsible for more deaths than any other cancer, except for lung cancer. However, microscopic evidence of Prostate cancer is disease common among men.

Since cancer ranks as the second leading cause of death in the world, survival analysis techniques can be used to measure the risk, hazards, and average survival time for cancer patients.

Survival analysis is the collection of statistical procedures for data analysis for which the outcome variable of interest is the time until an event occurs. The time could be in years, months, weeks, days etc. from the beginning of the follow-up of an individual until an event occurs. The event could be death, disease incidence, relapse from remission, recovery or any designated experience of interest.

Survival data include observations on survival times, response to a given treatment and patient characteristics related to response (medicine), lifetime of electronic devices (engineering), components or systems, felon's time to parole (criminology), duration of first marriage (sociology), length of newspaper or magazine subscription (marketing), workmen's compensation claims (insurance) and their various influencing risk or prognostic factors.

Survival analyses are used to investigate the time between entry into a study and the subsequent occurrence of an event. Although survival analyses were designed to measure differences between times to death in study groups, they are frequently used for time to other events including discharge from hospital;

disease onset; disease relapse or treatment failure; or cessation of an activity such as breastfeeding or use of contraception.

With data relating to time, a number of problems occur. The time to an event is rarely normally distributed and follow-up times for patients enrolled in cohort studies vary, especially when it is impractical to wait until the event has occurred in all patients. In addition, patients who leave the study early or who have had less opportunity for the event to occur need to be taken into account. Survival analyses circumvent these problems by taking advantage of the longitudinal nature of the data to compare event rates over the study period and not at an arbitrary time point.

Survival analyses are ideal for analyzing event data from prospective cohort studies and from randomized controlled trials in which patients are enrolled in the study over long time periods. The advantages of using survival analyses rather than logistic regression for measuring the risk of the event occurring are that the time to the event is used in the analysis and that the different length of follow-up for each patient is taken into account. This is important because a patient in one group who has been enrolled for only 12 months does not have an equal chance for the event to occur as a patient in another group who has been enrolled for 24 months. Survival analyses also have an advantage over regression in that the event rate over time does not have to be constant.

This study is going to be looking into analysis of survival time of prostate cancer patients from the point of admission to the time of discharge or time of death.

## **II. CENSORED AND UNCENSORED OBSERVATION**

Censoring occurs when we have some information about individual survival time, but we don't know the exact time. It is also the survival time of patients or animals (subjects) that dropped from the study that is; the survival times of subjects that are lost or die accidentally (unconnected with the event under study). If a subject survives longer than the period of study; such observation is called a censored observation. Censoring may occur in one of the following forms:

- Loss to follow-up (the patient may decide to move elsewhere)
- Dropout (a therapy may have bad effect such that it is necessary to discontinue treatment)
- Termination of the study
- Death due to a cause not under investigation (e.g., suicide).

Uncensored observation occurs when a subject or patient doesn't survive beyond the study period or occurrence of event during the study period.

Symbolically, let  $T$  the survival times for  $n$  patients, the survival times are  $t_{(1)}, t_{(2)}, \dots, t_{(n)}$  such that  $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(n)}$ .

Suppose the survival times for cancer patients are 2, 2+, 6, 10, 15+, 22 and 30+ (in days). The plus (+) sign indicates a censored observation. That is the patients survived for at least the period indicated before the + sign.

The survival times for the uncensored observations are times from the start of the experiment or study to their death. That is, the exact length of survival time of the individual is known. Uncensored observations are often referred to as exact observations.

Consider the survival times for the cancer patients given above, the times without plus (+) sign indicates the uncensored observations. For the purpose of this work, uncensored observations have been coded status 1 while the censored observations were coded status 0.

When a subject withdraw against medical advice (DAMA), or abscond treatment, or die accidentally or on referral services during the study period, the observation is said to be lost to follow-up and therefore is categorized as censored observation.

The research aimed at achieving the following objectives: determining the survival function of the disease, determining and comparing the survival function of the disease based on the type of treatment received by each patient, estimating the mean and median survival time of the treatment group using Kaplan-Meier product limit estimates, fitting Cox's regression model for the data and determining the Hazard Ratio for the treatment groups and the age and fitting Weibull distribution model for the remission data.

The data is a secondary data on prostate cancer in-patients which was collected from the University of Ilorin from 2007 to 2009, this period is referred to as the study period. The information available includes age, mode of treatment given to each patient, duration of treatment and outcome (which is often referred to as status). According to David Machin et al (2005), the term 'survival' is used because an early use of the associated statistical techniques arose from the insurance industry, which was developing methods of costing insurance premiums. The industry needed to know the risk or average survival time associated with a particular type of client. This 'risk' was based on that of a large group of individuals with a particular age, gender and possibly other characteristics; the individual was then given the risk for his or her group for the calculation of their insurance premium.

Historically, much of survival analysis has been developed and applied in relation to cancer clinical trials in which the survival time is often measured from the date of randomization or commencement of therapy until death. The seminar papers by Peto, Pike, Armitage et al. (1976, 1977) published in the British Journal of Cancer describing the design; conduct and analysis of cancer trials provide a landmark in the development and use of survival methods.

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. The data which are called survival data are times that measure the time to response, failure, death, relapse or development of a given disease.

In the course of study, subjects enter at different times, some are lost to follow-up, and some experience the event while others do not, at the end of the study. The exact survival time of the subjects who do not experience the event of interest and those lost to follow-up are not known. This is a special feature of survival data analysis and is known as censoring, the observations are known as censored observations or censored survival time.

### III. SOME TERMINOLOGY AND NOTATIONS

**Survival function:** also known as a survivor function or reliability function is a property of any random variable that maps a set of events, usually associated with mortality or failure of some system, onto time. It captures the probability that the system will survive beyond a specified time.

**Covariate:** a covariate is a variable that is possibly predictive of the outcome under study. A covariate may be of direct interest or it may be a confounding or interacting variable.

**Hazard ratio:** the hazard ratio (HR) in survival analysis is the effect of an explanatory variable on the hazard or risk of an event. For a less technical definition than is provided here, consider hazard ratio to be an estimate of relative risk and see the explanation on that page.

$$HR = e^{\beta} \text{ given that } X = 0,1$$

**Survival rate:** survival rate is a part of survival analysis indicating the percentage of people in a study or treatment group who are alive for a given period of time after diagnosis.

**T** denotes the random variable for a person's survival time. Since **T** denotes time, its possible values include all nonnegative numbers; that is, **T** can be any number equal to or greater than zero. **T = Survival time (T ≥ 0)**.

**t** denotes any specific value of interest for the random variable **T**.

**Greek letter delta (δ)** denote a (0,1) random variable indicating either failure or censorship. That is,  $\delta = 1$  for failure if the event occurs during the study period, or  $\delta = 0$  if the survival time is censored by the end of the study period.

$\delta = (0, 1)$  random variable

$$= \begin{cases} 1 & \text{if failure} \\ 0 & \text{if censored} \end{cases}$$

Note that if a person does not fail that is, does not get the event during the study period, censorship is the only remaining possibility for that person's survival time. That is,  $\delta = 0$  if and only if one of the following happens:

- a person survives until the study ends
- a person is lost to follow-up
- a person withdraws during the study period.

#### Survival Distributions

**1. Survivorship 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**. The survivor function is fundamental to a survival analysis, because obtaining survival probabilities for different values of **t** provides crucial summary information from survival data.

$$S(t) = P(\text{a subject survives longer than } t)$$

$$= P(T > t)$$

$$S(t) = 1 - F(t)$$

Where  $F(t)$  is defined as the cumulative distribution function of **T** i.e.  $P(T \leq t)$  the probability that a subject fails before time **t**.

$$S(t) = \frac{\text{Number of subjects surviving longer than } t}{\text{total number of subjects}}$$

The general formula is given as

$$S(t) = \exp\left(-\int_0^t h(u)du\right), \quad t \geq 0$$

#### Theoretical properties of survivorship function

According to David G. Kleinbaum, Mitchel Klein (2005); all survivor functions have the following characteristics:

- They are non-increasing; that is, they head downward as **t** increases

- At time  $t = 0$ ,  $S(t=0) = 1$ ; that is, at the start of the study, since no one has gotten the event yet, the probability of surviving past time 0 is one
- At time  $t \rightarrow \infty$ ,  $S(t \rightarrow \infty) = 0$ ; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually fall to zero.

The survivorship function is used to find the 50<sup>th</sup> percentile (median survival time).

**2. Hazard Function  $h(t)$**  is defined as the probability of failure during a very small time interval, assuming that the subject has survived to the beginning of the interval. The hazard function  $h(t)$  gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time  $t$ . It is a measure of proneness to failure.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{p(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

Where  $p(t \leq T < t + \Delta t | T \geq t)$   
 =  $p(\text{individual fails in the interval } [t, t + \Delta t] | \text{survival up to time } t)$  or  
 =  $p(\text{a subject of age } t \text{ fails in the interval } [t, t + \Delta t])$

Therefore,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{p(\text{a subject of age } t \text{ fails in the interval } (t, t + \Delta t))}{\Delta t}$$

$$h(t) = \frac{\text{number of subjects failing per unit time in the interval}}{\text{number of subjects not failing at time } t}$$

The general formula is given as

$$h(t) = - \left[ \frac{dS(t)/dt}{S(t)} \right]$$

The hazard function  $h(t)$  has the following characteristics:

- it is always nonnegative, that is, equal to or greater than zero
- it has no upper bound.

**3. Probability Density Function  $f(t)$**  is defined as the limit of the probability that a subject fails in the interval  $t$  to  $t + \Delta t$  per unit width  $\Delta t$

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{p(\text{a subject failing in the interval } (t, t + \Delta t))}{\Delta t}$$

$$f(t) = \frac{\text{number of subjects failing in the interval beginning at } t}{(\text{total number of subject})(\text{interval width})}$$

$$f(t) \geq 0 \text{ for all } t \geq 0$$

$$f(t) = 0 \text{ for } t < 0$$

#### IV. RELATIONSHIP BETWEEN THE SURVIVAL DISTRIBUTIONS

##### Survivorship function

$$S(t) = p[T > t]$$

From the definition of cumulative distribution  $F(t)$  of  $T$

$$S(t) = 1 - p[T \leq t]$$

$$S(t) = 1 - F(t)$$

Where  $F(t)$  is the probability that an individual fails on or before time  $t$ .

##### Probability Density Function [ $f(t)$ ]

$f(t)$  can be obtained from the derivative  $S(t)$

$$S^1(t) = -f(t)$$

$$f(t) = -S^1(t)$$

##### Hazard Function [ $h(t)$ ]

Hazard function is the ratio of the p.d.f. to the survivorship

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

$$= \frac{f(t)}{1 - F(t)}$$

$$= \frac{-S'(t)}{S(t)}$$

$$= \frac{d}{dt} (\ln[S(t)])$$

Therefore,

$$f(t) = h(t).S(t)$$

$$= h(t)[1 - F(t)]$$

### **Cox's Proportional Hazards Model**

Proportional hazards models are a class of survival models in statistics. Survival models relate the time that passes before some event occurs to one or more covariates that may be associated with that quantity. In a proportional hazards model, the unique effect of a unit increase in a covariate is multiplicative with respect to the hazard rate.

Cox's PH model

$$h(t, x) = h_0(t)e^{\sum_{i=1}^p \beta_i X_i},$$

Where  $h_0(t)$  is called the base line hazard

$X = x_1, x_2, \dots, x_p$  are explanatory/predictor variable

## **V. WEIBULL DISTRIBUTION MODEL**

Weibull distribution is a form of parametric model which does not assume a constant hazard rate and is characterized by two parameters,  $\lambda$  the value which determines the scaling) and  $\alpha$  (the value which determines the shape). When  $p=1$  the hazard rate remains constant as time increases, when  $p>1$  the hazard rate increases and  $p<1$  it decreases as time increases David G. Kleinbaum et al (2005). Thus, the Weibull distribution can be used to model the survival distribution of a population with increasing, decreasing or constant risk.

The probability density function is

$$f(t) = \lambda p t^{p-1} \exp(-\lambda t^p) \quad t \geq 0, \lambda > 0$$

the cumulative distribution function is

$$F(t) = 1 - \exp(-\lambda t^p)$$

The survivorship function is

$$S(t) = \exp(-\lambda t^p)$$

The hazard function is

$$h(t) = \lambda p t^{p-1}$$

For easy computation and accuracy, the computer packages used include STATA 10 and SPSS 15 in analyzing the survival data.

## **2. Methodology**

### **Methods of Estimation**

The data was grouped into two categories according to the treatment received by the patients which includes the surgery and chemotherapy groups and were coded 0 and 1 respectively. The indicator ( $\delta$ ) was also coded (0,1), where the censored observation is 0 and the event is 1.

#### **1. Estimation of Survival Function**

The Kaplan-Meier product limit estimate was used to estimate the median survival time and the survival function. The procedures are as follows:

Let

$P_1$  = Probability of surviving at least one day after admission

$P_2$  = Conditional probability of surviving the second day after having survived the first day.

.

.

.

$P_t$  = Conditional probability of surviving the (t)th day after having survived day (t-1).

Therefore the overall probability of surviving t days after admission,  $S(t)$ , is then given by the products of the following probabilities. Thus,

$$S(t) = P_1 \times P_2 \times \dots \times P_t$$

But

$$P_t = \frac{(n_t - d_t)}{n_t}$$

or

$$P_t = 1 - \frac{d_t}{n_t}$$

$$S(t) = \left(1 - \frac{d_1}{n_1}\right) \times \left(1 - \frac{d_2}{n_2}\right) \times \dots \times \left(1 - \frac{d_t}{n_t}\right)$$

$$S(t) = \prod_t \left(1 - \frac{d_t}{n_t}\right)$$

Since  $S(t)$  is the cumulative survival function, the median survival time can be estimated as follows:

Let  $m$  = median survival time,

$m$  = 50 percentile

$S(m) = 0.5$

Therefore, when  $S(t) = 0.5$ , the corresponding survival time ( $t$ ) is the median survival time.

Where,

$t$  denotes the start of a short time interval ending at  $(t + 1)$ .

$n_t$  is the number of patients alive at the start of the interval.

$d_t$  is the number of patient dying in the short time interval just after  $t$ .

## 2. Method of Comparing Two Survival Curves

Log-rank test also referred to as the Mantel-Cox test; Peto Richard, et al (1972): is the most widely used method for comparing two or more survival curves. Its hazard ratio estimated as well. The log-rank test is a large-sample chi-square test that uses as its test criterion a statistic that provides an overall comparison of the Kaplan-Meier curves being compared. This (log-rank) statistic, like many other statistics used in other kinds of chi-square tests, makes use of observed versus expected cell counts over categories of outcomes. The categories for the log-rank statistic are defined by each of the ordered failure times for the entire set of data being analyzed.

Log-rank  $\chi^2$  can be calculated as follows:

$$\chi^2 = \frac{(O_i - E_i)^2}{\text{var}(O_i - E_i)}$$

Where,

$$O_i - E_i = \sum_j^n (m_{ij} - e_{ij}), \quad i = 1, 2$$

$$\text{var}(O_i - E_i) = \sum_j^n \left( \frac{n_{1j}n_{2j}(m_{1j} + m_{2j})(n_{1j} + n_{2j} - m_{1j} - m_{2j})}{(n_{1j} + n_{2j})^2(n_{1j} + n_{2j} - 1)} \right), \quad i = 1, 2$$

expected cell counts:

$$e_{1j} = \left( \frac{n_{1j}}{n_{1j} + n_{2j}} \right) \times (m_{1j} + m_{2j})$$

$$e_{2j} = \left( \frac{n_{2j}}{n_{1j} + n_{2j}} \right) \times (m_{1j} + m_{2j})$$

$n_{ij}$  is the number of subject in the risk set at time  $t_j$

$m_{ij}$  is the number of subjects failing at time  $t_j$

$e_{ij}$  is the expected number at time  $j$

$O_i$  is the sum of observed subject at risk in group  $i$

$E_i$  is the sum of the expected number of subject failing at time  $j$  in group  $i$

### Hypothesis for comparing two survival curves

$H_0$ : the risk of getting the event is the same for both treatment groups against

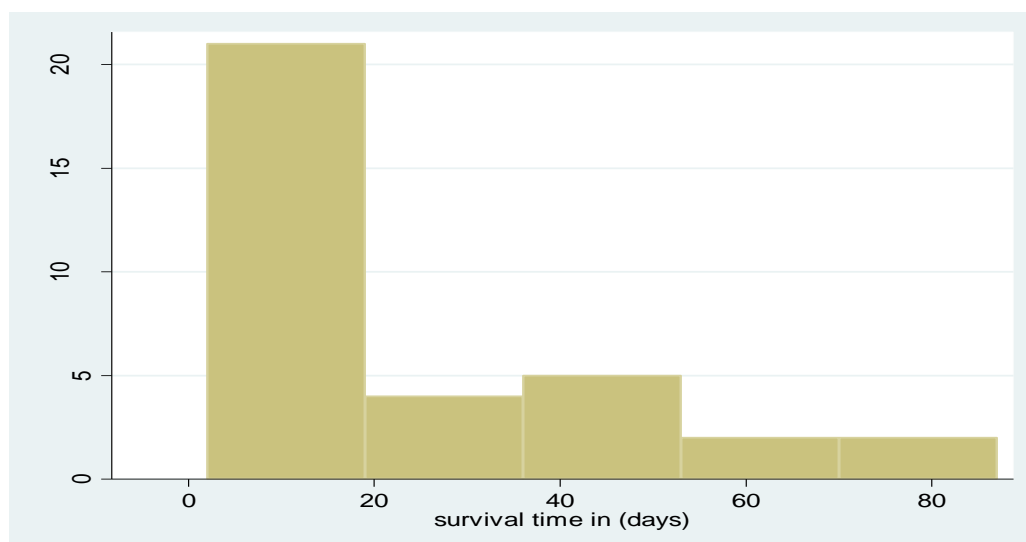
$H_1$ : The risk of getting the event is not the same for both treatment groups.

Reject  $H_0$ , if p-value is less than  $\alpha$ .

## 3. Data Presentation and Analysis

The following histogram shows the distribution of survival time of the in-patients suffering from prostate cancer.

Figure 3.1



The distribution is skewed to the right, in that the right-hand tail of the distribution is much longer than the left-hand tail. In this situation the mean is not a good summary of the ‘average’ survival time because it is unduly influenced by the extreme observations.

**Table 3.1 Case Processing Summary (prostate cancer)**

mode of treatment	Total N	N of Events	Censored	
			N	Percent
chemotherapy	20	8	12	60.0%
Surgery	14	6	8	57.1%
Overall	34	14	20	58.8%

The above table shows the summary of the groupings of the patients according to the type of treatment received.

**3.1 Estimation of Survivor Function for Prostate Cancer**

Below are the survival function table, median survival table and the Kaplan-Meier survival function.

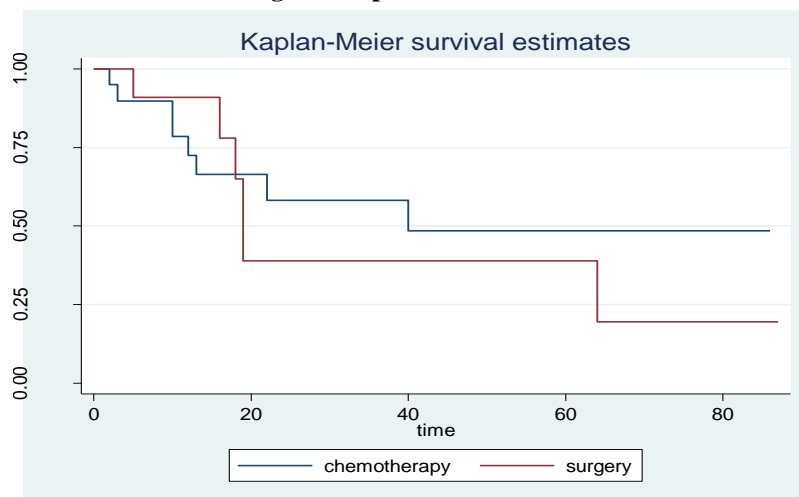
**Table 3.2 Survivor function table**

	Time	Beg. Total	Fail	Net lost	Survivor function	Std. error	[95% interval]	conf.
surgery	2	14	0	2	1.0000	-	-	
	3	12	0	1	1.0000	-	-	
	5	11	1	0	0.9091	0.0867	0.5081, 0.9867	
	6	10	0	1	0.9091	0.0867	0.5081, 0.9867	
	10	9	0	1	0.9091	0.0867	0.5081, 0.9867	
	15	8	0	1	0.9091	0.0867	0.5081, 0.9867	
	16	7	1	0	0.7792	0.1413	0.3544, 0.9418	
	18	6	1	0	0.6494	0.1671	0.2494, 0.8744	
	19	5	2	0	0.3896	0.1741	0.0920, 0.6891	
	40	3	0	1	0.3896	0.1741	0.0920, 0.6891	
	64	2	1	0	0.1948	0.1629	0.0116, 0.5486	
87	1	0	1	0.1948	0.1629	0.0116, 0.5486		
chemotherapy	2	20	1	1	0.9500	0.0487	0.6947, 0.9928	
	3	18	1	0	0.8972	0.0689	0.6475, 0.9733	
	5	17	0	1	0.8972	0.0689	0.6475, 0.9733	
	10	16	2	1	0.7851	0.0956	0.5227, 0.9137	
	12	13	1	0	0.7247	0.1056	0.4576, 0.8758	
	13	12	1	0	0.6643	0.1128	0.3975, 0.8341	
	16	11	0	1	0.6643	0.1128	0.3975, 0.8341	
	17	10	0	2	0.6643	0.1128	0.3975, 0.8341	
	22	8	1	0	0.5813	0.1256	0.3060, 0.7799	
	25	7	0	1	0.5813	0.1256	0.3060, 0.7799	
	40	6	1	0	0.4844	0.1370	0.2107, 0.7136	
	43	5	0	1	0.4844	0.1370	0.2107, 0.7136	

	44	4	0	1	0.4844	0.1370	0.2107, 0.7136
	49	3	0	1	0.4844	0.1370	0.2107, 0.7136
	69	2	0	1	0.4844	0.1370	0.2107, 0.7136
	86	1	0	1	0.4844	0.1370	0.2107, 0.7136

**Table 3.2** is an output of survival analysis from the STATA package which displays the list of survivor function. The table is separated into two treatment groups which include surgery and chemotherapy. From the chemotherapy group, the probability that a patient survives 2days after admission is 0.95 that is  $S(2)=0.95$ , at  $S(22)$  the probability of survival drops to 0.58 which means that 58 percent of patients on admission will be alive after 22days While the probability of surviving more than 40days is 0.48.

**Fig. 3.2 Kaplan-Meier Curves**



**Figure 3.2** describes the survival function of the data in table 4.2 in which the two curves of both treatment intercepts at  $S(t)=0.66$  which makes it difficult to conclude that one treatment is better than the other. But based on the estimated medians, the chemotherapy treatment group has a better median value of 40days while the surgery group has median value of 19days which means that patients receiving the chemotherapy treatment have a better chance of survival.

**Table 3.3 Medians for Survival Time**

Mode of treatment	Median			
	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Chemotherapy	40.000	.	.	.
Surgery	19.000	0.670	17.687	20.313
Overall	40.000	22.150	0.000	83.415

**3.2 Comparison of the Survival Curves**

Hypothesis

$H_0$ : the risk of event is the same for both treatment groups

$H_1$ : not  $H_0$

Significance level=0.05

Decision Rule: Reject  $H_0$  if p-value is less than the significance level

**Table 3.4: Log-rank test to compare the treatment modes**

Treatment modes	Events observed	Events expected	Hazard Ratio
Chemotherapy	8	8.60	0.8372
Surgery	6	5.40	1.1944
Total	14	14.00	

Chi square (1) = 0.11

Pr>Chi square = 0.7376

Decision: Do not reject the null hypothesis



Conclusion: from the table the p-value is greater than the significance level, indicating that there is equal risk of death in both groups of treatment.

### 3.3 Parametric Regression Model

The Weibull regression model gives the distribution for which the failure rate (death) is proportional to survival time which is explained by the shape parameter ( $p$ ).

**Table 3.5: Result of the Weibull regression model**

	<b>Haz. ratio</b>	<b>Std. Err</b>	<b>Z</b>	<b>P &gt;  z </b>	<b>[95% Conf. interval]</b>
<b>Trt</b>	<b>0.8242899</b>	<b>0.4451682</b>	<b>0.36</b>	<b>0.720</b>	<b>0.2860077, 2.375649</b>
<b>Age</b>	<b>1.024563</b>	<b>0.0269481</b>	<b>0.92</b>	<b>0.36</b>	<b>0.9730839, 1.078765</b>
<b>P</b>	<b>0.9697338</b>	<b>0.2017821</b>			<b>0.6449627, 1.458043</b>

#### Interpretation of the model:

- Treatment: the hazard ratio is 0.8243 which indicates that the survival rate of the chemotherapy treatment group is higher than that of the surgery group since  $HR < 1$  while the surgery is the reference category. However, this does not reach conventional statistical significance since the p-value = 0.720 which is greater than the significance level.
- Age: the hazard ratio is greater than one; indicating that the risk of getting the event increases as the ages increases. The test is not statistically significant since the p-value is greater than the significance level.
- From the above model, the estimate of shape parameter  $p$  is 0.97 which is less than one suggesting a decrease in hazard as the survival time increases.

### 3.4 Proportional Hazard Model (Semi-Parametric)

**Table 3.6: Result of the Cox's proportional hazard model**

	<b>Haz. Ratio</b>	<b>Std. Err.</b>	<b>Z</b>	<b>P&gt; Z </b>	<b>[95% Conf. Interval]</b>
<b>Treatment</b>	<b>0.835265</b>	<b>0.4534299</b>	<b>0.33</b>	<b>0.740</b>	<b>0.2882333, 2.420502</b>
<b>Age</b>	<b>1.019876</b>	<b>0.0252778</b>	<b>0.79</b>	<b>0.427</b>	<b>0.9715169, 1.070643</b>

#### Interpretation of the model:

- Treatment: the hazard ratio is 0.835 which suggests a considerable advantage of the chemotherapy over that of the surgery group. However, this does not reach conventional statistical significance since the p-value = 0.740 which is greater than the significance level.
  - Age: the hazard ratio is greater than one; indicating that the risk of getting the event increases as the ages increases. The test is not statistically significant since the p-value is greater than the significance level.
- It can be seen that the hazard ratios in the Weibull model is not so different from that of the Cox's model, so the two models described the data adequately well.

## 4. Summary, Discussions and Conclusion

### 4.1 Summary

Table below summarizes the number of cases for each treatment group, the number of patients that died (uncensored observations), the number of patients that survived (censored observation) and their respective percentages.

**Table 4.1 Summary**

	<b>Surgery group</b>	<b>%</b>	<b>Chemotherapy group</b>	<b>%</b>
<b>Number died</b>	<b>6</b>	<b>42.9</b>	<b>8</b>	<b>40</b>
<b>Number survived</b>	<b>8</b>	<b>57.1</b>	<b>12</b>	<b>60</b>
<b>Total number</b>	<b>14</b>	<b>100</b>	<b>20</b>	<b>100</b>

The summary table above shows that the percentage of people that died in the surgery group is more than those that died in the chemotherapy group.

With the aid of STATA10 package, the Logrank test was applied to test for the equality of survival function of the prostate cancer in-patients receiving different treatments. The set hypothesis (null hypothesis) at 5% level of significance was accepted indicating that the survival functions are the same. On a contrary, the hazards ratio from the fitted models indicated that the risk of getting the event in the surgery group is higher than that of chemotherapy group.

## 4.2 Discussions

From fig. 4.1 the distribution of the survival function was presented in a histogram which indicates that the normal curve is skewed to the right. In this situation the mean is not a good measure of central tendency. This property restricted the study from using the normal regression analysis.

Table 4.2 displays the Kaplan-Meier product limit estimate list of survival function which ranks the survival times according to the treatment received by the patients and their respective survival function.

The survival function of the table 4.2 was described by the Kaplan-Meier curves of figure 4.2 in which the curve of the chemotherapy treatment group crosses that of surgery group at about  $S(t) = 0.66$  makes it difficult to conclude that one treatment is better than the other.

The estimated hazard rate for surgery and chemotherapy groups are 0.019 and 0.016 respectively which means that the patients belonging to chemotherapy group have a lower risk of getting the event and longer survival time compared to that of surgery group. The estimated median survival time for surgery and chemotherapy treatment groups are 19 and 40days respectively is an evidence that the chemotherapy group have a longer survival time.

The median survival time can also be expressed in terms in terms of survival function as follows;  $S(t) = 0.5$  this implies that prostate cancer in-patients receiving chemotherapy treatment has the probability of 0.5 to survive longer than 40days while the in-patients undergoing surgery has the probability of 0.5 to survive longer than 19days.

Table 4.5 displays the result of the Weibull regression model with treatment  $HR = 0.82$ , the surgery group is the reference category, the chemotherapy treatment group can be said to have considerable advantage over the surgery group since the  $HR < 1$ . Also the  $HR$  for age is 1.02. This suggests that the risk of getting the event increases as the age increases.

Table 4.6 is the result from the Cox's regression model (semi parametric model) with treatment  $HR = 0.84$  and  $HR$  (age) = 1.02. This result gives almost the same values of HRs as that of Weibull regression model and the same conclusion can be drawn from both models.

## 4.3 Conclusion

Having certified with the analysis in the previous chapter, the data collected on the treatment modes and survival times of the prostate cancer in-patients of the university of Ilorin teaching hospital, the following conclusion were drawn:

The outcome of the Logrank test revealed that there is no difference between the survival curves of patients who received the chemotherapy treatment and those that undergo the surgical procedure.

Based on the estimated hazard rates of 0.016 and 0.019 death per day and corresponding median survival times of 40 and 19days for patients in chemotherapy and surgery groups respectively; shows that the prostate cancer in-patients treated with chemotherapy survives longer than in-patients that undergo surgery.

Also, the hazards ratio estimated from the Weibull and Cox's proportional hazards model shows that the risk of death is reduced in chemotherapy treatment group compared to that of surgery. Generally, results show that the risk of death increases as the age increases for all treatment groups (since  $HR_{(age)} > 1$ ). The shape parameter  $P = 0.96$  shows that hazards decreases as time increases.

Finally, the result of Cox's proportional hazards model reports almost the same values of Hazards ratio as that of Weibull proportional hazards model. This shows that for this study, both models are the same.

## REFERENCES

- [1] **David .G. Kleinbaum and Mitchel Klein (2005)**. Survival Analysis, A Self-learning Text, 2<sup>nd</sup> Edition, Springer-Verlag, New York.
- [2] **David Machin, Yin Bun Cheung, Mahesh .K.B. Parmar (2005)**. Survival Analysis: A Practical Approach, Second Edition. John Wiley and Sons, Ltd.
- [3] **Peto Richard and Julian Peto (1972)**. Asymptotically Efficient Rank Invariant Test procedures, Journal of the Royal Statistical Society. Series A (General), Vol.135, No 2, 185-207.
- [4] **Peto .R., Pike .M.C., Armitage P., Breslow N.E., Cox .D.R., Howard .S.V., Mantel N., MacPherson**
- [5] **K., Peto J. and Smith P.G. (1976)**. Design and Analysis of randomized clinical trials requiring Prolonged observation of each Patients I: introduction and design. British Journal of Cancer, 34, 585-612(1).
- [6] **Peto .R., Pike .M.C., Armitage P., Breslow N.E., Cox .D.R., Howard .S.V., Mantel N., MacPherson**
- [7] **K., Peto .J. and Smith P.G. (1976)**. Design and Analysis of Randomized Clinical Trials requiring prolonged observation of each Patient II: introduction and design. British Journal of Cancer, 35, 1-39(1).