Applied Mathematics and Computational Intelligence
Volume 13, No. 3, 2024 [143-175]

Survival Analysis on Lung Cancer Patients in Damaturu-Nigeria

Umar Yusuf Madaki[1] and Abubakar Abdullahi Gadaka[2]

[1,2]Department of Mathematics and Statistics, Faculty of Science, Yobe State University, Damaturu, Nigeria

Corresponding author: uymadaki@ysu.edu.ng, uymadaki84@gmail.com

Received: 12 November 2023
Revised: 12 July 2024
Accepted: 23 July 2024

ABSTRACT

Cancer is one of the leading causes of death across the globe, in both men and women accounting for 23% of all cancer deaths in 2019 according to the Centers for Disease Control and Prevention. One of the eccentric problems with lung cancer is that it usually has a poor prognosis. With such a deadly disease, it is crucial to predict the survival likelihood of cancer patients. However, this is not an easy task due to the many factors affecting the disease progression. Survival time has become an essential outcome of clinical trial, which began to emerge among the latter half of the 20th century. A present study was carried out on the survival analysis for patients with lung cancer. The data was obtained from Yobe State Specialist Hospital, Damaturu where each sample was collected from the recipients of the treatment of radical prostatectomy. The Kaplan Meier method was used to obtain and estimate the survival function and median. The log-rank test was used to test the differences in the survival curves. The cox proportional hazard (PH) model provided an effective covariate on the hazard function. As a result of cox PH model, the influence of standard clinical prognostic factors is based on the hazard rate of lung cancer patients. We performed rigorous cross-examination on each feature's relationship and the model for each feature type using data analysis information and survival analysis models. For each feature type, we used one representative survival analysis model from semi-parametric methods (Cox proportional hazards model), one from non-parametric methods (Kaplan-Meier estimator), and one from machine learning approaches (random survival forests). Using the results obtained from these different methods, we identified the best feature types and model combinations to get the top performance for various follow-up periods. The best model is Cox proportional hazards model based on the AIC and log-likelihood functions respectively.

Keywords: AIC, Cox proportional hazards model, Damaturu, Kaplan-Meier estimator, Lung Cancer, Yobe.

1 INTRODUCTION

The word "cancer" refers to a collection of illnesses in which cells exhibit aberrant growth and division patterns. There are over a hundred distinct forms of cancer. The male reproductive system's lung gland is where lung cancer first appears. Its job is to produce the fluid in semen that keeps sperm cells safe and nourished. There are several routes that a cancer cell might spread, including through tissue, the lymphatic system, and the blood (National Cancer Institute), [1]. Early diagnosis and

screening are among the most successful intervention strategies for lung cancer [2]. After lung cancer, lung cancer is the second most prevalent malignant cancer that kills males, and its frequency rises with age. Men who have lung cancer often have long lives since the disease progresses more slowly than other malignancies. Thankfully, half of newly diagnosed instances of lung cancer are in the early stages of the disease and remains restricted to the lung. On the other hand, aggressive lung malignancies can be extremely deadly in a large number of instances [2]. The term "metastasis" refers to the ability of a lung cancer cell to travel to other areas of the body, including lymph nodes and bones. The three main risk factors for lung cancer are family history, age, and race/ethnicity. Age has been determined to be the most significant factor among them, particularly for older men and women over 60 years respectively.

1.1 STATEMENT OF THE PROBLEM

One of the main causes of death is cancer. According to the Centers for Disease Control and Prevention, lung cancer specifically accounts for 23% of all cancer fatalities in 2019 and is the leading cause of cancer death in both men and women. One particular problem with lung cancer is that it usually has a poor prognosis. With such a deadly disease, it is crucial to predict the survival likelihood of cancer patients. However, this is not an easy task due to the many factors affecting the disease progression. In Nigeria, researches have been conducted on lung cancer among the populaces especial the native for monitoring and controlling lung cancer [1]. One of the most effective intervention tools for lung cancer is screening and early diagnosis [2]. However, the lack of knowledge on the disease and the low uptake of routine screening among men, especially those at risk of developing lung cancer make the problem a complex one.

1.2 RESEARCH QUESTIONS

What are the survival probabilities over time for individuals with different risk factors and hazard rates vary over time for individuals with different risk factors and what is the impact of each risk factor on survival outcomes for performing Kaplan-Meier analysis and estimating survival and hazard functions with all risk factors?

How accurate are the estimates of survival functions obtained using the Cox Model in the presence of censored data and do different methods for handling censored data affect the efficiency of survival function estimation and what are the advantages and limitations of using the Cox Model for survival analysis with censored data?

Which method, between Kaplan-Meier and Cox Proportional Hazard models, provides a better fit to the survival data based on AIC and How do the estimated survival functions obtained from Kaplan-Meier and Cox Proportional Hazard models differ in terms of model complexity and predictive accuracy based on AIC, offers the most parsimonious yet accurate representation of survival data among the compared techniques?

1.3 AIM AND OBJECTIVES OF THE STUDY

The aim of this study is to apply the survival analysis methods on lung cancer patients (case study of Yobe state Specialist Hospital, Damaturu)" having the following objectives:

To perform the Kaplan-Meier analysis by estimating survival and hazard functions with all the risks factors.

To examine the efficiency of the methods used to estimate survival functions in the presence of censored data using the Cox Model.

To compare the techniques of the estimated survival functions with the Kaplan-Meier and the Cox Proportional hazard models using Akaike Information Criterion (AIC).

1.4 SCOPE AND LIMITATION AND SIGNIFICANCE OF THE STUDY

This study was limited to Yobe State Specialist Hospital, Damaturu. In order to increase the authenticity of the study, health and occupation status will be used as surrogate for income and residence. The significance of this study is to show the knowledge levels, perception toward lung cancer and uptake of screening for lung cancer among men attending specialist Hospital Damaturu, Yobe state. The results of this study demand crucial health measures targeted at promoting specific knowledge levels on lung cancer and calls for positive behavioral changes towards avoiding risks for the development of lung cancer in men. The study demands the design of new screening strategies for lung cancer across the state, as early screening for lung cancer has been revealed to contribute meaningfully to the management of the disease. It is anticipated that the information generated will also be used by local cancer bodies, the Yobe state cancer control strategy, the Cancer Society of Yobe state, academicians, scientists for developing policies for control and prevention of lung cancer in Yobe state Damaturu-Nigeria.

2 SOME REVIEWS ON SURVIVAL ANALYSIS ON LUNG CANCER STUDY

Lung cancer is a significant health burden worldwide, and its incidence and mortality rates are increasing in Nigeria, reflecting global trends. Survival analysis plays a crucial role in understanding the prognosis, risk factors, and treatment outcomes of lung cancer patients in Nigeria. [3], conducted a retrospective study to assess survival outcomes and prognostic factors among lung cancer patients in Nigeria. They found that the overall survival rates were lower compared to global averages, with advanced stage at diagnosis being a significant predictor of poor prognosis. Additionally, the study identified socioeconomic factors, access to healthcare, and histological subtype as important determinants of survival. A study by [4], investigated the impact of treatment modalities on survival outcomes in lung cancer patients in Nigeria. The study utilized Kaplan-Meier analysis to assess survival probabilities among patients receiving chemotherapy, radiotherapy, or combined therapy. Their findings highlighted the importance of early initiation of treatment and comprehensive multidisciplinary care in improving survival rates among lung cancer patients in a study by [5], where they explored the influence of ethnic and genetic factors on survival outcomes among lung cancer patients of Nigerian descent. Using Cox proportional hazards regression analysis, they observed variations in survival rates among different ethnic groups, suggesting a potential role of genetic predisposition in lung cancer prognosis among Nigerians. [6], conducted a population-based study to investigate socioeconomic disparities in survival outcomes among lung cancer patients in Nigeria. Using Cox regression analysis, they identified income level, education, and urban/rural

residence as significant predictors of survival. The study emphasized the importance of addressing socioeconomic inequalities in access to healthcare and treatment outcomes for lung cancer patients. [7], faced significant challenges that kept them from bringing data together from different studies in order to assess disparities in results of treatment in various institutions. Initially, they were presented with different endpoints from the studies. Following this, they noticed that the different studies showed varying disease severity. Finally, usefulness of the results was limited by the differences in the techniques used to measure patient-focused outcomes. There are several clinical trials where survival analysis model was used. We present here some of the techniques of survival analysis for cancer data especially lung and breast. [8], showed the survival advantage in patients diagnosed with early breast cancer, treated with post-surgery radiation results showed that the best rates of survival were found with combined radiation and breast conserving surgery in all cases. The available data indicate that post-surgery radiation provides a survival advantage irrespective of the type of surgery in node positive patients. Likewise, survival advantage was observed with post-surgery radiation and breast-conserving procedure in node negative patients. [9], determined that surrogate endpoints for lung cancer specific survival may reduce the length of the clinical trials for a patient's lung cancer. A study performed by [10], provided a comparison of survival between African American and White men at the four distinct stages of lung cancer under the same treatment. Moreover, the study made it possible to estimate the average difference in survival between White and African American males diagnosed with lung cancer and addressed some of the critical issues related to the treatment lung cancer patients in survival cure models were discuss in [11] and [12] respectively.

3 MATERIALS AND METHODS

3.1 DATA COLLECTION

The data was obtained from Yobe State Specialist Hospital, Damaturu where each sample was collected from the recipients of the treatment of radical prostatectomy.

3.2 THE SURVIVAL FUNCTION

The survival function represents the probability that an event has not occurred by time t. Mathematically, it's defined as:

$$S(t) = P(T > t) \tag{1}$$

where T is a random variable representing the time until the event of interest (like death, failure, etc.). So, $S(t)$ gives the probability that the event has not occurred by time t.
Given a random variable T that denotes the survival time, the survival function denoted as $S(t)$ is defined as:

$$S(t) = P(T > t) = 1 - F(t) = 1 - \int_0^t f(u)\, du$$

where $f(t)$ and $F(t)$ are the probability density and the cumulative density functions respectively of a given distribution. The expression in (1) is the probability of surviving beyond time t. Note that $S(0) = 1$, $S(t) \to 0$ as $t \to \infty$. It is a downward sloping curve and can be estimated by using the Kaplan-Meier method.

3.3 THE HAZARD FUNCTION

The hazard function represents the instantaneous rate of occurrence of the event of interest at time t, given that it hasn't occurred before time t. Therefore, the hazard rate is defined as:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \tag{2}$$

Essentially, it's the conditional probability density function of the event occurring at time t, given survival up to time t, divided by the time interval. In simpler terms, it's the probability of the event happening at time t, given that the subject has survived up to t. Given a set containing individuals who are at a risk of experiencing a certain event denoted by $R(t)$ (risk set) or individuals who have not yet experienced the event by time t, the probability of an individual in the risk set experiencing the event in the small time interval $[t, t + \Delta t)$ is defined as $h(t)\Delta t$. Unlike the survival function which is a downward sloping curve for any type of survival data given, the hazard function takes on any shape of a non-negative function and it varies depending on the type of survival data given. This gives the cumulative hazard up to time t, which can be interpreted as the expected number of events that have occurred up to time t per unit time.

3.4 CUMULATIVE HAZARD FUNCTION (H(T))

The cumulative hazard function represents the cumulative risk of experiencing the event of interest up to time t. It's calculated by integrating the hazard function over the interval from 0 to t. Mathematically:

$$H(t) = \int_0^t h(u)\, du \tag{3}$$

The hazard function can be alternatively represented in terms of the cumulative hazard function $H(t)$. The name cumulative is due to the fact that the function is the accumulation of the hazard over time. According to [13], the cumulative hazard rate can be estimated using the Nelson-Aalen estimate and the increments of the Nelson-Aalen estimate can be smoothed to provide an estimate to the hazard rate. From equation (4), if $\hat{S}(t)$ is the Kaplan-Meier estimate to the survival function, then:

$$\widehat{H}(t) = -\sum_{i=1}^{k} ln\left(1 - \frac{di}{ni}\right)$$

is an estimate to the cumulative hazard function.
From Taylor series expansion:

$$ln\left(1 - \frac{di}{ni}\right) = -\frac{di}{ni} - \left[-\frac{di}{ni}\right] + \ldots = -\frac{di}{ni}$$

by ignoring higher order terms. The estimate to the cumulative hazard function is therefore given as:

$$\widehat{H}(t) = \sum_{i=1}^{k} \frac{di}{ni} \tag{4}$$

3.4.1 THE RELATIONSHIP BETWEEN THE HAZARD AND THE SURVIVAL FUNCTIONS

From equation (2) and (3) above, The relationship between the hazard function $h(t)$ and the survival function $S(t)$ is fundamental in survival analysis. The hazard function describes the instantaneous rate of failure at time t, while the survival function represents the probability of surviving beyond time t. These two functions are interconnected and can be derived from each other. The relationship between the hazard function $h(t)$ and the survival function $S(t)$ is given by:

$$h(t) = -\frac{d}{dt}\log\big(S(t)\big) \tag{5}$$

or equivalently:

$$S(t) = \exp\left(-\int_0^t h\,(u)du\right) \tag{6}$$

where: - $\frac{d}{dt}$ represents the derivative with respect to time t. - $\log(\cdot)$ is the natural logarithm. $\exp(\cdot)$ is an exponential function. $\int_0^t h\,(u)du$ is the cumulative hazard function, representing the cumulative risk up to time t.

3.4.2 Interpretation

The hazard function describes how the risk of an event changes with time. The survival function represents the probability of not experiencing the event up to time t. The hazard function and survival function provide complementary perspectives on the survival process: one focuses on the risk of experiencing the event at a specific time, while the other focuses on the probability of avoiding the event up to that time. Assuming that in the given sample of survival data none of the data points is censored and that also there exists no tied observations. The survival function denoted as $S\,(t)$ which is the probability that an individual survives beyond time t can be estimated by using the empirical survival function. The empirical function which is the estimate to the survival function in absence of censored data denoted as $\hat{S}(t)$ is given by:

$$\hat{S}(t) = \frac{\text{Number of individuals with survival time} \geq t}{\text{Number of individuals in the data set}}$$

$\hat{S}(t) = 1$ for all values of t before the first failure and $\hat{S}(t) = 0$ after the final failure or occurrence of event. The estimated survival function ($\hat{S}(t)$) is observed to be constant between two adjacent times and therefore its plot turns out to be a step function, this function decreases immediately after each observed event time [14,15].

3.5 NON-PARAMETRIC METHODS

3.5.1 The Kaplan-Meier Estimator (K-M)

The Kaplan-Meier estimator, also known as the product limit estimator was presented by [16]. It gives a simple and quick estimate of the survival function in the presence of censoring. It uses the exact failure time [14]. The Kaplan-Meier estimator (K-M) is a non-parametric method used to estimate the survival function from lifetime data in the presence of censored observations. It is commonly used in survival analysis to estimate the probability of surviving beyond a given time point.

Kaplan-Meier Estimator Formula

Suppose we have n observed survival times $t_1, t_2, \dots, t_n$, with corresponding censoring indicators $\delta_1, \delta_2, \dots, \delta_n$, where t_i is the time of event or censoring, and $\delta_i = 1$ if event occurred at t_i, and $\delta_i = 0$ if event was censored at t_i.

The Kaplan-Meier estimator of the survival function $S(t)$ at time t is given by:

$$\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right) \tag{7}$$

where: d_i is the number of events (deaths) at time t_i. n_i is the number of individuals at risk just before time t_i. In words, the Kaplan-Meier estimator calculates the probability of surviving beyond each event time, considering the individuals at risk just before each event. The Kaplan-Meier estimator estimates the survival function by calculating the probability of surviving beyond each observed event time, adjusting for censoring. The first point to consider is how censoring can be adjusted in the K-M method in order to estimate the survival function. As the K-M method makes no assumption about the shape of the underlying survival curve, it is categorized as a non-parametric method for estimating a survival function. However, using a non-parametric analysis typically generated much wider confidence bounds than those calculated via parametric analysis. Parametric analysis shows how predictions outside the range of observations are not possible with non-parametric analysis. The characterization of all the subjects of the survival analysis by K-M method can use only three variables [17]. The first variable is the serial time which begins with the commencement of the treatment and gets censored from the study when it reaches the end point. At the end of the serial time, the second variable consists of the patient's status. The third variable is the study groups the patients belong to. The idea of this method is based on the probability of surviving in k or more periods in the study and is a product of k probabilities when each period is observed under it. It is written by the following expression:

$$(k) = p_1 \times p_2 \times p_3 \times \dots \times p_k \tag{8}$$

In the above equation p_1 constitutes surviving proportion in the first period, p_2 is the proportion survived over the second period, and so on. The equation below gives the proportion of surviving for period i where they survived up to period $\dot{i}$:

$$p_i = \frac{r_i - d_i}{r_i} \tag{9}$$

where, r_i is the number of patients living at start of the period i, and d_i is the number of deaths, [18].

Order the observed event times: $t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(k)}$, where $t_{(i)}$ are distinct event times.

Initialize: $n_0 = n$ and $d_0 = 0$.

For each event time $t_{(i)}$:

Calculate the number of individuals at risk just before $t_{(i)}$: $n_i = n_{i-1} - d_{i-1}$.

Count the number of events at $t_{(i)}$: d_i.

Update $S(t_{(i)})$:

$$\hat{S}(t_{(i)}) = \hat{S}(t_{(i-1)}) \times \left(1 - \frac{d_i}{n_i}\right) \quad (10)$$

Survival function at t:

If $t < t_{(1)}, \hat{S}(t) = 1$.

If $t \geq t_{(k)}, \hat{S}(t) = 0$.

For $t_{(i)} < t \leq t_{(i+1)}, \hat{S}(t) = \hat{S}(t_{(i)})$.

3.6 SEMI-PARAMETRIC METHODS

3.6.1 Cox Proportional Hazard

The non-parametric methods are not useful for controlling the covariates and it requires categorical predictors. Therefore, multivariate approaches are used when we have several prognostic variables. The most widely applicable and broadly implemented multivariate method in the survival analysis is the regression model of the Cox proportional hazards. In the year 1972 Cox showed the first light to the Cox model [19], The explanatory variables and the response variables are combined. As any form can be adopted by the disturbance of the baseline, the nature of the model is semi-parametric [20].

The mathematical equation of the Cox model is:

$$h(t) = \exp\{h_0(t) + b_1x_1 + b_2x_2 + \cdots + b_px_p\} \quad (11)$$

3.6.2 Testing the Proportional Hazards Assumption

The assumption of proportional hazards (PH) function is the finest technique in the Cox model. This model helps clarify the idea that multiplicative effect of each covariate in the hazards function is constant over time [21]. Quite often the assumption of PH is substantially important.

3.7 ESTIMATION OF UNKNOWN PARAMETERS IN BOTH PARAMETRIC AND SEMI-PARAMETRIC REGRESSION MODELS

To demonstrate how to determine the estimates of the unknown parameters in both parametric and semi-parametric regression models we used the most common model in survival analysis, the Cox-PH model. The main objective in fitting the Cox proportional hazard model is to come up with estimates of the regression parameters (β s). Assuming that there are no tied event times, [19].

3.7.1 Cox Proportional Hazards Model

The Cox-PH model is a semi-parametric model that relates the time until an event occurs to one or more predictor variables. The hazard function $h(t|X)$ in the Cox model is given by [17, 19]:

$h(t|X) = h_0(t)\exp(X^\top\beta)$. $h(t|X)$ is the hazard function at time t given the covariates X. - $h_0(t)$ is the baseline hazard function, which is unspecified and depends only on time t. - X is the vector of covariates. - β is the vector of regression parameters to be estimated.

3.7.2 Estimation of Regression Parameters (βs)

To estimate the regression parameters β, [19] proposed maximizing the partial likelihood. The partial likelihood focuses on the ordering of events rather than their exact timing, allowing for the estimation of β without specifying $h_0(t)$.

3.7.3 Partial Likelihood Function

Given a sample of n individuals with observed event times $t_1, t_2, \dots, t_n$ and corresponding covariate vectors $X_1, X_2, \dots, X_n$, the partial likelihood function $L(\beta)$ is:

$$L(\beta) = \prod_{i=1}^{n}\left[\frac{\exp\left(X_i^\top\beta\right)}{\sum_{j\in R(t_i)}\exp\left(X_j^\top\beta\right)}\right]^{\delta_i}$$

where: δ_i is an indicator variable that is 1 if the event is observed for individual i, and 0 if the data is censored. - $R(t_i)$ is the risk set at time t_i, consisting of individuals who are still at risk just before time t_i.

The log partial likelihood function is:

$$\ell(\beta) = \sum_{i=1}^{n} \delta_i \left[X_i^{\top}\beta - \log\left(\sum_{j\in R(t_i)} \exp\left(X_j^{\top}\beta\right)\right)\right]$$
(12)

where $R(t_i) = \{j : t_j \geq t_i\}$, denotes the risk set at time t_i, n represents the number of individuals in the data set and m the observed survival times. Only event times contribute their factor to the numerator but both the censored and uncensored observations are included in the denominator where the sum over the risk set includes all individuals who are still at risk just before time t_i. It is easy to work with the partial log-likelihood which is given by

3.7.4 Maximizing the Partial Likelihood

To estimate β, we maximize the log partial likelihood function $\ell(\beta)$. This is typically done using numerical optimization techniques, such as the Newton-Raphson method or other iterative algorithms. The score function $U(\beta)$ and the observed information matrix $I(\beta)$ are derived from the log partial likelihood:

$$U(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} \delta_i \left[X_i - \frac{\sum_{j\in R(t_i)} X_j \exp\left(X_j^{\top}\beta\right)}{\sum_{j\in R(t_i)} \exp\left(X_j^{\top}\beta\right)}\right]$$
(13)

Let $\hat{\beta}$, denote the maximum partial likelihood estimate for β obtained by maximizing the partial log-likelihood function (11), the first derivative of $l(\beta)$ with respect to β is called a vector of efficient scores and is given by:

$$I(\beta) = -\frac{\partial^2 \ell(\beta)}{\partial \beta\, \partial \beta^{\top}} = \sum_{i=1}^{n} \delta_i \left[\frac{\sum_{j\in R(t_i)} X_j X_j^{\top} \exp\left(X_j^{\top}\beta\right)}{\sum_{j\in R(t_i)} \exp\left(X_j^{\top}\beta\right)} - \left(\frac{\sum_{j\in R(t_i)} X_j \exp\left(X_j^{\top}\beta\right)}{\sum_{j\in R(t_i)} \exp\left(X_j^{\top}\beta\right)} \right)^2 \right]$$
(14)

Using these, the Newton-Raphson update for β is:

$$\beta^{(k+1)} = \beta^{(k)} - \left[I\left(\beta^{(k)}\right)\right]^{-1} U\left(\beta^{(k)}\right)$$

where $\beta^{(k)}$ is the estimate at the k-th iteration.

To calculate the maximum likelihood estimates $\hat{\beta}$, we solve a nonlinear system $U(\beta) = 0$ and we use the Newton-Raphson algorithm [22]**.** The information matrix $I(\beta)$ is given by the negative of the second derivative of $l(\beta)$. The information matrix $I(\beta)$ is given by the negative of the second derivative of the log-likelihood function $\ell(\beta)$, and the maximum likelihood estimate $\hat{\beta}$ follows an asymptotic p-variate normal distribution.

Information Matrix $I(\beta)$

The information matrix $I(\beta)$ is defined as:

$$I(\beta) = -\frac{\partial^2 \ell(\beta)}{\partial \beta \, \partial \beta^{\mathrm{T}}} \quad (15)$$

The inverse of the information matrix is a consistent estimate of the covariance matrix of $\hat{\beta}$. It is used to construct confidence intervals for the components of β.

Asymptotic Normality

For large sample sizes, the maximum likelihood estimate $\hat{\beta}$ is asymptotically normally distributed with mean β and covariance matrix $I(\beta)^{-1}$.

$$\hat{\beta} \sim N(\beta, I(\beta)^{-1}) \quad (16)$$

Where: - $\hat{\beta}$ is the maximum likelihood estimate of β. - β is the true parameter vector. - $I(\beta)^{-1}$ is the inverse of the information matrix. For large samples, the maximum likelihood estimate $\hat{\beta}$ is known to follow asymptotic p-variate normal distribution:

Covariance Matrix of $\widehat{\boldsymbol{\beta}}$

The covariance matrix of $\hat{\beta}$ is estimated by the inverse of the information matrix:

$$\mathrm{Cov}(\hat{\beta}) = I(\hat{\beta})^{-1} \quad (17)$$

3.8 LOG RANK TEST

The log-rank is one commonly used non-parametric test for comparing two or more survival distributions of the patients; it is also called Mantel log-rank. Additionally, this method is useful when the risk of an event is always greater for one group than another in order to detect a difference between groups [18].

The calculation of the test is:

$$\mathrm{X}^2(\log rank) = \frac{(\mathrm{O1}-\mathrm{E1})}{\mathrm{E1}} + \frac{(\mathrm{O2}-\mathrm{E2})}{\mathrm{E2}} \quad (18)$$

Here, O_1 and O_2 stand for the number of total events that have been observed within the groups of 1 and 2 respectively. The expected number of events is represented by E_1 and E_2.

The *survdiff* function in R implements the log rank test. In our study, this method is useful to detect the difference between two groups of tumor - primary and metastasis.

3.9 LIKELIHOOD RATIO AND WALD TESTS

[13], argues that to test for a simple null hypothesis, one may use the likelihood-based tests. These tests include; The Likelihood ratio test, Wald's test and the Score test. These tests are asymptotically

equivalent and they all follow a chi-square distribution with p degrees of freedom. Where p is the dimension of the vector of the regression parameters. The Likelihood Ratio Test (LRT) and the Wald Test are commonly used to test hypotheses about the regression parameters (β) in the Cox Proportional Hazards (Cox-PH) model.

3.9.1 Likelihood Ratio Test (LRT)

The Likelihood Ratio Test compares the fit of two nested models: a full model with parameters β and a reduced model (null model) with parameters β_0, where some of the parameters are constrained to zero.

Calculate the Log Partial Likelihoods:

$\ell(\hat{\beta})$: Log partial likelihood of the full model.

$\ell(\widehat{\beta_0})$: Log partial likelihood of the reduced model.

Compute the Test Statistic:

$$\text{LR} = 2\left[\ell(\hat{\beta}) - \ell(\widehat{\beta_0})\right] \quad (19)$$

The test statistic follows a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between the full and reduced models.

Decision Rule:

Compare the test statistic to the critical value from the chi-squared distribution.

If LR is greater than the critical value, reject the null hypothesis.

Wald Test

The Wald Test assesses the significance of individual coefficients or sets of coefficients in the model.

Estimate the Coefficients:

$\hat{\beta}$: Estimated coefficients from the model.

Compute the Variance-Covariance Matrix:

$\hat{\Sigma}$: Estimated variance-covariance matrix of $\hat{\beta}$.

Compute the Test Statistic: For testing the null hypothesis $H_0: \beta_j = 0$:

$$W = \frac{\widehat{\beta_j}^2}{\text{Var}(\widehat{\beta_j})} \quad (20)$$

For testing multiple coefficients, $H_0: \beta = 0$:

$$W = \hat{\beta}^\top \hat{\Sigma}^{-1} \hat{\beta}$$

The test statistic follows a chi-squared distribution with degrees of freedom equal to the number of parameters being tested.

Decision Rule:

Compare the test statistic to the critical value from the chi-squared distribution. If W is greater than the critical value, reject the null hypothesis.

Applying LRT and Wald Tests in Cox-PH Model

Consider a Cox-PH model with two covariates X_1 and X_2:

$h(t|X) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2)$

Likelihood Ratio Test

Full Model:

$h(t|X) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2)$

Reduced Model:

$h(t|X) = h_0(t)\exp(\beta_1 X_1)$

Compute Log Partial Likelihoods:

$\ell(\hat{\beta})$: Log partial likelihood for the full model.

$\ell(\widehat{\beta_0})$: Log partial likelihood for the reduced model.

Test Statistic:

$\text{LR} = 2\left[\ell(\hat{\beta}) - \ell(\widehat{\beta_0})\right]$

Compare the test statistic to a chi-squared distribution with 1 degree of freedom (difference in the number of parameters).

Wald Test

Estimate Coefficients:

$\hat{\beta} = (\widehat{\beta_1}, \widehat{\beta_2})$

Variance-Covariance Matrix:

$$\hat{\Sigma} = \begin{pmatrix} \text{Var}(\widehat{\beta_1}) & \text{Cov}(\widehat{\beta_1}, \widehat{\beta_2}) \\ \text{Cov}(\widehat{\beta_1}, \widehat{\beta_2}) & \text{Var}(\widehat{\beta_2}) \end{pmatrix}$$

Test Statistic:

$W = \widehat{\beta_2}^2 / \text{Var}(\widehat{\beta_2})$
(21)

For testing both coefficients:

$W = (\widehat{\beta_1}, \widehat{\beta_2})^{\top} \hat{\Sigma}^{-1} (\widehat{\beta_1}, \widehat{\beta_2})$

Decision Rule:

Compare W to the chi-squared distribution with the appropriate degrees of freedom.

Reject the null hypothesis if W is greater than the critical value.

These tests help determine whether the covariates significantly contribute to the model, enhancing the understanding of the relationship between the predictors and the survival times.

3.10 VARIABLES AND BEST FITTING MODEL SELECTION SCHOENFELD RESIDUALS

In order to fit the standard cox-proportional hazard model, one has to be aware of one of its main assumptions. The model assumes that the hazard of the different strata formed by the levels of the covariates are proportional [24]. One can use the Kaplan-Meier plots to test for this assumption, but these graphical techniques may be inadequate in cases where the violation of the proportional hazard assumption is marginal. [25], presents the Goodness of fit (GOF) testing approach. This approach gives test statistics and a p-value for assessing the proportional hazard assumption. This test enables a researcher to make an objective decision than when using the graphical method. A number of tests have been presented in literature but for this research we used the one discussed by [26], Schoenfeld residuals are further discussed by [27]. The idea behind this statistical test is that if the PH assumption holds for a particular covariate, then the Schoenfeld residuals for that covariate will not be related to the survival time was discuss by many authors like [28, 29, 30 and 31] respectively.

3.11 AKAIKE INFORMATION CRITERIA

According to [32], Akaike Information Criterion in the Cox Proportional Hazards Model. The Akaike Information Criterion (AIC) is a measure used to compare the goodness of fit of statistical models while penalizing for the number of estimated parameters to prevent overfitting. It is widely used in model selection. For a given model, the AIC is calculated as follows:

$$\text{AIC} = -2\ell(\hat{\beta}) + 2k \quad (22)$$

where: - $\ell(\hat{\beta})$ is the log-likelihood of the model at its maximum likelihood estimates. - k is the number of estimated parameters in the model.

Application in the Cox Proportional Hazards Model

In the context of the Cox Proportional Hazards model, the AIC is used to compare different Cox models with varying numbers of covariates or different functional forms of the covariates. For a Cox-PH model, the AIC can be expressed as:

$$\text{AIC} = -2\ell(\hat{\beta}) + 2p$$

where: - $\ell(\hat{\beta})$ is the log partial likelihood of the Cox model evaluated at the estimated parameters $\hat{\beta}$. - p is the number of covariates (or parameters) in the model.

Model Comparison

As the measure of GOF of a model defined with parameters estimated by the maximum likelihood method. The value of the AIC will always increase if an unnecessary variable is included in the model. This therefore implies that the smaller the AIC the better the model. To compare different models using AIC: - Fit each model to the data. - Compute the AIC for each model. Select the model with the lowest AIC value, indicating the best trade-off between goodness of fit and model complexity.

4 RESULTS AND DISCUSSION

This research investigates the influence of standard clinical prognostic features on the survival time of lung cancer patients. Particularly it seeks independent variable patterns to determine the survival times and identify the correlations among the variables of interest. For this goal, the Cox model performed well, which identified covariates associated with survival. The result of each method was performed by statistical package in R, which was used to analyze the data. After applying the Kaplan–Meier (K-M) method to the RP data, the results are tabulated in Table 1, as shown below.

Table 1: Calculation for the K-M Estimate of the Survival Function.

Time	n. risk	n. event	Survival	Std. error	Lower 95% CI	Upper 95% CI
5	228	1	0.9956	0.00438	0.9871	1.000
11	227	3	0.9825	0.00869	0.9656	1.000
12	224	1	0.9781	0.00970	0.9592	0.997
13	223	2	0.9693	0.01142	0.9472	0.992
15	221	1	0.9649	0.01219	0.9413	0.989
26	220	1	0.9605	0.01290	0.9356	0.986
30	219	1	0.9561	0.01356	0.9299	0.983
31	218	1	0.9518	0.01419	0.9243	0.980
53	217	2	0.943	0.01536	0.9134	0.974
54	215	1	0.9386	0.01590	0.9079	0.970
59	214	1	0.9342	0.01642	0.9026	0.967
60	213	2	0.9254	0.01740	0.8920	0.960
61	211	1	0.9211	0.01786	0.8867	0.957
62	210	1	0.9167	0.01830	0.8815	0.953
65	209	2	0.9079	0.01915	0.8711	0.946
71	207	1	0.9035	0.01955	0.8660	0.943
79	206	1	0.8991	0.01995	0.8609	0.939
81	205	2	0.8904	0.02069	0.8507	0.932
88	203	2	0.8816	0.02140	0.8406	0.925
92	201	1	0.8772	0.02174	0.8356	0.921
93	199	1	0.8728	0.02207	0.8306	0.917
95	198	2	0.864	0.02271	0.8206	0.910
105	196	1	0.8596	0.02302	0.8156	0.906
107	194	2	0.8507	0.02362	0.8056	0.898
110	192	1	0.8463	0.02391	0.8007	0.894
116	191	1	0.8418	0.02419	0.7957	0.891
118	190	1	0.8374	0.02446	0.7908	0.887
122	189	1	0.833	0.02473	0.7859	0.883
131	188	1	0.8285	0.02500	0.7810	0.879
132	187	2	0.8197	0.02550	0.7712	0.871
135	185	1	0.8153	0.02575	0.7663	0.867
142	184	1	0.8108	0.02598	0.7615	0.863
144	183	1	0.8064	0.02622	0.7566	0.859
145	182	2	0.7975	0.02667	0.7469	0.852
180	180	1	0.7931	0.02688	0.7421	0.848
153	179	1	0.7887	0.02710	0.7373	0.844
156	178	2	0.7798	0.02751	0.7277	0.836
163	176	3	0.7665	0.02809	0.7134	0.824
166	173	2	0.7577	0.02845	0.7039	0.816
167	171	1	0.7532	0.02863	0.6991	0.811
170	170	1	0.7488	0.02880	0.6944	0.807
175	167	1	0.7443	0.02898	0.6896	0.803
176	165	1	0.7398	0.02915	0.6848	0.799
177	164	1	0.7353	0.02932	0.6800	0.795
179	162	2	0.7262	0.02965	0.6704	0.787
180	160	1	0.7217	0.02981	0.6655	0.783
181	159	2	0.7126	0.03012	0.6559	0.774
182	157	1	0.7081	0.03027	0.6511	0.770
183	156	1	0.7035	0.03041	0.6464	0.766
186	154	1	0.6989	0.03056	0.6416	0.761
189	152	1	0.6943	0.03070	0.6367	0.757
194	149	1	0.6897	0.03085	0.6318	0.753
197	147	1	0.6850	0.03099	0.6269	0.749
199	145	1	0.6803	0.03113	0.6219	0.744

201	144	2	0.6708	0.03141	0.6120	0.735
202	142	1	0.6661	0.03154	0.6071	0.731
207	139	1	0.6613	0.03168	0.6020	0.726
208	138	1	0.6565	0.03181	0.5970	0.722
210	137	1	0.6517	0.03194	0.5920	0.717
212	135	1	0.6469	0.03206	0.5870	0.713
218	134	1	0.6421	0.03218	0.5820	0.708
222	132	1	0.6372	0.03231	0.5769	0.704
223	130	1	0.6323	0.03243	0.5718	0.699
226	126	1	0.6273	0.03256	0.5666	0.694
229	125	1	0.6223	0.03268	0.5614	0.690
230	124	1	0.6172	0.03280	0.5562	0.685
239	121	2	0.6070	0.03304	0.5456	0.675
245	117	1	0.6019	0.03316	0.5402	0.670
246	116	1	0.5967	0.03328	0.5349	0.666
267	112	1	0.5913	0.03341	0.5294	0.661
268	111	1	0.5860	0.03353	0.5239	0.656
269	110	1	0.5807	0.03364	0.5184	0.651
270	108	1	0.5753	0.03376	0.5128	0.645
283	104	1	0.5698	0.03388	0.5071	0.640
284	103	1	0.5642	0.03400	0.5014	0.635
285	101	2	0.5531	0.03424	0.4899	0.624
286	99	1	0.5475	0.03434	0.4841	0.619
288	98	1	0.5419	0.03444	0.4784	0.614
291	97	1	0.5363	0.03454	0.4727	0.608
293	94	1	0.5306	0.03464	0.4669	0.603
301	91	1	0.5248	0.03475	0.4609	0.597
303	89	1	0.5189	0.03485	0.4549	0.592
305	87	1	0.5129	0.03496	0.4488	0.586
306	86	1	0.5070	0.03506	0.4427	0.581
310	85	2	0.4950	0.03523	0.4306	0.569
320	82	1	0.4890	0.03532	0.4244	0.563
329	81	1	0.4830	0.03539	0.4183	0.558
337	79	1	0.4768	0.03547	0.4121	0.552
340	78	1	0.4707	0.03554	0.4060	0.546
345	77	1	0.4646	0.03560	0.3998	0.540
348	76	1	0.4585	0.03565	0.3937	0.534
350	75	1	0.4524	0.03569	0.3876	0.528
351	74	1	0.4463	0.03573	0.3815	0.522
353	73	2	0.4340	0.03578	0.3693	0.510
361	70	1	0.4278	0.03581	0.03581	0.504
363	69	2	0.4154	0.03583	0.3508	0.492
364	67	1	0.4092	0.03582	0.3447	0.486
371	65	2	0.3966	0.03581	0.3323	0.473
387	60	1	0.3900	0.03582	0.3258	0.467
390	59	1	0.3834	0.03582	0.3193	0.460
394	58	1	0.3768	0.03580	0.3128	0.454
426	55	1	0.3700	0.03580	0.3060	0.447
428	54	1	0.3631	0.03579	0.2993	0.440
429	53	1	0.3563	0.03576	0.2926	0.434
433	52	1	0.3494	0.03573	0.2860	0.427
442	51	1	0.3426	0.03568	0.2793	0.420
444	50	1	0.3357	0.03561	0.2727	0.413
450	48	1	0.3287	0.03555	0.2659	0.406
455	47	1	0.3217	0.03548	0.2592	0.399
457	46	1	0.3147	0.03539	0.2525	0.392

460	44	1	0.3076	0.03530	0.2456	0.385
473	43	1	0.3004	0.03520	0.2388	0.378
477	42	1	0.2933	0.03508	0.2320	0.371
519	39	1	0.2857	0.03498	0.2248	0.363
520	38	1	0.2782	0.03485	0.2177	0.356
524	37	2	0.2632	0.03455	0.2035	0.340
533	34	1	0.2554	0.03439	0.1550	0.333
550	32	1	0.2475	0.03423	0.1475	0.325
558	30	1	0.2392	0.03407	0.1570	0.316
567	28	1	0.2307	0.03391	0.1550	0.308
574	27	1	0.2221	0.03371	0.1650	0.299
583	26	1	0.2136	0.03348	0.1571	0.290
613	24	1	0.2047	0.03325	0.1489	0.281
624	23	1	0.1958	0.03297	0.1407	0.272
641	22	1	0.1869	0.03265	0.1327	0.263
643	21	1	0.1780	0.03229	0.1247	0.254
654	20	1	0.1691	0.03188	0.1169	0.245
655	19	1	0.1602	0.03142	0.1091	0.235
687	8	1	0.1513	0.03090	0.1014	0.226
689	17	1	0.1424	0.03034	0.0938	0.216
705	6	1	0.1335	0.02972	0.0863	0.207
707	15	1	0.1246	0.02904	0.0789	0.197
728	14	1	0.1157	0.02830	0.0716	0.187
731	13	1	0.1068	0.02749	0.0645	0.177
735	12	1	0.0979	0.02660	0.0575	0.167
765	10	1	0.0881	0.02568	0.0498	0.156
791	9	1	0.0783	0.02462	0.0423	0.145
814	7	1	0.0671	0.02351	0.0338	0.133
883	4	1	0.0503	0.02285	0.0207	0.123
11	227	3	0.7798	0.02751	0.7277	0.836
12	224	1	0.7665	0.02809	0.7134	0.824
13	223	2	0.7577	0.02845	0.7039	0.816
15	221	1	0.7532	0.02863	0.6991	0.811
26	220	1	0.7488	0.02880	0.6944	0.807
30	219	1	0.7443	0.02898	0.6896	0.803
31	218	1	0.7398	0.02915	0.6848	0.799
53	217	2	0.7353	0.02932	0.6800	0.795
54	215	1	0.7262	0.02965	0.6704	0.787
59	214	1	0.7217	0.02981	0.6655	0.783
60	213	2	0.7126	0.03012	0.6559	0.774
61	211	1	0.7081	0.03027	0.6511	0.770
62	210	1	0.7035	0.03041	0.6464	0.766
65	209	2	0.6989	0.03056	0.6416	0.761
71	207	1	0.6943	0.03070	0.6367	0.757
79	206	1	0.6897	0.03085	0.6318	0.753
81	205	2	0.6850	0.03099	0.6269	0.749
88	203	2	0.6803	0.03113	0.6219	0.744
92	201	1	0.6708	0.03141	0.6120	0.735
93	199	1	0.6661	0.03154	0.6071	0.731
95	198	2	0.6613	0.03168	0.6020	0.726
105	196	1	0.6565	0.03181	0.5970	0.722
107	194	2	0.6517	0.03194	0.5920	0.717
110	192	1	0.6469	0.03206	0.5870	0.713
116	191	1	0.6421	0.03218	0.5820	0.708
118	190	1	0.6372	0.03231	0.5769	0.704
122	189	1	0.6323	0.03243	0.5718	0.699

131	188	1	0.6273	0.03256	0.5666	0.694
132	187	2	0.6223	0.03268	0.5614	0.690
135	185	1	0.6172	0.03280	0.5562	0.685
142	184	1	0.6070	0.03304	0.5456	0.675
144	183	1	0.6019	0.03316	0.5402	0.670
145	182	2	0.5967	0.03328	0.5349	0.666
147	180	1	0.5913	0.03341	0.5294	0.661
153	179	1	0.5860	0.03353	0.5239	0.656
156	178	2	0.5807	0.03364	0.5184	0.651
163	176	3	0.5753	0.03376	0.5128	0.645
166	173	2	0.5698	0.03388	0.5071	0.640
167	171	1	0.5642	0.03400	0.5014	0.635
170	170	1	0.5531	0.03424	0.4899	0.624
175	167	1	0.5475	0.03434	0.4841	0.619
176	165	1	0.5419	0.03444	0.4784	0.614
177	164	1	0.5363	0.03454	0.4727	0.608
179	162	2	0.5306	0.03464	0.4669	0.603
180	160	1	0.5248	0.03475	0.4609	0.597
181	59	2	0.5189	0.03485	0.4549	0.592
182	157	1	0.5129	0.03496	0.4488	0.586
183	156	1	0.5070	0.03506	0.4427	0.581
186	154	1	0.4950	0.03523	0.4306	0.569
189	152	1	0.4890	0.03532	0.4244	0.563
194	149	1	0.4830	0.03539	0.4183	0.558
197	147	1	0.4768	0.03547	0.4121	0.552
199	145	1	0.4707	0.03554	0.4060	0.546
201	144	2	0.4646	0.03560	0.3998	0.540
202	142	1	0.4585	0.03565	0.3937	0.534
207	139	1	0.4524	0.03569	0.3876	0.528
208	138	1	0.4463	0.03573	0.3815	0.522
210	137	1	0.4340	0.03578	0.3693	0.510
212	135	1	0.4278	0.03581	0.3631	0.504
218	134	1	0.4154	0.03583	0.3508	0.492
222	132	1	0.4092	0.03582	0.3447	0.486
223	130	1	0.3966	0.03581	0.3323	0.473
226	126	1	0.3900	0.03582	0.3258	0.467
229	125	1	0.3834	0.03582	0.3193	0.460
230	124	1	0.3768	0.03580	0.3128	0.454
239	121	2	0.3700	0.03580	0.3060	0.447
245	117	1	0.3631	0.03579	0.2993	0.440
246	116	1	0.3563	0.03576	0.2926	0.434
267	112	1	0.3494	0.03573	0.2860	0.427
268	111	1	0.3426	0.03568	0.2793	0.420
269	110	1	0.3357	0.03561	0.2727	0.413
270	108	1	0.3287	0.03555	0.2659	0.406
283	104	1	0.3217	0.03548	0.2592	0.399
284	103	1	0.3147	0.03539	0.2525	0.392
285	101	2	0.3076	0.03530	0.2456	0.385
286	99	1	0.3004	0.03520	0.2388	0.378
288	98	1	0.2933	0.03508	0.2320	0.371
291	97	1	0.2857	0.03498	0.2248	0.363
293	94	1	0.2782	0.03485	0.2177	0.356
301	91	1	0.2632	0.03455	0.2035	0.340
303	89	1	0.2554	0.03439	0.1962	0.333
305	87	1	0.2475	0.03423	0.1887	0.325
306	86	1	0.2392	0.03407	0.1810	0.316

310	85	2	0.2307	0.03391	0.1729	0.308
320	82	1	0.2221	0.03371	0.1650	0.299
329	81	1	0.2136	0.03348	0.1571	0.290
337	79	1	0.2047	0.03325	0.1489	0.281
340	78	1	0.1958	0.03297	0.1407	0.272
345	77	1	0.1869	0.03265	0.1327	0.263
348	76	1	0.1780	0.03229	0.1247	0.254
350	75	1	0.1691	0.03188	0.1169	0.245
351	74	1	0.1602	0.03142	0.1091	0.235
353	73	2	0.1513	0.03090	0.1014	0.226
361	70	1	0.1424	0.03034	0.0938	0.216
363	69	2	0.1335	0.02972	0.0863	0.207
364	67	1	0.1246	0.02904	0.0789	0.197
371	65	2	0.1157	0.02830	0.0716	0.187
387	60	1	0.1068	0.02749	0.0645	0.177
390	59	1	0.0979	0.02660	0.0575	0.167
394	58	1	0.0881	0.02568	0.0498	0.156
426	55	1	0.0783	0.02462	0.0423	0.145
428	54	1	0.0671	0.02351	0.0338	0.133
429	53	1	0.0503	0.02285	0.0207	0.123
433	52	1	0.7798	0.02751	0.7277	0.836
442	51	1	0.7665	0.02809	0.7134	0.824
444	50	1	0.7577	0.02845	0.7039	0.816
450	48	1	0.7532	0.02863	0.6991	0.811
455	47	1	0.7488	0.02880	0.6944	0.807
457	46	1	0.7443	0.02898	0.6896	0.803
460	44	1	0.7398	0.02915	0.6848	0.799
473	43	1	0.7353	0.02932	0.6800	0.795
477	42	1	0.7262	0.02965	0.6704	0.787
519	39	1	0.7217	0.02981	0.6655	0.783
520	38	1	0.7126	0.03012	0.6559	0.774
524	37	2	0.7081	0.03027	0.6511	0.770
533	34	1	0.7035	0.03041	0.6464	0.766
550	32	1	0.6989	0.03056	0.6416	0.761
558	30	1	0.6943	0.03070	0.6367	0.757
567	28	1	0.6897	0.03085	0.6318	0.753
574	27	1	0.6850	0.03099	0.6269	0.749
583	26	1	0.6803	0.03113	0.6219	0.744
613	24	1	0.6708	0.03141	0.6120	0.735
624	23	1	0.6661	0.03154	0.6071	0.731
641	22	1	0.6613	0.03168	0.6020	0.726
643	21	1	0.6565	0.03181	0.5970	0.722
654	20	1	0.6517	0.03194	0.5920	0.717
655	19	1	0.6469	0.03206	0.5870	0.713
687	18	1	0.6421	0.03218	0.5820	0.708
689	17	1	0.6372	0.03231	0.5769	0.704
705	16	1	0.6323	0.03243	0.5718	0.699
707	15	1	0.6273	0.03256	0.5666	0.694
728	14	1	0.6223	0.03268	0.5614	0.690
731	13	1	0.6172	0.03280	0.5562	0.685
735	12	1	0.6070	0.03304	0.5456	0.675
765	10	1	0.6019	0.03316	0.5402	0.670
791	9	1	0.5967	0.03328	0.5349	0.666
814	7	1	0.5913	0.03341	0.5294	0.661

KEY:

Time = Survival time of patients
n.risk = Number of risk
n.event = Number of events occurred
Survival = Survival Probability
Std. Error= Standard Error of Estimates
Lower 95%Cl = Lower 95 percent Confidence interval
Upper 95%Cl = Upper 95 percent Confidence interval

Here, we are interested in "time" and "status" as they play an important role in the analysis. Time represents the survival time of patients. Since patients survive, we will consider their status as dead or non-dead (censored).

4.1 SURVIVAL TIMES OF THE PATIENTS.

Here, the x-axis specifies "Number of days" and the y-axis specifies the "probability of survival ". The dashed lines are upper confidence interval and lower confidence interval. We also have the confidence interval which shows the margin of error expected i.e in days of surviving 200 days, upper confidence interval reaches 0.76 or 76% and then goes down to 0.60 or 60%. Table 1, presents the K-M Estimate of The Survival Function of the 95% confidence intervals from Table 2, fitting the univariate cox-proportional hazard model. The variables that were found significant by using the likelihood ratio test at 0.05 level of significance.

Table 2: Cox Fit Model

coef	exp (coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	z	p	-log2(p)	
age	0.01	1.01	0.01	-0.01	0.03	0.99	1.03	0.92	0.36	1.48
sex	-0.55	0.58	0.20	-0.94	-0.16	0.39	0.85	-2.74	0.01	7.36
ph.ecog	0.73	2.08	0.22	0.30	1.17	1.35	3.23	3.29	<0.005	9.95
ph.karno	0.02	1.02	0.01	0.00	0.04	1.00	1.05	2.00	0.05	4.45
pat.karno	-0.01	0.99	0.01	-0.03	0.00	0.97	1.00	-1.54	0.12	3.02
meal.cal	0.00	1.00	0.00	-0.00	0.00	1.00	1.00	0.13	0.90	0.16
wt.loss	-0.01	0.99	0.01	-0.03	0.00	0.97	1.00	-1.84	0.07	3.94

Concordance	0.65
Partial AIC	1011.50
log-likelihood ratio test	28.33 on 7 df
-log2(p) of ll-ratio test	12.35

Interpretation of the Summary Table 2:

There are 168 observations

There were 121 deaths which occurred (events observed)

coef-> gives the result of the model coefficients

In case of Cox-ph models, the model coefficients can be measures without measuring the baseline hazard function i.e., h(t0)h(t0). In general terms, the baseline hazard function is unspecified.

exp(coef) -> The hazard ratio

As a con of the Cox-PH model, we cannot estimate the survival. If we are not getting an estimate of the intercept, then we cannot in turn measure the hazard, which again means we cannot measure the survival function. But we can measure the HR. Interpretation of the HR: Considering the sex column, at a given instance of time, someone who has a sex is male, is 0.58 times as likely to die as someone who is not male, adjusting for mismatch level. In terms of %, it can be interpreted as (1-HR) % i.e. here it would be (1-0.58) % = 42%.

se(coef) -> The Standard Error in the observation of the coefficients

coef lower 95% and coef upper 95% -> The 95% confidence interval, which suggests that we are 95% confident that the true value of model coefficients is within this interval.

exp(coef) lower 95% and exp(coef) upper 95% -> The 95% confidence interval, which suggests that we are 95% confident that the true value of HR is within this interval.

z-score -> value of wald test obtained as coefse(coef)coefse(coef)

Concordance -> Goodness of fit for survival analysis, which is the fraction or percentage of the pairs of observations which are concordant. (Basically, how well the model is performing with respect to the actual scenario). Higher the value of Concordance, better is the fit.

AIC -> The Akaike information criterion (AIC) is an estimator of out-of-sample prediction error and thereby relative quality of statistical models for a given set of data. AIC estimates the relative amount of information lost by a given model: the less information a model loses, the higher the quality of that model.

Statistically significant values: From the summary table above, we can see the pp values for the different attributes.

Statistically Significant: A column will be regarded as statistically significant when the p-value is < 0.05. In this case, we can see the attributes, "sex" and "ph.ecog" have pp-values less than 0.05. We can also see attributes like age having a HR of 1.01, which suggests there is only 1% increase in risk factor for a higher age group. In general terms, it can be said that there is no significant difference between different age groups.

Inference: From this it can be understood while grouping the data during analysis, "sex" and "ph.ecog" should be the attributes on which we focus more. Here, we can see that the pp-value for "sex" is 0.01 and the HR = 0.58. This indicates, a strong relationship between the patients' sex and the decreased risk of death. In general terms, considering the other covariates to be constant, a

female (sex=2) patient has a higher chance of survival compared to a male patient. Again, for the attribute "ph.ecog", the pp-value is <0.005 and the HR is 2.08.This indicates a strong relationship between the ph.ecog value and the increased risk of death in a patient. In general terms, considering the other covariates to be constant, a higher value of "ph.ecog" is associated with more risk in survival. In this case, a patient with higher ph.ecog value has about 109% higher risk of death.

5 COMPARISON OF THE DIFFERENT GROUPS OF THE ATTRIBUTES USING THE KAPLAN MEIER CURVE

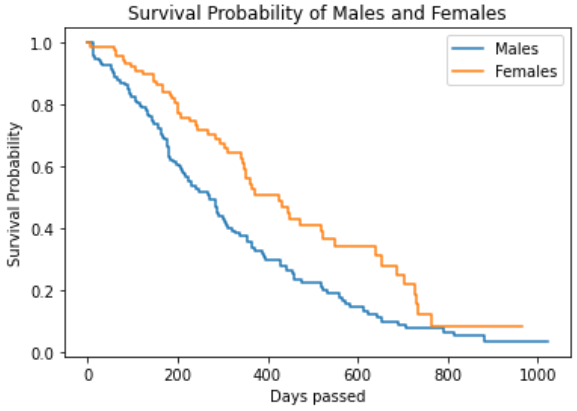

Figure.1: Survival Probability of Males and Females

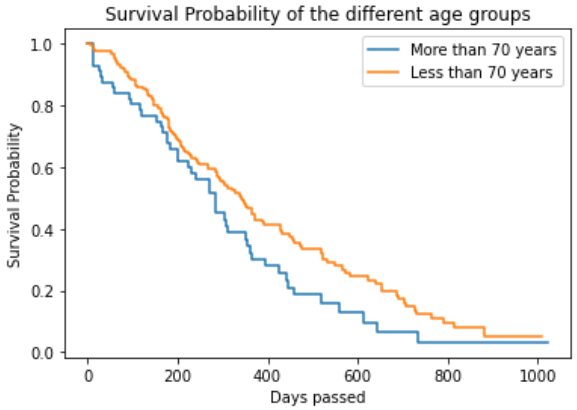

Figure 2: Survival Probability of the Different Age Groups

From Figure 2 above, it can be inferred that subjects who are older than 70 years have a comparatively lower chances of survival compared to subjects younger than 70 years.

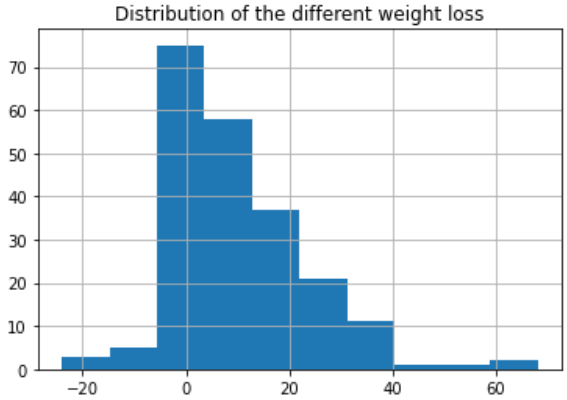

Figure 3: Distribution of the Different Weight Loss

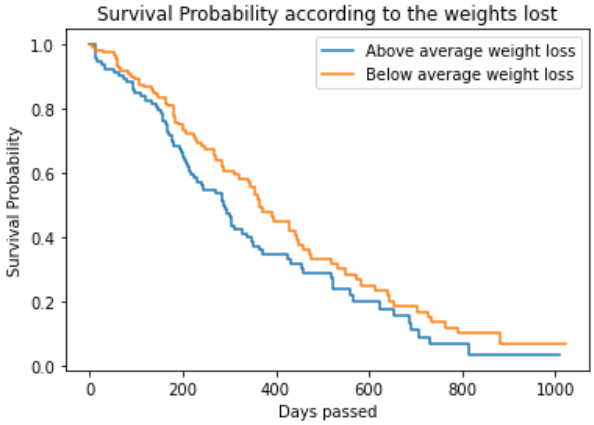

Figure 4: Survival Probability According to the Weight Loss

From Figure 4 above, it can be inferred that subjects having above average weight loss have lower chances of survival.

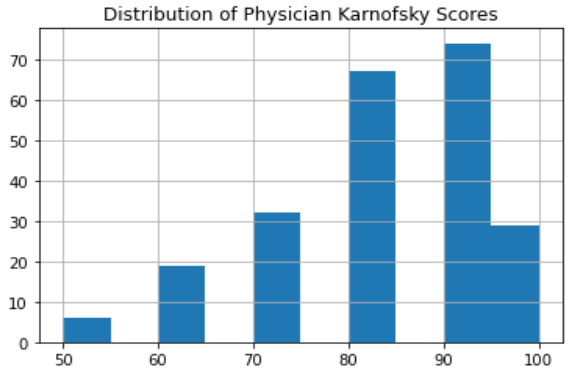

Figure 5: Distribution of Physician Karnofsky Scores

A ph.karno score of >= 80 will consider all patients who are able to carry on normally activity without special care. People with less than 80 are considered actually ill with people having > 40 scores as really sick.

Considering this, the two groups are made as:

People with <= 80 pat.karno score

People with > 80 pat.karno score

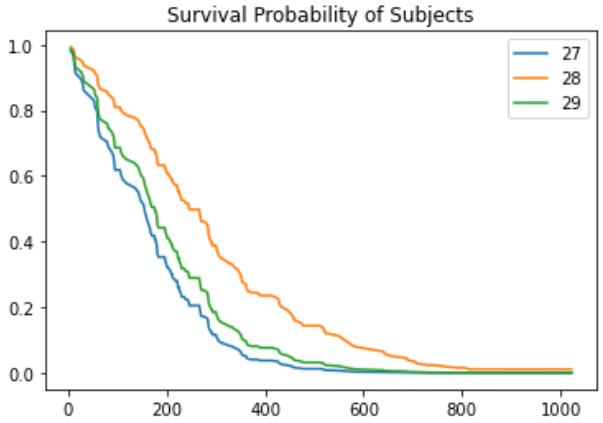

Figure 6: Survival Probability of Subjects

From the above Figure 6, it can be seen that the subject 28 has the highest chances of survival and subject 27 has the lowest chances of survival. Looking into the ph.ecog values in the table above, it can be seen that subject 28 has a low ph.ecog value (1.0) compared to subject 27 (3.0). This concords with the fact that higher the ph.ecog value lesser are the chances of survival.

Table 3: Estimate Coefficients of the Lung Cancer Covariates

coef	exp(coef)	se(coef)	z Pr(>|z|)
inst	-3.037e-02	9.701e-01	1.312e-02 -2.315 0.020619
age	1.281e-02	1.013e+00	1.194e-02 1.073 0.283403
sex	-5.666e-01	5.674e-01	2.014e-01 -2.814 0.004890
ph.ecog	9.074e-01	2.478e+00	2.386e-01 3.803 0.000143
ph. karno	2.658e-02	1.027e+00	1.163e-02 2.286 0.022231
pat. karno	-1.091e-02	9.891e-01	8.141e-03 -1.340 0.180160
meal.cal	2.602e-06	1.000e+00	2.677e-04 0.010 0.992244
wt.loss	-1.671e-02	9.834e-01	7.911e-03 -2.112 0.034647

Table 4: Lung Cancer Patient Using the Cox PH Model.

coef	exp(coef)	exp(-coef)	lower .95	upper .95
inst	0.9701	1.0308	0.9455	0.9954
age	1.0129	0.9873	0.9895	1.0369
sex	0.5674	1.7623	0.3824	0.8420
ph.ecog	2.4778	0.4036	1.5523	3.9552
ph. karno	1.0269	0.9738	1.0038	1.0506
pat. karno	0.9891	1.0110	0.9735	1.0051
meal.cal	1.0000	1.0000	0.9995	1.0005
wt.loss	0.9834	1.0169	0.9683	0.9988

Concordance	= 0.648 (se = 0.03)
Likelihood ratio test	= 33.7 on 8 df, p=5e-05
Wald test	= 31.72 on 8 df, p=1e-04
Score (logrank) test	= 32.51 on 8 df, p=8e-05

In the result, there are two tables: In Table 3, the second column presents the regression coefficient. The sign of the coefficients is an important issue to consider since a positive sign means the hazard ratio for this variable is higher, while the negative sign will decrease the hazard risk (risk of death). The column z in Table 3 records the ratio of each regression coefficient to its standard error; a wald statistic is asymptotically standard normal under the hypothesis that the corresponding coefficient is zero. Finally, p-value shows the significance of the explanatory variable. In Table 4, the asymptotically equivalent tests of the omnibus null hypothesis that all of the coefficients are zero are likelihood ratio, wald score, chi-square statistic at bottom of the output. We can conclude that in cox model, if the coefficient is negative the hazard will decrease, but if the coefficient is positive the hazard will increase. The plot of survival curves based on the cox model and Kaplan-Meier Estimates for the model is presented in Figure 1 to Figure 7 respectively. The estimated distribution of survival times for cox model is illustrated below by using *survefit* function graph (function to calculate survival time). It is illustrated by the estimate survival function.

The exp(coef) column contains eβ1eβ1 (see background section above for more info). This is the hazard ratio – the multiplicative effect of that variable on the hazard rate (for each unit increase in that variable).

For a categorical variable like sex, going from male (baseline) to female results in approximately ~40% reduction in hazard.

We could also flip the sign on the coef column, and take exp (0.531), which can be interpreted as:

- Males have a 1.71.7-fold increase in hazard, or that
- Males die at approximately 1.7×1.7× the rate per unit time as females (females die at 0.588×0.588× the rate per unit time as males).

Note,

- HR=1HR=1: No effect
- HR>1HR>1: Increase in hazard
- HR<1HR<1: Decrease in hazard

There is a p-value on the sex term, and a p-value on the overall model.

That 0.001110.00111 p-value is really close to the p=0.00131p=0.00131 p-value we saw on the Kaplan-Meier plot.

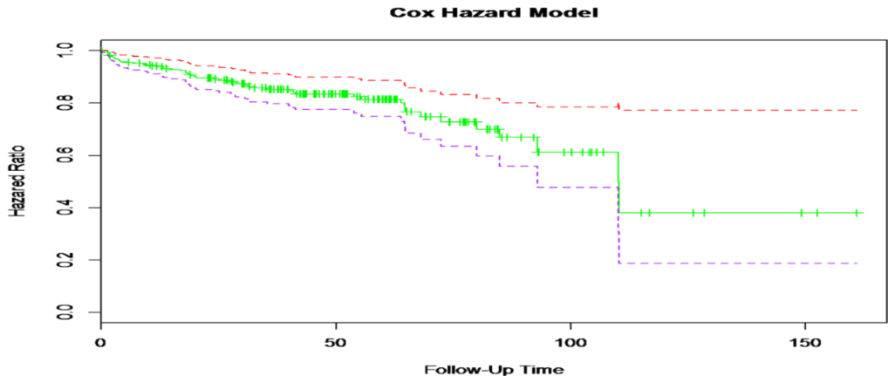

Figure 7: The Cox proportional hazard (PH) with error bars shows 95% confidence intervals.

The important step after fitting the model for Cox is to evaluate the adequacy of the fitted model. As we mention in section 3, the methods above, the model that checks the analysis is based on residuals. In the analysis for the cox model four major criteria of residuals have been described, they are the Cox Snell residual, the deviance residual, martingale residual and the Schoenfeld residual.

5.1 INTERPRETATION OF FIGURES

From the above Figures 1 to Figure 7 it can be inferred that males have a comparatively lower chances of survival compared to females. The construction of a table is a necessary first step in order to analyze the K-M estimate, which requires three elements to function. These elements are serial time (survival time by month), status at serial time (1= death, 0=censored), and group (1 = male and 2= female). An Excel spreadsheet was used to build the table, beginning with the shortest times for each group and sorted by ascending serial time, which is shown in Table 1 above. The initial table is preparation for K-M analysis to be used by statistical program R [28]. The plot of survival curves is

an important part of survival analysis for each group of interest. However, the comparison between two groups is represented by log rank test. The plot Figure 1 to Figure 7 for the K-M estimate of the survival function plays the role of a step function rather a smooth curve, which is between two times (times at adjacent deaths and the interval only decrease at each death). In the curve in the survival duration for the interval is represented by the lengths of the horizontal lines along the X-axis of serial times. Moreover, the cumulative probability of surviving a given time is seen on the Y-axis. In addition, the vertical distances between horizontals are important because they illustrate the change in cumulative probability. When the event of interest occurs, the interval is terminated. Some subjects are censored (patients did not die during the follow up) and they are shown as vertical bar marks in the graph; these do not terminate the interval. The figure shows the median of survival time and the survival rate. Presently, we will look at the censored subject as shown in the curve. The line of the group 1 curve ends with censored subject as seen in the plot. That provides us with a warning in terms of interpreting anything beyond this point, because the subjects might have the event (death) a few hours later. In contrast, the line of group 2 has no subjects left and the curve drops to zero after the seventeen intervals.

Table 5: The Log-Likelihoods and Akaike Information Criterion (AIC)

Distribution	Loglikelihood	K	C	AIC
Kaplan Meire	-256.5	14	1	589.0159
Cox model	-253.5	14	2	585.0454

We compared by using statistical criteria (Maximum likelihood (ML) test and AIC). According to these criteria, with the AIC (the smaller AIC is better) and higher log likelihood value. The Figure 1 and Figure 2 present Kaplan-Meier curves that show how the risk of death at the prevalence of age is distributed across the categories of a given covariates. However, it is not possible to include all the Kaplan-Meier plots in this thesis for all the covariates included in the study but by fitting a univariate Cox proportional hazard model, the hazard ratios have enough information to give about the distribution of the hazard on a given covariate.

5.2 DISCUSSION OF RESULTS

Several statistical models have been suggested for analyzing special types of data, which is referred to as censored data in the survival analysis literature. Non-parametric, semi-parametric, and parametric survival models are mainly used in many clinical trials. These models direct the form of the conditional hazard function for a given set of variables of the survival time.

The Cox proportional hazards model makes several assumptions. Thus, it is important to assess whether a fitted Cox regression model adequately describes the data.

Here, we discuss three types of diagnostics for the Cox model:

- Detecting non-linearity in relationship between the log-hazard and the covariates.
- Examining influential observations (or outliers).

- Testing the proportional hazards assumption.

The Kaplan-Meier method gives very good estimations of survival probabilities. The pattern of this method in assuming on censoring is independent of the survival time as shown in Figure 7. Each group of tumors has a pattern independent of the survival time [28, 30]. The present study has demonstrated that the patients with a primary tumor have a lesser risk than those with metastatic considering the latter have the spread of cancer cells in the body. Therefore, their survival time will be decreased. The results have provided curve of K-M and the table of the survival time which may be useful in comparing the survival time of each group and checking the censored data. The survival time for almost 10 years is 0 for metastatic tumors while the survival time for those with primary tumors is 0.56 which is evidence of survival. The K-M graph displays the cumulative survival function on a linear scale by tumor (Figure 7). The survival curve of primary tumor patients was lower than that of metastatic tumor patients, which means that primary have a higher probability of surviving (not experiencing an event) [31]. Table 2 presented the calculations from the log-rank test to show that there is significant evidence of difference in survival times for groups (primary, and metastatic) since the p-value is less than 0.05. That means there is no significant relation between the survival times of each group of tumors. The most popular method of examining the effect of explanatory variables on survival is the Cox PH model [25]. This model requires the assumption of proportional hazards between strata formed by the combinations of levels of the different explanatory variables [26, 27]. Hence, we found the model that only includes the four significant variables, which was chosen with p-value and AIC criteria. Additionally, we can conclude that the Cox model was performed to evaluate the joint prognostic significance factors. The proportional hazards (PH) assumption can be checked using statistical tests and graphical diagnostics based on the scaled Schoenfeld residuals. In principle, the Schoenfeld residuals are independent of time. A plot that shows a non-random pattern against time is evidence of violation of the PH assumption [28, 29].

6 CONCLUSIONS

Survival analysis in lung cancer studies in Nigeria provides valuable insights into the factors influencing prognosis and treatment outcomes in this population. These studies underscore the need for comprehensive approaches addressing socioeconomic disparities, access to healthcare, and the role of ethnicity and genetics in lung cancer survival. Further research is warranted to improve early detection, treatment strategies, and support systems for lung cancer patients in Nigeria.

7 RECOMMENDATIONS

The community should be educated to appreciate modern health care and influenced to abandon, or at least modify some of their harmful traditional practices. Vital basic services should be made available and accessible to members of the communities. The government should carry out an in-depth study on the effect of parent's education on infant to health and survival. The research study recommends that a greater focus along the lung cancer care pathway in Nigeria, with emphases on improving access to early diagnosis at early age. The study suggests that the Ministry of Health and

Social Services should draft up the country's first national policy on lung cancer diagnosis and management, because at the moment there is none. The medical health record of the Specialist Hospital Damaturu, should enhance their record system, hence making available of relevant information on maternal mortality, so that subsequent researcher will collect more information (data) than the one used in this research work. A doctor should try to reduce the value of ph.ecog in patients by providing more relevant medicines. The recommendations of this study on improving uptake of screening and promoting information dissemination on lung cancer should also go a long way in significantly improving the efficient and effective health management of lung cancer at all stages.

8 ACKNOWLEDGEMENTS

The authors are very grateful to Yobe State University, Damaturu for sponsoring the fund together with the referee for suggestions and comments which significantly improved the quality of the manuscript.

REFERENCES

[1] J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray. Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, vol. 149, no. 4, pp. 778-789. 2021. DOI: 10.1002/ijc.33588. Epub ahead of print. PMID: 33818764.

[2] G. A. O. Magoha, and Z. W. W. Ngumi. Cancer of the penis at Kenyatta National Hospital. *East African Medical Journal*, vol. 77, no. 10, pp. 2000. DOI:10.4314/eamj.v77i10.46706

[3] F. A. Olopade, O.A. Awolude and A.O. Adisa. Survival Outcomes and Prognostic Factors in Lung Cancer Patients in Nigeria. *Journal of Global Oncology*, vol. 4, pp. 1-9. DOI: 10.1200/JGO.18.00112. 2018. DOI: 10.4103/lungindia.lungindia_408_21. PMID: 35259791; PMCID: PMC9053916.

[4] T. M. Akande, R. A. Arogundade and T. S. Akinwande. Treatment Modalities and Survival Outcomes in Lung Cancer Patients in Nigeria. *Nigerian Journal of Clinical Practice*, vol. 23, no. 8, pp. 1059-1065. DOI: 10.4103/njcp.njcp_191_20. 2020.

[5] A. O. Adeyinka, O.O. Ogunleye and O. Salako. Ethnic and Genetic Factors in Survival Analysis of Lung Cancer Patients in Nigeria. *African Journal of Medicine and Medical Sciences*, vol. 48, no. 4, pp. 419-427, 2019.

[6] E. R. Ezeome, O. Lawal and O. Ohuche. Socioeconomic Disparities in Survival Outcomes Among Lung Cancer Patients in Nigeria: A Population-Based Study. *Nigerian Journal of Clinical Practice*, vol. 20, no. 9, pp. 1137-1142. DOI: 10.4103/njcp.njcp_112_17. 2017.

[7] M. S. Litwin, D. J. Pasta, J. Yu, M. L. Stoddard, and S. C. Flanders. Urinary Function and Bother After Radical Prostatectomy or Radiation For Prostate Cancer: A Longitudinal, Multivariate Quality Of Life Analysis From The Cancer Of The Prostate Strategic Urologic Research

Endeavor. *The Journal of Urology*, vol. 164, no. 6, pp. 1973-1977, 2000. DOI: 10.1016/s0022-5347(05)66931-5. PMID: 11061894.

[8] V. Vinh-Hung, C. Verschraegen, D. I. Promish, G. Cserni, J. Van de Steene, P. Tai and M. Royce. Ratios of involved nodes in early breast cancer. *Breast Cancer Research*, vol. 6, pp. 1-9. 2004. DOI: 10.1186/bcr934. Epub 2004 Oct 6. PMID: 15535850; PMCID: PMC1064081.

[9] M. E. Ray, K. Bae, M. H. Hussain, G. E. Hanks, W. U. Shipley and H. M. Sandler. Potential surrogate endpoints for prostate cancer survival: analysis of a phase III randomized trial. *JNCI: Journal of the National Cancer Institute*, vol. 101, no. 4, pp. 228-236, 2009. DOI: 10.1093/jnci/djn489. Epub 2009 Feb 10. PMID: 19211454; PMCID: PMC2734115.

[10] Y. M. Chan. Statistical Analysis and Modeling of Prostate Cancer (Doctoral dissertation, University of South Florida). 2019.

[11] A. Amaro, A. I. Esposito, A. Gallina, M. Nees, G. Angelini, A. Albini, and U. Pfeffer. Validation of proposed prostate cancer biomarkers with gene expression data: a long road to travel. *Cancer and Metastasis Reviews*, pp. 1-15, 2014. DOI: 10.1007/s10555-013-9470-4. PMID: 24477410; PMCID: PMC4113682.

[12] O. S. Balogun, M. R. Role, and O. O. Dawodu. Survival Analysis of lung Cancer in Ilorin, Kwara State. *Survival*, vol. 4, no. 06, 2014.

[13] O. Aalen, O. Borgan and H. K. Gjessing. Nonparametric analysis of survival and event history data, *Survival and Event History Analysis*, pp. 69–130, 2008.

[14] D. Collett, (2023). *Modelling survival data in medical research*. Chapman and Hall/CRC.

[15] D. W. Hosmer Jr, S. Lemeshow and R. X. Sturdivant. *Applied logistic regression*. John Wiley and Sons. 2013.

[16] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, *53*(282), 457-481.1958. https://doi.org/10.1080/01621459.1958.10501452

[17] J. T. Rich, J. G. Neely, R. C. Paniello, C. C. Voelker, B. Nussenbaum and E. W. Wang. A practical guide to understanding Kaplan-Meier curves. *Otolaryngology—Head and Neck Surgery*, vol. 143, no. 3, pp. 331-336, 2010. DOI: 10.1016/j.otohns.2010.05.007. PMID: 20723767; PMCID: PMC3932959.

[18] V. Bewick, L. Cheek and J. Ball. Statistics review 12: survival analysis. Critical Care-London-, 8, 389-394. 2004.

[19] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187-202, 1972. DOI: https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

[20] F. E. Harrell Jr, and F. E. Harrell. Cox proportional hazards regression model. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis, pp. 475-519, 2015.

[21] X. Xue, X. Xie, M. Gunter, T. E. Rohan, S. Wassertheil-Smoller, G. Y. Ho and H. D. Strickler. Testing the proportional hazards assumption in case-cohort analysis. BMC medical research methodology, vol. 13, pp. 1-10, 2013. DOI: https://doi.org/10.1186/1471-2288-13-88.

[22] R. I. Jennrich and S. M. Robinson. A Newton-Raphson algorithm for maximum likelihood factor analysis. *Psychometrika*, vol. 34, no. 1, pp. 111-123, 1969. DOI: https://doi.org/10.1007/BF02290176.

[23] W. W. M. Abeysekera and M. R. Sooriyarachchi. Use of Schoenfeld's global test to test the proportional hazards assumption in the Cox proportional hazards model: an application to a clinical study. 2009. DOI: 10.4038/jnsfsr.v37i1.456.

[24] D. G. Kleinbaum, K. Dietz, M. Gail and M. Klein. *Logistic regression* (p. 536). New York: Springer-Verlag. 2002.

[25] D. A. Schoenfeld. Sample-size formula for the proportional-hazards regression model. *Biometrics*, pp. 499-503. 1983. DOI: https://doi.org/10.2307/2531021

[26] P. M. Grambsch and T. M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, vol. 81, no. 3, pp. 515-526, 1994. DOI: https://doi.org/10.1093/biomet/81.3.515

[27] T. G. Clark, M. J. Love, S. B. Bradburn, and D. G. Altman. Survival analysis part I: basic concepts and first analyses. *British Journal of Cancer*, vol. 89, no. 2, pp. 232, 2003. DOI: 10.1038/sj.bjc.6601118. PMID: 12865907; PMCID: PMC2394262.

[28] N. K. Dakhil, Y. M. Al–mayali, and M. A. Al–A'bidy. Analysis of Breast Cancer Data using Kaplan–Meier Survival Analysis. *Journal of Kufa for Mathematics and Computer*, vol. 1, no. 6, 2012. DOI: https://doi.org/10.31642/JoKMC/2018/010602

[29] U. Y. Madaki, B. I. Babura, M. Sani and I. Abdullahi. (2023). Cure Fraction Models on Survival Data and Covariates with Bayesian Parametric Estimation Methods. *Applied Mathematics and Computational Intelligence (AMCI)*, vol. 12, no. 1, pp. 17–29. 2023.

[30] U. Y. Madaki and M. R. B. A. Bakar. A Bayesian estimation on right censored survival data with mixture and non-mixture cured fraction model based on beta-Weibull distribution. In *AIP Conference Proceedings*, vol. 1739, no. 1, pp. 020079, AIP Publishing LLC. 2016.

[31] U. Y. Madaki and M. R. B. A. Bakar. Cure Models based on Weibull Distribution with and without Covariates using Right Censored Data. *Indian Journal of Science and Technology*, vol. 9, pp. 28, 2016.

[32] H. Akaike. Factor analysis and AIC. *Psychometrika*, 52, 317-332. 1987.