Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.755138

# ANALYSING PREDICTORS OF HEALTHCARE RESOURCE CONSUMPTION AMONG CANCER PATIENTS: A COMPREHENSIVE REGRESSION MODELING STUDY

**Dr. Mrinal Deka**
Faculty of Science, Statistics, Assam down town University, Guwahati
Email: mrinal.deka@adtu.in

**Abstract**
In this particular study, various risk factors such as socio-economic status, demographics, and clinical factors have been successfully identified for lung and oral cancer. The research indicates that patient age, tumour size, node size and blood sugar levels adversely affect cancer patient survival times, with increased values corresponding to decreased survival times. Furthermore, the study addresses the critical issue of selecting an appropriate survival model and concludes that the Weibull survival model is the most suitable option. This conclusion is drawn based on the lower Akaike Information Criterion (AIC) values obtained compared to other models across all cancer types. The findings suggest that survival times estimated from this model are reliable, facilitating predictions of cancer patient survival times based on available data.

**Key Words:** Lung Cancer, Oral Cancer, Risk factors, Parametric survival model, Bayesian Model

## Introduction

Cancer poses a significant public health challenge, causing considerable suffering and reducing the lifespan of affected individuals, a fact often underestimated (Chu et al., 2008). However, it would be imprudent to solely analyse socio-economic factors without considering demographic and clinical factors.

Cancer comprises a group of diseases characterized by abnormal tissue growth, forming tumours (Baghestani et al., 2015). The nature and consequences of cancer vary depending on its location within the body. Yet, there has been limited research into understanding this heterogeneity, which could greatly inform treatment strategies and medical management. Before delving into specific types of cancer, such as oral and lung cancer, a comprehensive analysis of their incidence is necessary (Ali et al., 2011).

Lung cancer is prevalent in men and ranks as the third most common cancer in women (Ali et al., 2011). Smoking remains the primary risk factor for lung cancer (Aggarwal et al., 2016), with approximately 7,300 nonsmokers dying from lung cancer annually due to exposure to passive smoke in the United States (Aggarwal et al., 2016). Variations in smoking behaviour, such as the type of tobacco and inhalation patterns, are cited by the World Health Organization as significant contributors to lung cancer development (Payne, 2004).

Oral cancer, a subtype of head and neck cancer, affects various areas including the throat, tongue, and gums, and is associated with tobacco use, heavy alcohol consumption, HPV infection, and weakened immune systems (Sharma et al., 2014).

Understanding the occurrence and treatment of different cancers requires an examination of various risk factors, including socio-economic, demographic, and clinical

factors. Age, for instance, is a universally recognized risk factor for cancer incidence, with most cancers becoming more prevalent with advancing age (White et al., 2014). Additionally, research indicates disparities in cancer mortality rates between genders, with men more susceptible to certain types of cancers (Kim et al., 2018). Economic factors also play a crucial role, as the cost of cancer treatment often leads to financial strain, especially for economically disadvantaged individuals, impacting their access to early detection and treatment (Nair et al., 2014).

Early detection is paramount in preventing cancer progression, yet lack of awareness about risk factors and limited access to healthcare among economically disadvantaged populations often result in delayed diagnosis (Caplan, 2014; Walter et al., 2015). Adequate emotional and financial support can facilitate early diagnosis and treatment initiation, thereby improving survival rates (Zhai et al., 2019).

Analysing disease incidence data using appropriate statistical methods yields valuable insights even in the face of uncertainty (Mohamad et al., 2007). Consequently, this study employs various survival models to investigate the impact of risk factors on cancer patient survival, aiming to identify the most suitable model for the available data and predict survival times effectively.

**Review of Literature**

Undertaking a comprehensive review of cancer research materials presents a significant challenge due to the vast amount of available literature. However, focusing on studies investigating various risk factors associated with cancer provides valuable insights into the occurrence of different cancer types (White et al., 2014). For instance, research conducted in Leicester city (England) examined the use of alcohol, tobacco, and pan among males from various Asian communities, shedding light on their knowledge and attitudes towards oral cancer risk factors and prevention (Vora et al., 2000). Similarly, tobacco consumption emerges as a significant concern in North Eastern states in India, where a substantial portion of cancer cases, particularly among males, is attributed to tobacco use (Ngaihte et al., 2019). Furthermore, incorporating healthy dietary habits, such as consuming fruits and vegetables, is shown to mitigate the risk of oral squamous cell carcinoma (Scully, 2005). Other studies explore the relationship between diabetes and oral cancer, revealing potential connections (Aiuto et al., 2017).

Research also delves into the association between nutritional factors and the survival rates of oral cancer patients (Liu et al., 2006). Additionally, comparative analyses investigate various treatment modalities, including surgical procedures, radiotherapy, and combinations thereof, to determine their efficacy (Rogers et al., 2012). Long-term survival rates and influencing factors among oral cancer patients are examined, with findings suggesting differences based on treatment approaches (Christian et al., 2015). Moreover, research endeavours compare different survival functions using non-parametric and parametric models to identify optimal approaches for analysing cancer patient outcomes (Kottabi, 2012).

While existing literature provides valuable insights, there remains a gap in studies that simultaneously consider multiple risk factors, including socio-economic, demographic, and clinical factors, especially in regions like South Assam. Additionally, predictive models for cancer patient survivability incorporating these factors are lacking for such high-risk areas. Therefore, the current study aims to address these gaps by investigating the interplay of various

risk factors and demographic information in predicting cancer patient outcomes in South Assam.

## Method and Material

The study relies on secondary data sourced from the Cachar Cancer Hospital and Research Centre, situated in Silchar, Assam, spanning from 2014 to 2019. The hospital's primary aim is to deliver high-quality treatment to its patients while maintaining affordable service costs. It's noted that 80% of the patients are employed in daily wage jobs, with 50% earning minimal wages *(https://cacharcancerhospital.org)*. The majority of hospitalized patients typically hail from various districts across Assam and other northeastern states such as Tripura, Manipur, Meghalaya, and Mizoram (Ngaihte et al., 2019 and Mathur et al., 2020).

Considering the prevalent occurrence of cancer types in the region, 386 with lung cancer and 398 with oral cancer were chosen from a total of 2604 cancer patients.

The present study aims to explore the correlation between socio-economic, demographic and medical factors and the survival rates of lung and oral cancer patients. Econo-demographic factors such as age, gender, marital status, religion, and consumption habits are considered alongside medical factors like tumour size, node size, blood sugar level, and treatment techniques. These factors are treated as independent variables (risk factors), while the survival time of cancer patients serves as the dependent variable in constructing the model.

Before formally integrating these factors into survival models, it is essential to assess whether some independent variables play redundant roles, potentially affecting the reliability of the models. Therefore, correlations are computed among both the dependent and independent variables, as well as among the independent variables themselves. Point Bi-serial correlation and Karl Pearson correlation are utilized for categorical and continuous variables respectively. Additionally, Collinearity Diagnostics are conducted to gauge the degree of multicollinearity within the model.

Parametric survival models including the Exponential, Weibull, and Gaussian models are applied to analyse the effects of various factors on the survival times of lung and oral cancer patients. The selection of the appropriate survival model for the dataset is determined using the Akaike Information Criterion (AIC), where a lower AIC value indicates a better-fitted model for the data.

Furthermore, Bayesian regression modeling is employed to validate the estimates obtained from the selected survival model. Bayesian linear regression, stemming from the Bayesian approach, treats uncertainty as probability, contrasting with the frequentist approach. This approach integrates prior knowledge about parameters before observing the data, using prior distributions. The posterior distribution, representing the updated beliefs after observing the data, is obtained using the likelihood function. Markov Chain Monte Carlo (MCMC) techniques are commonly employed to approximate the posterior distribution. Mathematically, the posterior distribution can be defined as the distribution of unknown parameters given the observed data.

$$Posterior \propto Likelihood \times Prior \ldots (1)$$

We can also write equation (1) as

$$P(\theta|D) \propto P(D|\theta) \times P(\theta) \ldots (2)$$

Thus, based on Bayes theorem, we can write the following equation

$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{\sum P(D|\theta) \times P(\theta)} \dots (3)$$

It is more convenient if we write the equation (3) as follows

$$P(Model|New_{Data}) = \frac{P(New_{Data}|Model) \times P(Model)}{P(New_{Data})} \dots (4)$$

In this study, a non-informative prior was employed due to a lack of sufficient knowledge about the parameters, thereby considering them neutral.

$$\beta_j \sim N(\mu_j, \sigma_j^2) \dots (5)$$

In our analysis, we utilized the 'brms' package, version 4.0.2, in R to derive posterior distributions for Bayesian Regression Models using Stan.

Our primary objective was to forecast survival times employing appropriate survival models. To achieve this, we adopted a holdout sample technique, segregating 30 samples as holdout samples for each cancer site (Lung, and Oral cancer), while the remaining samples served as training data. Subsequently, we computed estimated survival times from the holdout samples, distinctively for each cancer site, and juxtaposed them with actual survival times. Furthermore, we calculated the Mean Square Error (MSE) to gauge the average squared difference between the estimated and actual survival times, providing insight into the predictive performance of our models.

**Result**

Initially, correlations were computed to assess redundancy among the dependent and independent variables, resulting in the construction of (Table 1). While prior literature indicates the importance of all variables as potential risk factors, this study will focus solely on Age, Tumour size, Node size, and Blood Sugar Level, as they exhibit significant correlations with the survival time of patients for each cancer site (as shown in Table 1). Other factors will be omitted from the model, as their correlations with the dependent variable are not statistically significant, suggesting potential redundancy in the model (Uyanik et al., 2013).

**Table 1: Correlation between dependent and independent variables varies across three distinct cancer sites**

| Type of cancer | Factor | Correlation Coefficient | *p* value |
|---|---|---|---|
| Lung cancer | Surv_Time and Age | -.060 | .029 |
| | Surv_Time and Tumour Size | -.225 | .001 |
| | Surv_Time and Node Size | -.502 | .035 |
| | Surv_Time and Blood Sugar Level | -.181 | .001 |
| | Surv_Time and Gender | -.856 | .089 |
| | Surv_Time and Marital Status | .789 | .074 |
| | Surv_Time and Religion | .653 | .096 |

| Type of cancer | Factor | Correlation Coefficient | *p* value |
|---|---|---|---|
| | Surv_Time and Consumption Habit | -.589 | .085 |
| | Surv_Time and Treatment | .489 | .088 |
| | Surv_Time and Economic Status | .659 | .093 |
| Oral cancer | Surv_Time and Age | -.016 | .005 |
| | Surv_Time and Tumour Size | -.220 | .001 |
| | Surv_Time and Node Size | -.530 | .001 |
| | Surv_Time and Blood Sugar Level | -.159 | .001 |
| | Surv_Time and Gender | -.523 | .062 |
| | Surv_Time and Marital Status | .632 | .095 |
| | Surv_Time and Religion | .854 | .091 |
| | Surv_Time and Consumption Habit | -.362 | .085 |
| | Surv_Time and Treatment | .259 | .069 |
| | Surv_Time and Economic Status | -.658 | .086 |

Again, we assess the correlation among the chosen independent variables to identify and eliminate redundancy within the regression model, resulting in the construction of the subsequent table.

**Table 2: Correlations between the selected independent variables across two distinct cancer sites**

| Type of cancer | Independent Variable | Correlation Coefficient | *p* value |
|---|---|---|---|
| Lung cancer | Age and Tumour size | .014 | .633 |
| | Age and Blood sugar level | -.053 | .125 |
| | Age and Node size | .012 | .512 |
| | Tumour size and Blood sugar level | -.705 | .214 |
| | Tumour size and Node size | .818 | .255 |
| | Node size and Blood sugar level | -.745 | .214 |

| | | | |
|---|---|---|---|
| Oral cancer | Age and Tumour size | .026 | .373 |
| | Age and Blood sugar level | -.112 | .214 |
| | Age and Node size | .028 | .248 |
| | Tumour size and Blood sugar level | -.699 | .258 |
| | Tumour size and Node size | .809 | .366 |
| | Node size and Blood sugar level | -.746 | .198 |

The correlations between the selected independent variables for each cancer site as displayed in Table 2, indicate non-significance, with p-values exceeding 0.05. Therefore, for this study, we consider Age, Tumor Size, Node Size, and Blood Sugar level as the independent variables. Collinearity diagnostics between these variables were also conducted, yielding the following table.

**Table 3: Collinearity diagnostics for the chosen independent variables across various cancer sites**

| Type of Cancer | Independent Variable | VIF value | Condition Index (CI) |
|---|---|---|---|
| Lung Cancer | Age | 1.589 | 9.620 |
| | Tumour size | 2.653 | 8.962 |
| | Node size | 1.960 | 3.520 |
| | Blood sugar level | 1.025 | 4.587 |
| Oral Cancer | Age | 1.569 | 5.201 |
| | Tumour size | 2.853 | 6.321 |
| | Node size | 2.330 | 8.951 |
| | Blood sugar level | 1.623 | 8.741 |

The table (Table 3) presents collinearity diagnostics for the chosen independent variables across various cancer sites, indicating that multicollinearity is effectively managed, with Variance Inflation Factor (VIF) values below 5 and Condition Indices below 15. Subsequently, parametric survival models such as the 'Exponential,' 'Weibull,' and 'Gaussian' models were employed to evaluate the impact of factors on the survival times of lung and oral cancer patients. To determine the most suitable survival model for the dataset, the Akaike Information Criterion (AIC) was utilized, where a lower AIC value indicates a better-fitted model for the specific data (Kleinbaum and Klein, 2012). The following table is constructed to aid in this assessment.

**Table 4: Calculated AIC values for the three parametric survival models**

| Type of Cancer | Survival Model | AIC Value |
|---|---|---|
| Lung Cancer | Weibull Model | 2664.338 |
| | Exponential Model | 3293.062 |
| | Gaussian Model | 2711.861 |

| Type of Cancer | Survival Model | AIC Value |
|---|---|---|
| Oral Cancer | Weibull Model | 2616.320 |
| | Exponential Model | 3237.275 |
| | Gaussian Model | 2661.897 |

Table 4 indicates that the Weibull survival model is the most suitable for this study, as it yields lower AIC values compared to other survival models across all types of cancer. Consequently, we solely focus on the regression estimates derived from the Weibull survival model.

**Table 5: Outcomes derived from the Weibull survival model for various cancer sites**

| Type of Cancer | Factor | Estimate | Standard Error | p value | 95 % Credible Interval | |
|---|---|---|---|---|---|---|
| | | | | | LCL | UCL |
| Lung Cancer | Intercept | 5.2474 | 0.1065 | .029 | 5.0387 | 5.4562 |
| | Age | -0.0021 | 0.0010 | .001 | -0.0042 | 0.0209 |
| | Tumour Size | -0.0063 | 0.0159 | .035 | -0.0375 | 0.3377 |
| | Node Size | -0.0082 | 0.0247 | .001 | -0.0089 | 0.0099 |
| | Blood Sugar Level | -0.0143 | 0.0008 | .001 | -0.0156 | 0.1762 |
| Oral Cancer | Intercept | 5.3265 | 0.1080 | .005 | 5.1147 | 5.5384 |
| | Age | -0.0044 | 0.0011 | .001 | -0.0066 | -0.0022 |
| | Tumour Size | -0.0155 | 0.0157 | .001 | -0.0464 | 0.0154 |
| | Node Size | -0.0235 | 0.0185 | .002 | -0.0255 | 0.0102 |
| | Blood Sugar Level | -0.0136 | 0.0007 | .003 | -0.0150 | -0.0123 |

Based on the findings outlined in Table 5, we have identified independent variables that wield a significant influence on the dependent variable, with p-values below 0.05. Across all 'Types of Cancer,' it's noteworthy that all independent variables exhibit significance. Notably, the coefficients associated with the independent variables—patient age, tumour size, node size, and blood sugar level—are consistently negative. This suggests a negative impact on the survival time of lung and oral cancer patients. In essence, as patient age, tumour size, node size and blood sugar level increase, the survival times of cancer patients tend to decrease.

Furthermore, to bolster the reliability of our estimates derived from the Weibull survival model, we have employed a Bayesian regression model. This additional analysis aims to validate the robustness of our findings, thereby providing further credibility to our results. The results of this Bayesian regression model are summarized in the subsequent table.

**Table 6: Estimates derived from both the Weibull survival model and Bayesian regression model across various cancer sites**

| Type of Cancer | Factor | Estimate (Obtained from Weibull model) | Estimate (Obtained from Bayesian model) |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Lung Cancer | Intercept | 5.2474 | 5.1400 |
| | Age | -0.0021 | -0.0018 |
| | Tumour Size | -0.0063 | -0.0069 |
| | Node Size | -0.0082 | -0.0078 |
| | Blood Sugar Level | -0.0143 | -0.0102 |
| Oral Cancer | Intercept | 5.3265 | 5.2200 |
| | Age | -0.0044 | -0.0039 |
| | Tumour Size | -0.0155 | -0.0159 |
| | Node Size | -0.0235 | -0.0243 |
| | Blood Sugar Level | -0.0136 | -0.0129 |

After obtaining parameter estimates from both models, as depicted in Table 6, we note that the estimates are nearly identical. Consequently, our confidence in the unknown parameter estimates for each cancer site is bolstered by the new knowledge acquired from observed data. Our primary objective is to forecast survival times using the Weibull survival model. To accomplish this, we employ the holdout sample technique, setting aside 30 samples for each cancer site (Lung and Oral cancer) while utilizing the remainder as training samples. We then compute estimated survival times for the holdout samples of each cancer site, comparing them with the actual survival times. Additionally, we calculate the Mean Square Error (MSE) to gauge the average squared difference between the estimated and actual survival times. The subsequent figures illustrate the comparison of estimated and actual survival times for Lung and Oral cancer.

**Fig. 1: Comparison of estimated and actual survival times for Lung cancer**

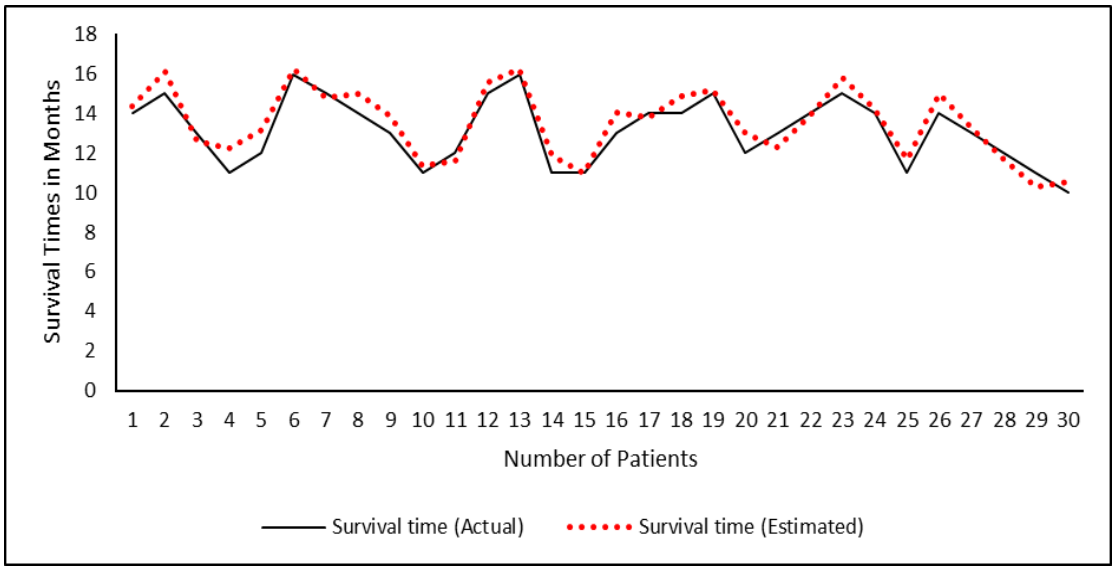


Figure 1 indicates that there is negligible disparity between the actual and estimated survival times for lung cancer (mean squared error = 0.4698). Consequently, we can infer that the estimated survival times derived from the model are dependable, and thus, the model holds promise for predicting survival times for lung cancer patients using the available dataset.

**Fig. 2: Comparison of estimated and actual survival times for Oral cancer**

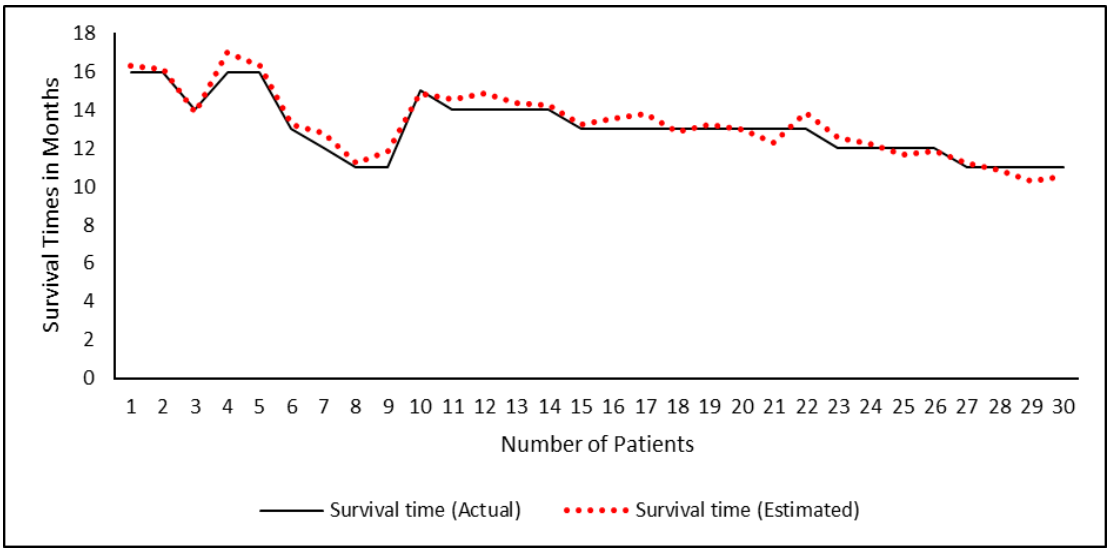


Figure 2 yields analogous outcomes, indicating no discernible variance between actual and estimated survival durations for oral cancer (MSE=0.2487). Hence, it is deduced that the model's estimated survival times are dependable, affirming its viability for predicting survival times of oral cancer patients using the provided dataset.

**Conclusion**

The research has effectively identified associations between various risk factors, including socio-economic, demographic, and clinical factors, with lung, and oral cancer. Our findings reveal that patient age, tumour size, node size and blood sugar levels have a detrimental impact on cancer patient survival times. Specifically, as these factors increase, survival times decrease. Furthermore, the study addresses the critical question of selecting an appropriate survival model for our dataset. We determine that the Weibull survival model is the most suitable option, as it yields a lower AIC value compared to other models across all cancer types. Additionally, we assert that the estimated survival times derived from this model are dependable and can be utilized for predicting survival times among cancer patients based on the available data.

**Reference**

Abdelaal M. A, A. M. (2020). Estimation of Out-of-Field Dose Variation using Markus Ionization Chamber Detector. *SciMedicine Journal*, 8-15.

Aggarwal A, L. G. (2016). The State of Kung Cancer Research: A Global Analysis. *Journal of Thoracic Oncology*, 1040-1050.

Aiuto D. F, D. G. (2017). Evidence Summary: The Relationship between Oral Diseases and Diabetes. *British Dental Journal*, 944-948.

Ali I, W. W. (2011). Cancer Scenario in India with Future Perspectives. *Cancer Therapy*, 50-70.

Baghestani A, M. S. (2015). Survival Analysis of Patients with Breast Cancer using Weibull Parametric Model. *Asian Pacific Journal of Cancer Prevention*, 8567-8571.

Baskar R, L. A. (2012). Cancer and Radiation Therapy: Current Advances and Future Directions. *International Journal of Medical Sciences*, 193-199.

Caplan L. (2014). Delay in Breast Cancer: Implications for Stage at Diagnosis and Survival. *Frontiers in Public Health*, 1-5.

Carter C. L, A. C. (1989). Relation of Tumour Size, Lymph Node Status and Survival in 24740 Breast Cancer Cases. *Cancer*, 181-187.

Christian D, N. C. (2015). Survival Analysis of Dental Implants and Implant-Retained Prostheses in Oral Cancer Patients up to 20 Years. *Clin Oral Invest*, 1347-1352.

Chu P, W. J. (2008). Estimation of Life Expectancy and the Expected Years of Life Lost in Patients with Major Cancers: Extrapolation of Survival Curves under High-Censored Rates. *Value in Health*, 1102-1109.

Daoud I. J. (2017). Multicollinearity and Regression Analysis. *IOP Conf. Series: Journal of Physics: Conf. Series 949 012009* (pp. 1-6). Malaysia: IOP Publishing.

Fidarova E. (2020). *Cancer*. Retrieved from World Health Organization: www.who.int/health-topics/cancer

Francis R. E. (2017). The Application of Binary Logistic Regression Analysis on Staff Performance Appraisal. *Science Journal of Applied Mathematics and Statistics*, 164-168.

Hoegg T. (2013). *Statistical Modelling of Breast Cancer Risk for British Columbian Women.* Columbia: University of British Columbia.

Jim S. L, P. J. (2015). Religion, Spirituality and Physical Health in Cancer Patients: A Meta-Analysis. *Cancer*, 3760–3768.

Kim I. H, H. L. (2018). Sex Differences in Cancer: Epidemiology, Genetics and Therapy. *Biomolecules & Therapeutics*, 335-342.

Kleinbaum D.G, K. M. (2012). *Survival Analysis, A Self-Learning Text* (3rd ed.). Springer.

Koehrsen W. (2018). *Introduction to Bayesian Linear Regression; An Explanation of the Bayesian Approach to Linear Modeling*. Retrieved from towards data science: https://towardsdatascience.com/introduction-to-bayesian-linear-regression-e66e60791ea7

Kottabdi Z. (2012). *Statistical Modeling and Analysis of Breast Cancer and Pancreatic Cancer.* Florida: University of South Florida.

Kroman N, J. B. (2000). Factors Influencing the Effect of Age on Prognosis in Breast Cancer: Population Based Study. *British Medical Journal*, 474-479.

Liu S, T. W. (2006). Nutritional Factors and Survival of Patients with Oral Cancer. *Journal of the Sciences and Specialties of the Head and Neck*, 998-1007.

Langovain K. (2008). Survival Analysis for Clinical Studies. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.*, 303-307.

Mathur P, S. K. (2020). Cancer Statistics, 2020: Report From National Cancer Registry Programme, India. *American Society of Clinical Oncology*, 1063-1075.

Mohamad P, E. H. (2007). Comparing Cox Regression and Parametric Models for Survival of Patients with Gastric Carcinoma. *Asian Journal of Cancer Prevention*, 412-416.

Nair S. K, R. S. (2014). Cost of Treatment for Cancer : Experiences of Patients in Public Hospitals in India. *Asian Pacific Journal of Cancer Prevention*, 5049-5054.

Ngaihte P, E. Z. (2019). Cancer in the NorthEast India: Where We are and What Needs to Be Done? *Indian Journal of Public Health*, 251-253.

Payne S. (2004). *Gender in Lung Cancer and Smoking Research.* World Health Organization.

Rogers N. S, B. C. (2012). The Prosthodonic Pathway of the Oral Cancer Patient. *Dental Update*, 98-106.

Scully C. (2005). Oral Cancer; The Evidence for Sexual Transmission. *British Dental Journal*, 203-207.

Sharma D. J, K. M. (2014). Cancer: Scenario and Relationship of Different Geographical Areas of the Globe with Special Reference to North East-India. *Asian Pacific Journal of Cancer Prevention*, 3721-3729.

Sun S. Y, Z. Z. (2017). Risk Factors and Preventions of Breast Cancer. *International Journal of Biological Sciences*, 1387-1397.

Uyanik K. G, G. N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, 234-240.

Uysal E. (2017). Top 100 Cited Classic Articles in Breast Cancer Research. *European Journal of Breast Health*, 129-137.

Vora R. A, C. M. (2000). Alcohol, Tobacco and Paan Use and Understanding of Oral Cancer Risk among Asian Males in Leicester. *British Dental Journal*, 444–451.

Walter M. F, G. R. (2015). Symptoms and other Factors Associated with Time to Diagnosis and Stage of Lung Cancer: A Prospective Cohort Study. *British Journal of Cancer*, 6-13.

Wang W, Y. L. (2017). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conf. Series: Materials Sci. Eng. 2018;324(1).* (pp. 1-10). Shanghai, China: IOP Publishing.

White C. M, D. M. (2014). Age and Cancer Risk: A Potentially Modifiable Relationship. *American Journal of Preventive Medicine*, 7-15.

Yavari K. (2018). Anti-Angiogenesis Therapy of Cancer Cells using 153Sm-Bevasesomab. *Emerging Science Journal*, 130-139.

Yeole B. B, K. A. (2003). An Epidemiological Assessment of Increasing Incidence and Trends in Breast Cancer in Mumbai and Other Sites in India, During the Last Two Decades. *Asian Pacific Journal of Cancer Prevention*, 51-56.

Zare H. (2019). Effects of Salvia Officinalis Extract on the Breast Cancer Cell Line. *SciMedicine Journal*, 25-29.

Zhai Z, F. Z. (2019). Effects of Marital Status on Breast Cancer Survival by Age, Race and Hormone Receptor Status: A Population-Based Study. *Cancer Medicine*, 4906–4917.