Analysis of Lung Cancer Data – Investigation of Eastern Cooperative Oncology Group (ECOG) performance score as a predictor for survival lung cancer patients

Table of Contents

Abstract

Introduction: Lung cancer is the leading cause of cancer-related deaths in many of the developed countries. One of the reasons lung cancer is at the top of the list is that it is often not diagnosed until the cancer is at an advanced stage. Thus, the earliest diagnosis of lung cancer is crucial to ensure earlier treatment. Data containing subjects with advanced lung cancer from the North Central Cancer Treatment Group was thereby analyzed to investigate if the Eastern Cooperative Oncology Group (ECOG) performance score is a predictor for survival lung cancer patients.

Method: The data was analyzed using R programming language. The descriptive statistics of the population characteristics was carried out and the influence of the several features of the dataset in predicting survival of lung cancer was explored using Cox regression analysis, with major focus on ECOG score. Univariate Cox regression analysis was first explored to identify significant features. The significant features were then supplied to the multivariate Cox regression model to describe how the factors jointly impact on survival. The survival plot of the patients over time was plotted via Kaplan-Meier (K-M) survival curves while the survival curves patients based on sex and ECOG were compared.

Results: From the study, the Cox regression model revealed sex and ECOG as the only significant features for predicting the survival of lung cancer with a p-value of less than 0.05. The p-value for ECOG score is 6.56e-06, with a hazard ratio HR = 1.53, indicating a strong relationship between the ECOG value and increased risk of death. Holding the other covariates constant, a higher value of ECOG score is associated with a poor survival.

Introduction

Background of the Study

Lung cancer is a widespread and lethal malignancy impacting the lungs, disrupting their essential role in oxygen and carbon dioxide exchange. Globally, it is a leading cause of cancer-related deaths, with geographic variations in incidence linked to smoking prevalence. Tobacco smoke, a primary risk factor, significantly heightens the likelihood, with risks escalating based on smoking duration and intensity. Secondhand smoke, occupational exposures, and genetic factors also contribute to lung cancer development. Symptoms often emerge in later stages, encompassing cough, chest pain, shortness of breath, wheezing, weight loss, and fatigue. Prognosis hinges on factors such as cancer stage, overall health, and cancer type, emphasizing the importance of early detection for successful treatment.

In addressing the global health concern of lung cancer, a study analyzes data from the North Central Cancer Treatment Group (NCCTG). The focus is on exploring the predictive role of Eastern Cooperative Oncology Group (ECOG) performance status in lung cancer patient survival. ECOG status, reflecting a patient's daily activity capability, is recognized as a potential prognostic factor in various cancers, including lung cancer. The primary goal is to investigate the reliability of ECOG performance status as a predictor for lung cancer patient survival. This analysis aims to provide critical insights into disease progression, potentially guiding clinicians in making well-informed decisions about patient care.

Research Questions

1. What is the predicting effect of ECOG performance score on survival of lung cancer?
2. Are there any other potential confounding features that influence the survival of patients with lung cancer.

Methods

Study Design

The population used in the study contains 300 lung cancer patients documented in the NCCTG database from a specified time period. The inclusion criteria for the study are patients diagnosed with lung cancer, availability of complete data on ECOG performance status and adequate information on patient demographics, clinical characteristics, and survival outcomes. The dataset measured nine features which include age, sex, institution, status, time, ph.ecog, ph.karno, pat.karno, and meal.cal. The description of these features was provided in appendix I.

Participants

The participants used in this study include 300 patients diagnosed with advanced lung cancer.

Outcome

The outcome of the study revealed that ECOG performance score is a significant predictor of survival of lung cancer ($p < 0.05$). The Cox regression model fitted on survival time using ECOG score as a predictor variable produced a hazard ratio (HR) greater than 1. This indicated that higher value of ECOG score is associated with a poor survival from lung cancer.

Covariates

ECOG performance score is the main independent variable which served as the predictor variable for survival time of patients. However, sex and age were found to be potential confounding variables. Further analysis with multivariate Cox regression revealed only sex as a significant covariate which has impact on survival.

Statistical Analysis Methods

The study utilized various statistical analysis methods. Descriptive statistics was used to characterize study participants, employing frequency counts and percentages for categorical variables and median, mean, IQR, standard deviation, minimum, and maximum for continuous variables.

Survival analysis, focusing on time-dependent survival rates, employed the Kaplan-Meier curve to illustrate the survival function. Cox proportional-hazards model, or Cox regression analysis (Cox, 1972) was employed to investigate the association between patient survival time and predictor variables. The hazard function (h(t)) in the Cox model represents the risk of dying at time 't' and is expressed as

$$h(t) = h_0(t) * \exp\left(b_1x_1 + b_2x_2 + \cdots + b_px_p\right)$$

where 't' is survival time, 'h(t)' is the hazard function, 'xi' are covariates, 'bi' are coefficients, and 'h_0' is the baseline hazard. The analysis aims to assess the impact of covariates on survival risk over time.

Results

A. **Demographic and Clinical Characteristics of the Study Participants**

The study population was characterized using descriptive statistics. The participants characteristics were divided into quantitative characteristics and categorical characteristics as shown in table 1a and 1b respectively. Table 1a revealed the mean, median and other statistics for age, ph.karno, pat.karno and meal.cal. It was also observed that there are missing values for ph.karno, pat.karno and meal.cal. The patients represented in the study are between 39 and 82 years of age with a mean age of 60.

Table 1a. Summary Statistics for Quantitative Factors

Variable	N	Median	IQR	Mean	SD	Min	Max
Age	300	61	53 – 68	60	9.5	39	82
ph.karno	299	80	75 – 90	82.1	12.3	50	100
pat.karno	297	80	70 – 90	80	14.6	30	100
meal.cal	253	975	675 – 1150	922.6	371.8	96	2600

From table 1b, it was observed that there are more male patients (59.33%) represented in the study than female patients (40.67%). The frequency count of ECOG categories revealed that a large number of the cancer patients are symptomatic but completely ambulatory. This is followed by asymptomatic patients with a percentage of 26.33% and patients that use less than 50 days in bed (21.67%). Only 3 patients (0.01%) spend more than 50 days in bed.

Table 1b. Summary Statistics for Categorical Factors

Variables	Category	Count	Percentage
Sex	Male	178	59.33%
	Female	122	40.67%
	Missing	0	0.0%
ECOG	0 – Asymptomatic	79	26.33%
	1 – Ambulatory	152	50.67%
	2 - <50 days in bed	65	21.67%
	3 - >50 days in bed	3	0.01%
	Missing	1	0.003%

B. **Trend of Survival Time for the Study Population**

Kaplan-Meier curve was generated to reveal the patients' overall survival with lung cancer as shown in Figure 1. The survival plot revealed the median survival of the overall cohort is 329 days (95% CI: 301, 353), which means 50% of the cohort only survived approximately 10 months and few days. The 1-year survival probability is 40.2% (95% CI: 34%, 47%) which indicated that only 40.2% of the patients survived past 1 year. The survival of patients over time based on their sex and ECOG scores were also explored using K-M plots and presented in appendix II.

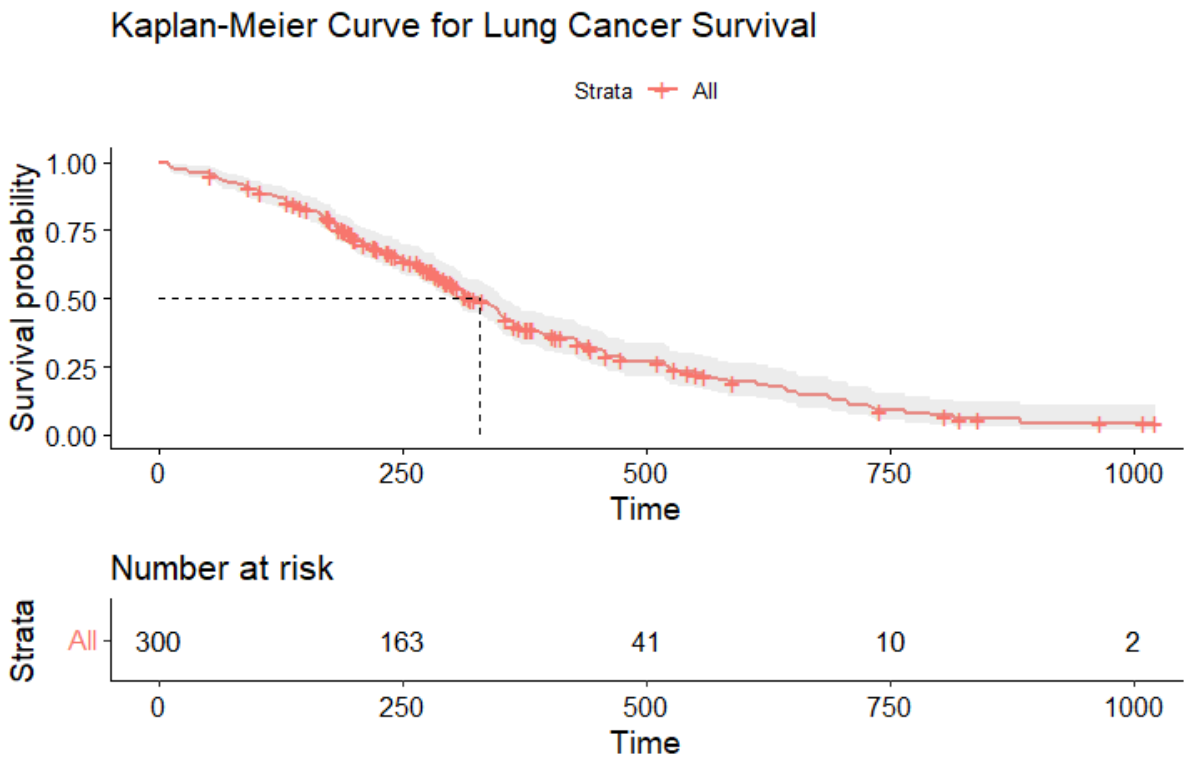

Fig. 1: Kaplan-Meier curve showing the survival of the 300 patients with a median line.

C. **Univariate Cox regression**

Univariate Cox regression was performed on each of the covariates to analyze the influence of each variable on overall survival. This helped determine which variables to include in the multivariate model using a p-value cutoff of 0.05. The null hypothesis for a univariate Cox regression is that the covariates do not affect overall survival: $\beta=0$.

Table 2: Univariate Cox Regression

Variables	Beta	HR (95% CI for HR)	p-value
age	0.016	1 (1-1)	0.038
sex	-0.52	0.59 (0.44-0.8)	0.00046
ph.karno	-0.01	0.99 (0.98-1)	0.058
ph.ecog	0.43	1.5 (1.3-1.9)	9.2e-06
pat.karno	-0.016	0.98 (0.97-0.99)	0.00099
meal.cal	-8.8e-05	1 (1-1)	0.68

From table 1, it was observed that age, Karnofsky score rated by patient, sex, and ECOG are significant covariates with univariable *p*-values less than 0.05. However, age and Karnofsky score by patients have hazard ratio of 1 and 0.98 respectively. Although significant, HR = 1 indicated no effect therefore these variables will be dropped for multivariate cox regression. Sex and ECOG scores are significant and also has HR less than 1 and greater than 1 respectively. These two variables are considered for multivariate cox regression.

D. **Multivariate Cox Regression**

The multivariate Cox regression model in Table 3 reveals significant associations between survival and ECOG score as well as sex. The ECOG score, with a p-value of 0.00000656 and a hazard ratio of 1.53, indicates a strong correlation between higher ECOG values and increased death risk. Patients with elevated ECOG scores have, on average, 1.5 times the chance of dying from lung cancer compared to those with lower scores, signifying the impact of disease debilitation on survival. Similarly, for the sex covariate, with a p-value of 0.0001417 and a hazard ratio of 0.592 (95% CI: 0.4427, 0.7922), being female is associated with a 40.8% lower risk of death compared to males. This suggests that being female correlates with a more favorable prognosis for lung cancer compared to being male.

Table 3. Sex and ECOG Cox Regression Model Output

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
sex	-0.5239	0.5922	0.1484	-3.529	0.0001417***
ecog	0.4278	1.5339	0.0949	4.507	0.00000656***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
	HR	lower 0.95 HR	upper 0.95 HR		
sex	0.5922	0.4427	0.7922		
ecog	1.5340	1.2735	1.8476		

It was concluded from the Cox regression analysis that ECOG is a significant predictor of survival time of patients with lung cancer. Similarly, sex was found to be a confounding variable and also influence the survival time. Therefore, the relevant point estimates of the patients and their associated 95% confidence intervals based on the ECOG score and sex was visualized and presented in figure 2.

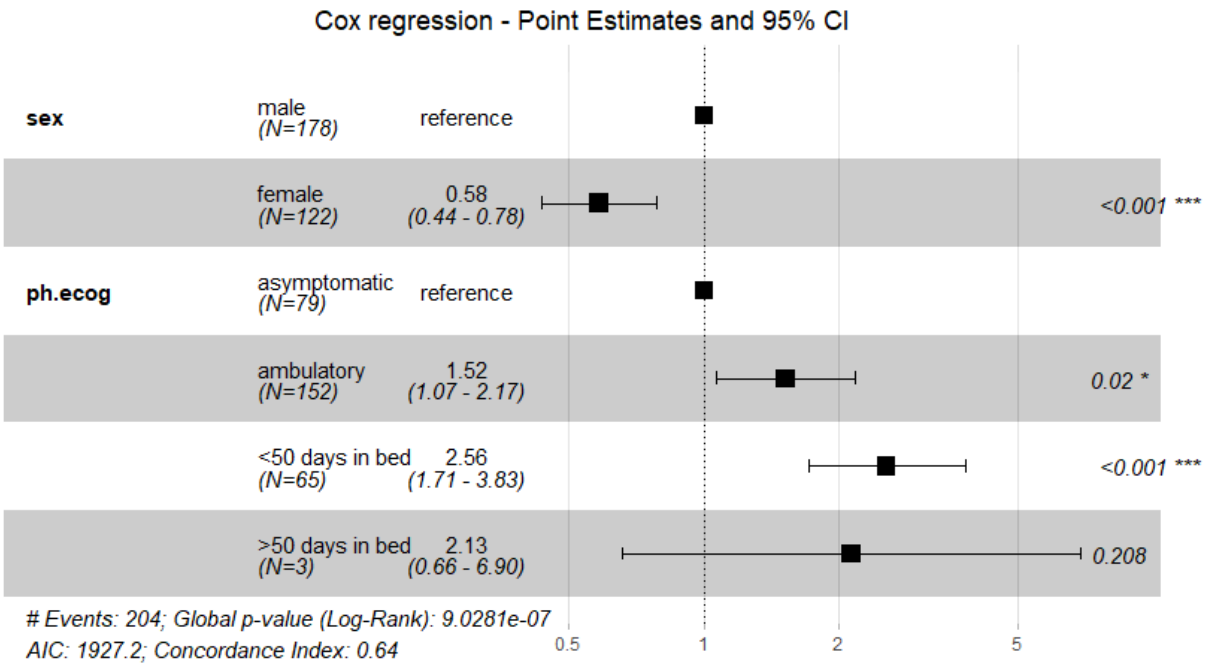

Fig. 2: Point estimates of Cox regression model with 95% confidence interval.

As seen from fig 2, female patients have a HR of 0.58 (95% CI: 0.44 – 0.78) from the male patients which served as baseline. This indicated that female patients have a 40.8% lower risk of death from lung cancer compared to male patients. As for ECOG performance score, taking asymptomatic patients as the baseline, patients with higher ECOG score are at higher risk of death from lung cancer. 50% of symptomatic patients but completely ambulatory is likely to die from lung cancer when compared to asymptomatic patients. Similarly, patients with higher ECOG score 2 and 3 are at twice risk of death as compared to the asymptomatic patients.

Discussion of Results

The study presents comprehensive insights into the demographic, clinical characteristics, survival trends, and influential factors in lung cancer patients.

Demographic and Clinical Characteristics: The study population, described through descriptive statistics, exhibits a mean age of 60, with individuals aged between 39 and 82. This shows that lung cancer is common among elderly people which is similar to the observation of Brown et al. (1996). Quantitative variables such as ph.karno, pat.karno, and meal.cal were characterized, highlighting missing values in some instances. Categorically, the majority of patients were male (59.33%). This is in line with several studies who reported higher incidence of lung cancer in male (Islami et al., 2015; Mohan et al., 2020). The ECOG categories showed a significant representation of symptomatic but completely ambulatory patients.

Survival Trends: The Kaplan-Meier curve illustrates the overall survival of lung cancer patients. The median survival of approximately 329 days and a 1-year survival probability of 40.2% underscore the challenges and limited prognosis associated with this malignancy.

Univariate Cox Regression: Univariate Cox regression identified age, sex, ph.ecog, pat.karno, and ECOG as significant covariates impacting overall survival. Notably, age and Karnofsky score were dropped for multivariate analysis due to negligible hazard ratios.

Multivariate Cox Regression: In the multivariate Cox regression model, ECOG and sex emerge as significant predictors of survival time. Higher ECOG scores correlate with a 1.5 times higher risk of death, emphasizing the impact of disease debilitation. Study carried out by Sehgal et al. (2021) maintained a similar observation that higher ECOG scores is responsible for low disease control. Female patients exhibit a 40.8% lower risk of death compared to males, implying a more favorable prognosis for women with lung cancer.

The point estimates and 95% confidence intervals of the Cox regression model based on ECOG score and sex indicated a significant result. Female patients demonstrate a significantly lower risk of death, while higher ECOG scores correspond to elevated mortality risk (Sehgal et al., 2021). Symptomatic but completely ambulatory patients have a 50% chance of death compared to asymptomatic patients, indicating the severity of symptoms on survival.

Strength and Limitation

Comprehensive analysis, including multivariate Cox regression and clear visualizations, provides valuable insights into predictors of lung cancer patient survival, enhancing clinical relevance. However, missing data, retrospective design, limited generalizability due to single-center data, and reliance on a specific cancer treatment group introduce biases and may impact the broader applicability of the findings.

Conclusion

The comprehensive analysis underscores the significance of ECOG performance status and sex, a confounding variable as predictors of survival in lung cancer patients. These findings contribute valuable insights for clinicians, aiding in risk stratification and informing personalized treatment strategies for improved patient outcomes. The study's rigorous statistical approach provides a robust foundation for understanding the complex interplay of factors influencing the trajectory of lung cancer.

References

Brown, J. S., Eraut, D., Trask, C., and Davison, A. G. (1996). Age and the treatment of lung cancer. *Thorax*, *51*(6), 564–568. https://doi.org/10.1136/thx.51.6.564

Cox DR (1972). Regression models and life tables (with discussion). J R Statist Soc B 34: 187–220

Islami, F., Torre, L. A., and Jemal, A. (2015). Global trends of lung cancer mortality and smoking prevalence. *Translational lung cancer research*, *4*(4), 327–338. https://doi.org/10.3978/j.issn.2218-6751.2015.08.04

Mohan, A., Garg, A., Gupta, A., Sahu, S., Choudhari, C., Vashistha, V., Ansari, A., Pandey, R., Bhalla, A. S., Madan, K., Hadda, V., Iyer, H., Jain, D., Kumar, R., Mittal, S., Tiwari, P., Pandey, R. M., and Guleria, R. (2020). Clinical profile of lung cancer in North India: A 10-year analysis of 1862 patients from a tertiary care center. *Lung India : official organ of Indian Chest Society*, *37*(3), 190–197. https://doi.org/10.4103/lungindia.lungindia_333_19

Sehgal, K., Gill, R. R., Widick, P., Bindal, P., McDonald, D. C., Shea, M., Rangachari, D., & Costa, D. B. (2021). Association of Performance Status with Survival in Patients with Advanced Non-Small Cell Lung Cancer Treated With Pembrolizumab Monotherapy. *JAMA network open*, *4*(2), e2037120. https://doi.org/10.1001/jamanetworkopen.2020.37120

Appendix

Appendix 1 – Description of the Dataset and Variables

The data contain subjects with advanced lung cancer from the North Central Cancer Treatment Group. It includes the following variables:

- inst: Institution code
- time: Survival time in days
- status: censoring status 1=censored, 2=dead
- age: Age in years
- sex: Male=1 Female=2
- ph.ecog: ECOG performance score as rated by the physician. 0=asymptomatic, 1= symptomatic but completely ambulatory, 2= in bed <50% of the day, 3= in bed > 50% of the day but not bedbound, 4 = bedbound
- ph.karno: Karnofsky performance score (bad=0-good=100) rated by physician
- pat.karno: Karnofsky performance score (0 = bad, 100 = good) as rated by patient
- meal.cal: Calories consumed at meals

Appendix II – Kaplan-Meier plots of Survival of patients with time based on ECOG and Sex

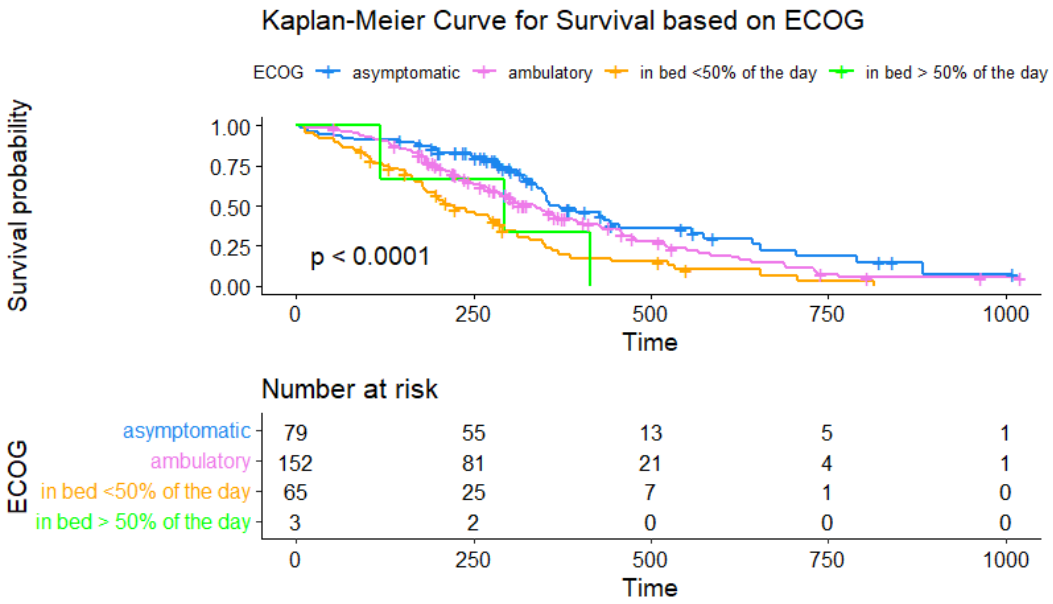

Fig. 3: Kaplan-Meier plots of Survival of patients with time based on ECOG. Higher ECOG scores is associated with poor survival of the lung cancer patients.

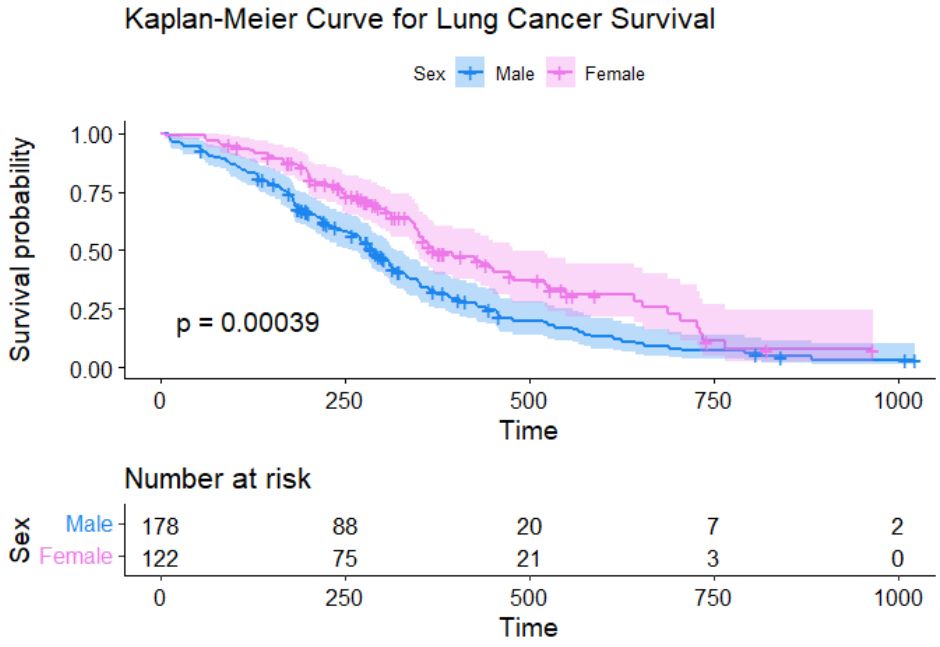

Fig. 4: Kaplan-Meier plots of Survival of patients with time based on sex. Female patients have better chance