



## Estimating the Survival of Patients With Lung Cancer: What Is the Best Statistical Model?

Siavosh Abedi<sup>1</sup>, Ghasem Janbabaee<sup>2</sup>, Mahdi Afshari<sup>3</sup>, Mahmood Moosazadeh<sup>4</sup>, Masoumeh Rashidi Alashti<sup>5</sup>, Akbar Hedayatzadeh-Omran<sup>5</sup>, Reza Alizadeh-Navaei<sup>5</sup>, Ehsan Abedini<sup>4</sup>

<sup>1</sup>Department of Internal Medicine, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran; <sup>2</sup>Gastrointestinal Cancer Research Center, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran; <sup>3</sup>Department of Community Medicine, Zabol University of Medical Sciences, Zabol, Iran; <sup>4</sup>Health Sciences Research Center, Addiction Institute, Mazandaran University of Medical Sciences, Sari, Iran; <sup>5</sup>Gastrointestinal Cancer Research Center, Mazandaran University of Medical Sciences, Sari, Iran

**Objectives:** Investigating the survival of patients with cancer is vitally necessary for controlling the disease and for assessing treatment methods. This study aimed to compare various statistical models of survival and to determine the survival rate and its related factors among patients suffering from lung cancer.

**Methods:** In this retrospective cohort, the cumulative survival rate, median survival time, and factors associated with the survival of lung cancer patients were estimated using Cox, Weibull, exponential, and Gompertz regression models. Kaplan-Meier tables and the log-rank test were also used to analyze the survival of patients in different subgroups.

**Results:** Of 102 patients with lung cancer, 74.5% were male. During the follow-up period, 80.4% died. The incidence rate of death among patients was estimated as 3.9 (95% confidence [CI], 3.1 to 4.8) per 100 person-months. The 5-year survival rate for all patients, males, females, patients with non-small cell lung carcinoma (NSCLC), and patients with small cell lung carcinoma (SCLC) was 17%, 13%, 29%, 21%, and 0%, respectively. The median survival time for all patients, males, females, those with NSCLC, and those with SCLC was 12.7 months, 12.0 months, 16.0 months, 16.0 months, and 6.0 months, respectively. Multivariate analyses indicated that the hazard ratios (95% CIs) for male sex, age, and SCLC were 0.56 (0.33 to 0.93), 1.03 (1.01 to 1.05), and 2.91 (1.71 to 4.95), respectively.

**Conclusions:** Our results showed that the exponential model was the most precise. This model identified age, sex, and type of cancer as factors that predicted survival in patients with lung cancer.

**Key words:** Survival, Lung neoplasms, Pulmonary cancer, Statistical models, Iran

### INTRODUCTION

Lung cancer is one of the most important health problems worldwide, due to its high incidence, poor prognosis, high fatality rate, and the high burden it imposes on society [1-3]. Lung cancer is classified as small cell carcinoma (SCLC) and non-small cell carcinoma (NSCLC). The latter type is classified into squamous cell carcinoma, adenocarcinoma, and giant cell carcinoma [1,4-7].

The survival of patients with lung cancer is one of the main indicators used to assess cancer control programs [1]. Few

Received: June 8, 2017 Accepted: January 21, 2019

**Corresponding author:** Mahmood Moosazadeh, PhD  
Health Sciences Research Center, Addiction Institute, Mazandaran University of Medical Sciences, Khazarabad Road, Sari 4815733971, Iran

E-mail: [mmoosazadeh1351@gmail.com](mailto:mmoosazadeh1351@gmail.com)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

studies have investigated the survival of lung cancer patients in developing countries, but this has emerged as a topic of interest in recent years because of the short survival time of these patients. Generally, regression models are applied to detect factors related to lung cancer survival, with options including Cox, Weibull, exponential, and Gompertz regression models. This study aimed to determine which model would best fit the survival data of patients with lung cancer and to identify the factors most strongly associated with survival.

## METHODS

This retrospective cohort study was carried out among patients suffering from lung cancer referred to the Tooba Clinic in Sari, the capital city of Mazandaran Province, Iran. All patients recruited for the study and their required information were registered with the Comprehensive Research Center for Cancer at Mazandaran University of Medical Sciences (ethical approval No. 1689). Informed consent was provided by each patient before entering the study.

All information necessary for this study, including the date of diagnosis, sex, age, type of cancer, and survival, was obtained from the database. No more information was obtained or used in the analyses, despite phone follow-up of patients' family members. In addition, since tumor, node and metastasis stage information was only recorded for 41 patients, it was not applied in the models.

The date of diagnosis was considered as the time of entry. To determine patients' up-to-date survival information, all patients were traced using their addresses and phone numbers. Each death after the date of diagnosis was considered as an event, and the date of death was considered as the event time. The follow-up time was considered to extend to the date when the most recent information was collected. If a patient's survival status could not be determined, he or she was considered to have missing records. Moreover, patients were right-censored if they did not experience the event of interest through the end of the follow-up period.

The mean age of patients was compared between the patients who died and those who survived using the *t*-test, since the data were normally distributed. A Kaplan-Meier table was used to illustrate patients' survival. The median survival time was compared by sex and type of lung cancer using the log-rank test.

To select the best model for determining the factors associated with patients' survival, 4 regression models were set up,

including Cox, Weibull, exponential and Gompertz models, and the highest fitness was detected based on which model yielded the lowest values of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Before applying these models, the proportional hazard assumption (equality of the risk of event during the time in the study variables) was tested using graphical and Schoenfeld residuals goodness-of-fit analysis. In addition, the constant hazard assumption during the time period of the study was assessed in order to apply the Weibull and exponential models. After conducting univariate analyses, multivariate models were applied to identify factors predictive of survival through measurements of hazard ratios (HRs) with 95% confidence intervals (CIs). All statistical tests were 2-sided, and *p*-values less than 0.05 were considered to indicate statistical significance. Stata version 14 (Stata-Corp., College Station, TX, USA) was used for data analysis.

## RESULTS

We investigated the survival of 102 patients with lung cancer, of whom 74.5% were male. The mean  $\pm$  standard deviation age of the patients was  $62.9 \pm 12.9$  years (range, 28.0 to 84.0), and the average age of males and females was  $63.2 \pm 13.1$  years and  $62.1 \pm 12.8$  years, respectively ( $p=0.71$ ).

The patients had been diagnosed between 2008 and 2014. The follow-up time was 2108.3 person-months. During this period, 82 (80.4%) patients died. The incidence rate (95% CI) of death was estimated as 3.9 (3.1 to 4.8) per 100 person-months. The cumulative survival proportion at the end of the first, second, third, fourth, fifth, sixth, seventh, and eighth years was 57%, 34%, 29%, 17%, 17%, 14%, 9%, and 9%, respectively. Moreover, the 5-year survival rate of males, females, NSCLC cases, and SCLC cases was 13%, 29%, 21%, and 0%, respectively. The median (interquartile range [IQR]) survival time of all patients, males, and females was 13 (6 to 39) months, 12 (6 to 28) months, and 16 (7 to 60) months, respectively ( $p=0.04$ ).

Eighty-one (79.4%) of the patients had NSCLC and 21 (20.6%) had SCLC. The median (IQR) survival period for these 2 groups was 16 (8 to 47) months and 6 (6 to 12) months, respectively ( $p=0.001$ ).

The proportional hazard assumption showed no difference in the risk of death between the study groups such as sex ( $p=0.47$ ), age ( $p=0.74$ ), and type of cancer ( $p=0.89$ ), making a Cox regression model suitable. The constant hazard assumption was also met ( $p=0.31$ ) for applying the exponential,

Weibull, and Gompertz models. The AIC, BIC, and likelihood ratio were 201.9, 208.8, and 8.8, respectively, for the Cox regression model; 122.7, 133.0, and 14.2, respectively, for the Weibull regression model; 121.7, 130.3, and 13.2, respectively, for the exponential regression model; and 123.6, 133.9, and 11.6, respectively, for the Gompertz model. The lowest values of the AIC and BIC and the highest likelihood ratio were found for the exponential model. Therefore, this model was applied to determine the factors associated with survival.

Univariate analysis showed that with each 1-year increase in age, the hazard of death increased by about 3% ( $p=0.002$ ). Moreover, risk of death in females was 48% lower than that in males ( $p=0.011$ ). SCLC cases had a 2.59-fold greater risk of death than NSCLC cases ( $p<0.001$ ). These associations were also statistically significant according to the multivariate models (Table 1).

## DISCUSSION

In this study, we assessed the precision of 4 regression models in determining the survival of patients with lung cancer and identifying factors related to survival. Ultimately, the exponential model was chosen to detect these factors. The mean

survival time for all patients, males, and females was estimated as 12.7 months, 12.0 months, and 16.0 months, respectively. The corresponding figures for NSCLC and SCLC cases were 16.0 months and 6.0 months, respectively.

Our multivariate models controlling for the potential confounders, showed a 3% increase in the death rate per each 1-year age increase. In addition, the risk of death was 44% lower in females than in males, while the risk of death was 2.91-fold higher in SCLC cases than in NSCLC cases.

The 5-year survival rate of our patients was lower than that of Chinese patients, but was higher than the rates estimated in studies conducted in Spain, Sweden, and France. In addition, the median survival time of patients in the current study was higher than those reported from West Azerbaijan and Yazd Provinces (Iran), Spain, Australia, and France. The higher rates of death among males, SCLC cases, and older patients are in accordance with the studies listed in Table 2 [1,6,8-13], but are in contrast to those observed in the study conducted in Azerbaijan Province. However, the difference is not considerable.

In a study conducted in Sweden, the 5-year survival rate among males and females was 11.5% and 20.1%, respectively, which is higher than the estimates of the current study. How-

**Table 1.** Univariate and multivariate exponential regression model of factors associated with death among patients with lung cancer

Variables	Univariate exponential regression		Multivariate exponential regression	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Age	1.03 (1.01, 1.05)	0.002	1.03 (1.01, 1.05)	<0.001
Sex	0.52 (0.31, 0.86)	0.011	0.56 (0.33, 0.93)	0.025
Type of lung cancer	2.59 (1.55, 4.33)	<0.001	2.91 (1.71, 4.95)	<0.001

HR, hazard ratio; CI, confidence interval.

**Table 2.** Comparison of survival rates and related factors across other studies

Authors [Ref]	Area of study	Publication year	Median survival time (mo)	Five-year overall survival rate (%)	HR (95% CI)			Type of model used
					Sex	Age (y)	Type of lung cancer	
Abazari et al. [1]	Iran	2015	4.8	-	M/F: 1.14 (0.85, 1.54)	60-70/<60: 1.37 (1.02, 1.83)	SCLC/NSCLC: 1.04 (0.66, 1.66)	Cox regression
Zhang et al. [6]	China	2017	-	35.2	-	-	-	-
Prim et al. [8]	Spain	2010	11.1	8.9	-	-	-	-
Svensson et al. [9]	Sweden	2013	-	13.6	M/F: 1.21 (1.06, 1.37)	60-69/<60: 1.07 (0.89, 1.30)	-	Cox regression
Biswas et al. [10]	USA	2014	-	-	M/F: 1.10 (1.01, 1.10)	51-70/25-50: 1.40 (1.20, 1.80)	-	Cox regression
Ball et al. [11]	Australia	2013	6.1	-	-	-	-	-
Zahir et al. [12]	Iran	2012	8.5	-	-	-	-	-
Grivaux et al. [13]	France	2009	7.0	10.4	-	-	-	-

HR, hazard ratio; CI, confidence interval; SCLC, small cell lung carcinoma; NSCLC, non-small cell lung carcinoma; M, male; F, female.

ever, the median survival time among NSCLC cases (16.5 months) and SCLC cases (7.5 months) was similar to our results [9].

A reason for discrepancies between our findings and those of other studies may be the exclusion of 127 cases with lung cancer from the current study due to their unknown survival status. These patients probably had poorer survival outcomes than those included in the study. Another explanation may be related to differences in when these studies were conducted. It seems that the survival of lung cancer patients is improving. Other factors that could be responsible for different results across studies include variation in the type of treatment, stage of disease, tumor grade, socioeconomic status, comorbidities, and cigarette smoking. However, insufficient evidence is available to prove that such factors are major sources of variation.

Unfortunately, we were not able to assess associations between survival and factors such as tumor size, comorbidities, smoking status, stage and grade of the tumor, body mass index, job status, location of the tumor, and treatment type. This is a limitation of the current study that should be addressed in future research.

In conclusion, our study showed that the exponential regression model was the most precise model for assessing the survival of patients with lung cancer and for identifying factors related to mortality. This model also showed that sex, age, and type of lung were predictors of survival.

## CONFLICT OF INTEREST

The authors have no conflicts of interest associated with the material presented in this paper.

## ACKNOWLEDGEMENTS

The authors would like to thank to the research deputy and Gastrointestinal Cancer Research Center of the Mazandaran University of Medical Science.

## ORCID

Siavosh Abedi <https://orcid.org/0000-0003-4453-3661>  
Ghasem Janbabaie <https://orcid.org/0000-0003-2651-6200>  
Mahdi Afshari <https://orcid.org/0000-0002-3159-8741>  
Mahmood Moosazadeh <http://orcid.org/0000-0002-5452-514X>  
Masoumeh Rashidi Alashti <https://orcid.org/0000-0002-3299-2193>

3299-2193  
Akbar Hedayatizadeh-Omran <https://orcid.org/0000-0002-9219-883X>  
Reza Alizadeh-Navaei <https://orcid.org/0000-0003-0580-000X>  
Ehsan Abedini <https://orcid.org/0000-0003-3767-0780>

## REFERENCES

1. Abazari M, Gholamnejad M, Roshanaei G, Abazari R, Roosta Y, Mahjub H. Estimation of survival rates in patients with lung cancer in west Azerbaijan, the northwest of Iran. *Asian Pac J Cancer Prev* 2015;16(9):3923-3926.
2. Pakzad R, Mohammadian-Hafshejani A, Ghoncheh M, Pakzad I, Salehiniya H. The incidence and mortality of lung cancer and their relationship to development in Asia. *Transl Lung Cancer Res* 2015;4(6):763-774.
3. Rafiemanesh H, Mehtarpour M, Khani F, Hesami SM, Shamlou R, Towhidi F, et al. Epidemiology, incidence and mortality of lung cancer and their relationship with the development index in the world. *J Thorac Dis* 2016;8(6):1094-1102.
4. Lüchtenborg M, Riaz SP, Lim E, Page R, Baldwin DR, Jakobsen E, et al. Survival of patients with small cell lung cancer undergoing lung resection in England, 1998-2009. *Thorax* 2014;69(3):269-273.
5. Sen E, Kaya A, Erol S, Savas I, Gonullu U. Lung cancer in women: clinical features and factors related to survival. *Tuberk Toraks* 2008;56(3):266-274 (Turkish).
6. Zhang C, Yang H, Zhao H, Lang B, Yu X, Xiao P, et al. Clinical outcomes of surgically resected combined small cell lung cancer: a two-institutional experience. *J Thorac Dis* 2017;9(1):151-158.
7. Kukulj S, Popović F, Budimir B, Drpa G, Serdarević M, Polić-Vižintin M. Smoking behaviors and lung cancer epidemiology: a cohort study. *Psychiatr Danub* 2014;26 Suppl 3:485-489.
8. Prim JM, Barcala FJ, Esquete JP, Reino AP, López AF, Cuadrado LV. Lung cancer in a health area of Spain: incidence, characteristics and survival. *Eur J Cancer Care (Engl)* 2010;19(2):227-233.
9. Svensson G, Ewers SB, Ohlsson O, Olsson H. Prognostic factors in lung cancer in a defined geographical area over two decades with a special emphasis on gender. *Clin Respir J* 2013;7(1):91-100.
10. Biswas T, Walker P, Podder T, Rosenman J, Efirid J. Important prognostic factors for lung cancer in tobacco predominant Eastern North Carolina: study based on a single cancer regis-

- try. *Lung Cancer* 2014;84(2):116-120.
11. Ball D, Thursfield V, Irving L, Mitchell P, Richardson G, Torn-Broers Y, et al. Evaluation of the Simplified Comorbidity Score (Colinet) as a prognostic indicator for patients with lung cancer: a cancer registry study. *Lung Cancer* 2013;82(2):358-361.
  12. Zahir ST, Mirtalebi M. Survival of patients with lung cancer, Yazd, Iran. *Asian Pac J Cancer Prev* 2012;13(9):4387-4391.
  13. Grivaux M, Zureik M, Marsal L, Asselain B, Peureux M, Chavailon JM, et al. Five year survival for lung cancer patients managed in general hospitals. *Rev Mal Respir* 2009;26(1):37-44 (French).