

## RESEARCH ARTICLE

# Cox Regression and Parametric Models: Comparison of How They Determine Factors Influencing Survival of Patients with Non-Small Cell Lung Carcinoma

Elahe Khaksar<sup>1</sup>, Mohsen Askarishahi<sup>1\*</sup>, Seyedhossein Hekmatimoghaddam<sup>2</sup>, Hassanali Vahedian-Ardakani<sup>3</sup>

### Abstract

**Background and objectives:** The present study of survival rate of patients with non-small cell carcinoma (NSCLC) compared the efficiency of Cox semi-parametric vs. parametric models in determination of influencing factors. **Methods:** In this retrospective cohort study, data were gathered from 190 patients with a confirmed diagnosis of NSCLC referred to Shahid Sadoughi and Shohadaye Kargar Hospitals in Yazd, Iran during 2005 to 2014. To identify and compare factors influencing the survival rate, a Cox semi-parametric model was fitted to the data. Data analysis was performed using the R software version R3.3.1, and the significance level was set at 0.05. **Results:** The average age was 64.5 years. About 40% of patients had stage 4 disease. The median survival was 8 months. After comparing the models, the more efficient was the log-normal distribution (AIC=889.3829), with which disease stage, type of therapy, and age were significant factors. Among the different types of therapy, chemotherapy and radiotherapy yielded higher survival rates, and increased age was associated with lower survival. **Conclusion:** The most efficient model was a log-normal model. Implementation of optimal therapies at early stages can improve the survival of patients.

**Keywords:** Cox proportional hazards model- influential factors- non-small cell lung carcinoma- parametric model

*Asian Pac J Cancer Prev*, **18** (12), 3389-3393

### Introduction

Lung cancer is the most common cancer killer worldwide, with a five-year survival rate of only 15% (Jemal et al., 2009). In the United States, there are approximately 226,000 new cases annually with an estimated 160,000 deaths (Siegel et al., 2012), and it cost an estimated US\$ 12 billion in 2010 (Mariotto et al., 2011). Non-small cell lung carcinoma (NSCLC) is one of two major types of lung cancer which accounts for 80–90% of all lung cancer cases (Siegel et al., 2014). Researches on the survival rate of these patients aim at identification of influential demographic and clinical factors. Rapp et al., (1988) evaluated the patients with NSCLC and showed that their median survival is about 4-5 months, with one-year survival rate of only 10%. Another similar study showed high mortality of these people, with an average survival of 7.150 months (Wao et al., 2013). A research in Germany in 2013 on 285 patients with lung cancer concluded that quitting smoking increases the efficacy of treatment and improves their quality of life (Andreas et al., 2013). In a study in the Netherlands on 451 patients with NSCLC who underwent surgery, they found that

surgery greatly reduces mortality rates (Battafarano et al., 2002). A study on the effect of sex on survival of patients with NSCLC who underwent surgery in Canada revealed that female gender can be a favorable factor for a better survival in stage one (Noris et al., 2013). Efficiency and validity of assessment tools are very important. Statistical methods for survival analysis are efficient instruments in identifying effective factors.

Studies have been conducted to compare parametric models and semi-parametric models in survival data analysis. Cox proportional hazards model (as a semi-parametric method), accelerated failure time (AFT) models such as Weibull, exponential, log-normal, and logistic-log (as parametric methods) and non-parametric Kaplan-Meier method are the most important approaches to analyze survival data. Various studies have been conducted on different clinical datasets to investigate the efficiency of different models. According to the results of these studies, different models do not have similar efficiency on different clinical data so that different data analyses require exploration among different models to find more efficient models. In a simulation study, it was shown that log-normal models and logistic logic models

<sup>1</sup>Department of Biostatistics and Epidemiology, School of Health, <sup>2</sup>Department of Laboratory Sciences, School of Paramedicine, <sup>3</sup>Department of Internal Medicine, School of Medicine, Shahid Sadoughi University of Medical Sciences, Safayeh, Yazd, Iran. \*For Correspondence: m.askari@ssu.ac.ir

are more efficient than Cox model (Orbe et al., 2002). So far, the studies on patients with NSCLC have been done with semi-parametric method. Therefore, conducting a study using both parametric and semi-parametric methods, in addition to determining risk factors, may lead to the introduction of a more efficient model. The purpose of the present study was to find the survival rate of patients with NSCLC, to determine prognostic factors which improve the survival rate (with the fitness of parametric and semi-parametric models) and finally, to select the best model to analyze survival.

## Materials and Methods

In a retrospective cohort study during the period 2005 to 2014, totally 200 records with NSCLC were extracted, 190 patients of whom were eligible to enter the study. These patients had visited Shahid Sadoughi and Shohadaye Kargar hospitals in Yazd, Iran. Data were collected from medical records of the patients, and completed with the latest health status of patients asked by telephone. The recorded demographic and clinical variables included age (23-65 or 66-91 years), the type of treatment (either surgery, chemotherapy or chemotherapy plus radiotherapy), the stage of the cancer at diagnosis (II, III or IV), history of smoking, duration of smoking, gender, histopathologic type (adenocarcinoma, squamous cell carcinoma [SCC], or large cell carcinoma [LCC]), patients' place of residence (the city of Yazd, other locations within the province of Yazd, or cities in other provinces), occurrence of death, the length of time between diagnosis and the commencement of treatment (in months), and the length of time between diagnosis and death (in months). To determine the survival rate, Kaplan-Meier non-parametric test was used. Also, univariate tests were done using log rank test to identify effective (significant) factors in survival rate. Then, for comparison of the multivariate parametric and Cox semi-parametric models, Akaike information criterion (AIC), the most efficient model was selected to determine influential factors. AIC is a measure of the goodness of fit in which each model with a smaller value is the proposed model which fits the data well. All of the analyses in this study were performed using the software R3.3.1, and all tests were carried out two-sided. P-values less than 0.05 were assumed significant.

## Results

Among 190 patients in this retrospective cohort, their median age was 64.500 years. Males constituted 71.600% of them (136 patient), and 76 (40%) of the patients were in stage IV at the time of diagnosis. Treatment methods undertaken were surgery alone in 53 patients (27.900%), chemotherapy alone in 66 patients (34.700%), and chemotherapy plus radiotherapy in 71 patients (37.400%). Results of evaluation of the survival rate at each level of variables that became significant in univariate log rank test are shown in Table 1.

The median survival of the patients was 8 months (range 1-48 months) (Figure 1).

Table 1. Demographic Characteristics of Patients with Non-Small Cell Lung Carcinoma and Results of Log Rank Test

Variable	n (percent)	p-value
Age (year)		0.001
23-65	101.00 (53.200)	
66-91	89.00 (46.800)	
Type of treatment		0
Surgery	53.00 (27.900)	
Chemotherapy	66.00 (34.700)	
Chemotherapy plus radiotherapy	71.00 (37.400)	
Stage of the disease		0
II	58.00 (30.500)	
III	56.00 (29.500)	
IV	40.00 (76.000)	
History of smoking		0.323
Yes	110.00 (57.900)	
No	80.00 (42.100)	
Duration of smoking		0.544
0	80.00 (42.100)	
<1year	50.00 (26.300)	
≥1year	60.00 (31.600)	
Gender		0.264
Man	136.00 (71.600)	
Woman	54.00 (28.400)	
Histopathologic subtype		0.107
Adenocarcinoma	77.00 (44.500)	
Squamous cell carcinoma	55.00 (28.900)	
Large cell carcinoma	58.00 (30.500)	
Place of residence		0.364
City of Yazd	91.00 (47.900)	
Other locations in the province of Yazd	34.00 (17.900)	
Cities in other provinces	65.00 (34.200)	

The 6-months survival rate for each of the age groups was as follows: 0.689 for the group 23-65 years (with standard error, SE= 0.047 and confidence interval, CI=0.603-0.788); 0.543 for the group 66-91 years (with SE= 0.053 and CI=0.448-0.658). The 18-months

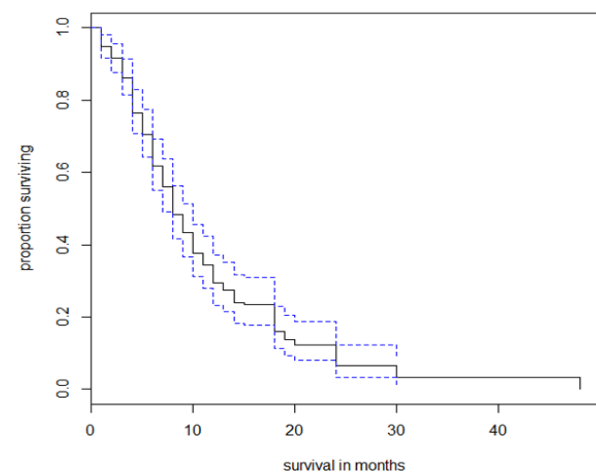


Figure 1. Kaplan-Meier Plot of Survival of Patients with Non-Small Cell Lung Carcinoma

Table 2. Multivariate Analysis of Prognostic Factors with Cox and Parametric Models

Variable	Cox HR (SE)	Exponential RR (SE)	Weibull RR (SE)	Log-Normal RR (SE)	Logistic log RR (SE)
Stage of the disease	2.519 (0.117)*	1.425 (0.098)*	1.494 (0.044)*	1.344 (0.045)*	0.744 (0.036)*
Type of treatment	0.286 (0.116)*	0.547 (0.103)*	0.578 (0.045)*	0.510 (0.048)*	1.879 (0.040)*
Age	1.580 (0.167)*	1.336 (0.162)	1.234 (0.074)*	1.203 (0.074)*	0.912 (0.058)
AIC	1207.588	1036.422	911.059	889.383	971.996

\* Significant difference at 0.05; HR, hazard ratio; RR, relative risk; SE, standard error; AIC, Akaike information criterion.

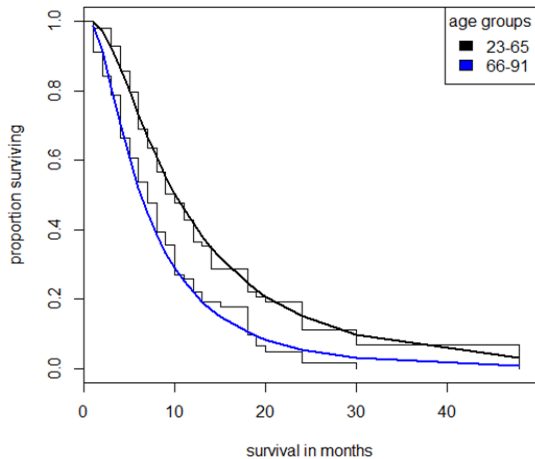


Figure 2. Log-Normal Plot of the Survival by the Age Groups

survival rates for the two age groups were 0.222 and 0.097, respectively. The 6-months survival rate for each type of treatment modalities were as follows: 0.140 for surgery (with SE=0.052 and CI=0.068-0.288); 0.576 for chemotherapy (with SE=0.062 and CI=0.467-0.712); and 0.986 for chemotherapy plus radiotherapy (with SE=0.014 and CI=0.959-1). The 18-months survival rates for the above-mentioned treatment groups were 0.037, 0.055 and 0.323, respectively. The 6-months survival rates for each of the clinical stages of the disease were as follows: 0.860 at stage II (with SE= 0.046 and CI=0.775-0.955); 0.784 at stage III (with SE= 0.055 and CI=0.683-0.900); 0.298 at stage IV (with SE=0.055 and CI=0.208-0.428).

Table 3. Multiple Log- Normal Model Classification

Variable	coef.	RR (HR)	LCL 95%	UCL 95%	p-value
Stage of the disease					
II	Ref				
III	-0.282	1.325	1.101	1.595	<0.010
IV	-0.6	1.823	1.525	2.178	<0.010
Type of treatment					
Surgery	Ref				
Chemotherapy	0.754	0.47	0.391	0.566	<0.010
Chemotherapy plus radiotherapy	1.352	0.259	0.214	0.312	<0.010
Age					
23-65	Ref				
66-91	-0.173	1.189	1.028	1.375	0.02

RR, relative risk; HR, hazard ratio; LCL, lower confidence limit; UCL, upper confidence limit; Ref, reference

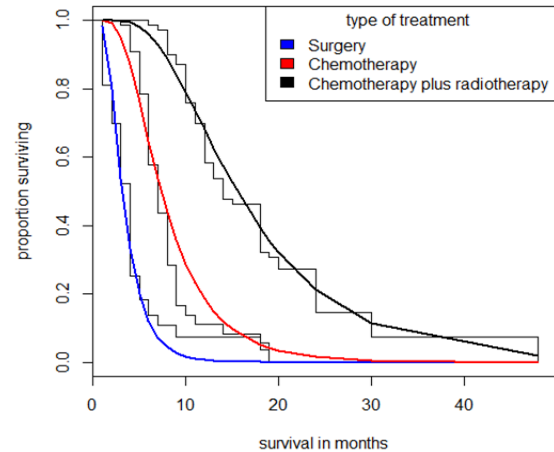


Figure 3. Log- Normal Plot of the Survival by the Type of Treatment

The 18-months survival rates for the stage II, III and IV were 0.308, 0.177 and 0.020, respectively.

Multivariate models of Cox, exponential, Weibull, log-normal, and logistic log were used for final modeling and comparison of the efficiency of models in predicting significant factors. Multivariate models are reported in Table 2.

Stage of the disease and the type of treatment had significant relationships with patients' survival at  $p < 0.05$ . Also, age was significant in all models, except in exponential and logistic log models. Also, according to the reported AIC values, log-normal model with the minimum value (AIC=889.383) showed better fitness. Therefore, the

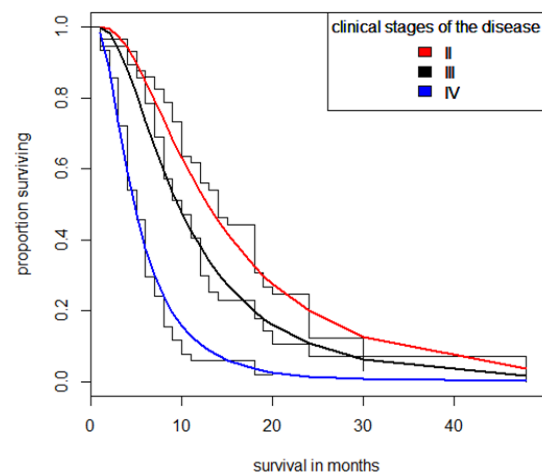


Figure 4. Log- Normal Plot of the Survival by the Clinical Stage

results of log-normal model for analysis at different levels of factors that influence survival rates are reported as is seen in Table 3. Figures 2, 3, 4 show log-normal plots of the survival by influential factors.

The hazard ratios (HR) divide the death rate in patients with the characteristic by the death rate in patients without the characteristic (reference group) as a measure of relative risk, according to the reported relative risk of patients in Stage III showed the risk of death in Stage III is 1.325 times higher than risk of death in Stage II (reference group).

## Discussion

The present study was conducted to determine the survival of patients with non-small cell lung carcinoma, to find prognostic factors of survival, and to determine the most efficient model for analysis of survival among parametric and non-parametric models. Stage of the disease, the type of treatment and age were identified by log-normal test as significant elements which affect survival. Multivariate analyses of Cox, Weibull, exponential, log-normal, and logistic log models were fitted on the data. Finally, log-normal model was selected as the most efficient model using Akaike's information criteria.

There are different views on the most efficient model in prediction of survival and its determinants in patients with cancer. In a study conducted to compare Cox model vs. parametric models on 197 children with acute leukemia in Iran, it was indicated that parametric models are preferred, and the Weibull method is the most efficient model among them (Hosseini Teshnizi and Ayatollahi, 2017). In another study that compared Cox model and parametric models on 484 patients with gastrointestinal cancer, parametric models were found preferred, and logistic log model was the most efficient model (Ghadimi et al., 2011). In again another work on comparison of Cox model vs. parametric models in 3421 patients, the findings showed that parametric models are the best, and gamma model was the most efficient model among the latter (Ravangard et al., 2011). Still another study conducted on 408 patients with gastric cancer to compare Cox model and parametric models, parametric models were preferred, especially Weibull model (Roshanaei et al., 2011). Among non-neoplastic diseases, a study on 529 patients with liver transplant showed that parametric models were performing better than Cox model, and log-normal model was the most efficient model among parametric models (Adelian et al., 2015). In a study carried out in Ireland by Ramzi et al., (2012) on 292 patients, they stated that modern chemotherapy can have positive effects on survival of patients with NSCLC who have poor performance. Chinese researchers studied 221 operated patients with NSCLC in 2012, and stated that postoperative radiotherapy can significantly improve the survival rate of the patients, which is consistent with our study (Dai et al., 2011). Hsu et al., (2012) studied patients 45 years of age and younger with NSCLC; there was a significant relationship between survival and stage of the disease. They deduced that stage IV is a negative

influential factor. A direct relationship between age and survival and also between histological grade and survival was found in a research carried out in China on prognostic factors influencing the survival of patients with NSCLC; that is, old age and poor histological differentiation were associated with poor survival, the former being similar to the results of our study (Wang et al., 2011). A study in Brazil found that survival in NSCLC patients is independent of gender and histologic type (Santoro et al., 2011). All of the results of the above-mentioned studies are consistent with our findings.

However, there are yet other studies that have yielded findings inconsistent with the present study. In a study performed in the United States they found that there is inverse relationship between the quantity of cigarette smoking in patients with NSCLC and their survival rate after diagnosis (Janjigian et al., 2010). However, in our study the survival of patients was independent of the quantity of tobacco smoking. In a study in Norway they stated that in general, women have a better chance of survival, and this factor is independent of age, treatment modality, histologic type and number of years after diagnosis (Pitz et al., 2013). In Taiwan a similar conclusion was stated by those who studied survival of patients 45 years of age and younger with NSCLC (Hsu et al., 2012). Both of the above studies are inconsistent with our study in which there was a significant relationship between survival and gender. For reasons such as lack of environmental conditions or proper facilities, evaluation of some factors was not possible in our work, which could potentially lead to interesting results: 1) skin color, as was shown in the United States when the researchers studied patients with NSCLC (Shirvani et al., 2012); 2) blood group, according to Sun et al., (2015) in Taiwan who studied 339,432 individuals, among them 1,235 patients had lung cancer, and concluded that the blood group can be influential.

In conclusion, overall, we concluded that according to reported AIC values the most efficient model is log-normal (parametric) model; the age group, the clinical stage of the disease, and the type of treatment of the individual are the most influential predictive factors regarding survival of patients with NSCLC. Identifying these important influential factors on the survival of patients could help in better management. In terms of survival rate, the best type of treatment, provided that all other variables were constant, was chemotherapy plus radiotherapy. The higher the clinical stage of the disease, the lower their survival rate. Also, the age had inverse relationship with the survival rate. Therefore, early diagnosis in younger ages and also a proper method of therapy for patients can reduce the risk of death.

## References

- Adelian R, Jamali J, Zare N, et al (2015). Comparison of Cox's regression model and parametric models in evaluating the prognostic factors for survival after liver transplantation in Shiraz during 2000–2012. *Int J Org Transplant Med*, 6, 119-25.
- Andreas S, Rittmeyer AC, Hinterthaler M, Huber RM (2013).

- Smoking cessation in lung cancer-Achievable and effective. *Dtsch Arztebl Int*, **43**, 719-24.
- Battafarano RJ, Piccirillo JF, Meyers BF, et al (2002). Impact of comorbidity on survival after surgical resection in patient with stage I non small cell lung cancer. *J Thorac Cardiovasc Surg*, **123**, 270-80.
- Dai H, Hul ZH, Wei Ji, et al. (2011). Postoperative radiotherapy for resected pathological stage 3A-N2 non small cell lung cancer: a retrospective study of 221 cases from a single institution. *Oncologist*, **16**, 641-50.
- Ghadimi MR, Mahmoodi M, Mohammad K, et al (2011). Family history of the cancer on the survival of the patients with gastrointestinal cancer in northern Iran, using frailty models. *BMC Gastroenterol*, **11**, 104.
- Hosseini Teshnizi S, Ayatollahi SMT (2017). Comparison of cox regression and parametric models: Application for assessment of survival of pediatric cases of acute leukemia in Southern Iran. *Asian Pac J Cancer Prev*, **18**, 981-5.
- Hsu CL, Chen KY, Shih JY, et al (2012). Advanced non small cell lung cancer in patients aged 45 years or younger: outcomes and prognostic factors. *BMC Cancer*, **12**, 241.
- Janjigian YY, McDonnell K, Kris MG, et al (2010). Pack years of cigarette smoking as a prognostic factor in patients with stage 3B/4 non small cell lung cancer. *Cancer*, **116**, 670-5.
- Jemal A, Siegel R, Ward E, et al (2009). Cancer statistics. *CA Cancer J Clin*, **59**, 225-49.
- Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML (2011). Projections of the cost of cancer care in the United States: 2010-2020. *J Natl Cancer Inst*, **103**, 117-28.
- Pitz MW, Musto G, Navaratnam S (2013). Sex as an independent prognostic factor in a population-based, non small cell lung cancer cohort. *Can Respir J*, **20**, 30-4.
- Orbe J, Ferreira E, Núñez-Antón V (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Stat Med*, **21**, 3493-510.
- Rapp E, Pater JL, Willan A, et al (1988). Chemotherapy can prolong survival in patients with advanced non-small-cell lung cancer- report of a Canadian multicenter randomized trial. *J Clin Oncol*, **6**, 633-41.
- Ravangard R, Arab M, Rashidian A, et al (2011). Comparison of the results of cox proportional hazards model and parametric models in the study of length of stay in a Tertiary Teaching Hospital in Tehran, Iran. *Acta Med Iran*, **49**, 650-8.
- Roshanaei G, Kazemnejad A, Sadighi S (2011). Assessment of risk factors affecting recurrence of patients with gastric cancer in the presence of informative censoring in Iran. *Asian Pac J Cancer Prev*, **12**, 2443-6.
- Salloum RG, Smith TJ, Jensen GA, Lafata JE (2012). Survival among non small cell lung cancer patients with poor performance status after first line chemotherapy. *Lung Cancer*, **77**, 545-9.
- Santoro IL, Roamos RP, Franceschini J, Jamnik S, Fernandes ALG (2011). Non small cell lung cancer in never smokers: a clinical entity to be identified. *Clinics (Sao Paulo)*, **66**, 1873-7.
- Scaglia NC, Chatkin JM, Pinto JA, et al (2013). Role of gender in survival of surgical patient with non small cell lung cancer. *Ann Thorac Med*, **8**, 142-7.
- Siegel R, Ma J, Zou Z, Jemal A (2014). Cancer statistics. *CA Cancer J Clin*, **64**, 9-29.
- Siegel R, Naishadham D, Jemal A (2012). Cancer statistics. *CA Cancer J Clin*, **62**, 10-29.
- Shirvani SM, Jiang J, Chang JY, et al (2012). Comparative effectiveness of 5 treatment strategies for early-stage non-Small cell lung cancer in the elderly. *Int J Radiat Oncol Biol Phys*, **84**, 1060-70.
- Sun W, Wen CP, Lin J, et al (2015). ABO blood types and cancer risk-A cohort study of 339,432 subjects in Taiwan. *Cancer Epidemiol*, **39**, 150-6.
- Wang Wl, Shen-tu Y, Wang ZQ (2011). Prognostic factors for survival of stage IB upper lobe non small cell lung cancer patient: A retrospective study in Shanghai China. *China J Cancer Res*, **23**, 265-70.
- Wao H, Mhaskar R, Kumar A, Miladinovic B, Djulbegovic B (2013). Survival of patients with non small cell lung cancer without treatment: a systematic review and meta-analysis. *Syst Rev*, **2**, 10.