Orjinal Araştırma / Original Article

Evaluation of Death Risks of Lung Cancer Patients with Frailty Models

Akciğer Kanseri Hastalarının Ölüm Risklerinin Zayıflık Modelleri ile Değerlendirilmesi

Özge Pasin[1], Şirin Çetin[2], İsa Dede[3]

[1]Department of Biostatistics, Faculty of Medicine, Istanbul University, Fatih, Istanbul, Turkey
[2]Department of Biostatistics, Faculty of Medicine, Tokat Gaziosmanpaşa University, Tokat, Turkey
[3]Medical Oncology, Faculty of Medicine, Mustafa Kemal University, Antakya, Turkey

Abstract

Objective: The aim of this study is to investigate the prognostic factors of lung cancer by evaluating the most appropriate survival model with a selection criteria.

Material and Method: In the study, the data of 185 patients diagnosed with lung cancer from the Medical Oncology Outpatient Clinic of Mustafa Kemal University Faculty of Medicine were retrospectively obtained from the patient files. The frailty models with different distributions were used for evaluating the heterogeneity between patients. Model selections were made according to AIC and BIC criteria.

Results: The median survival time of patients with lung cancer in the study was 11 months (95% confidence interval 9.57-12.42). The best frailty models' frailty distribution was lognormal and the basic hazard function distribution was loglogistic. The best model results showed that, the effect of the albumin variable on the risk of death of lung cancer patients was statistically significant (p = 0.018).

Conclusions: Generally, environmental and genetic factors that affect the survival time of lung cancer patients can not be evaluated. Thus, the term of the frailty resulting from the heterogeneity of factors when assessing individuals influencing survival of patients with lung cancer should be taken into account.

Keywords: Lung cancer, frailty, survival time

Öz

Amaç: Bu çalışmanın amacı; akciğer kanseri verisini analiz etmek için en uygun sağkalım modelini seçim kriterleri ile değerlendirerek akciğer kanseri verisinin prognostik faktörlerini ortaya çıkarmaktır.

Gereç ve Yöntem: Çalışmada Mustafa Kemal Üniversitesi Tıp Fakültesi Tıbbi Onkoloji Polikliniğinden 185 akciğer kanseri tanısı almış hastaya ait veriler geriye dönük olarak hasta dosyalarından elde edilmiştir. Bireylerin arasındaki heterojenliklerin değerlendirilmesi amacıyla farklı dağılımlara sahip zayıflık modelleri kurulmuştur. Modeller AIC ve BIC kriterlerine göre değerlendirilmiştir.

Bulgular: Çalışmadaki akciğer kanserli hastalara ait medyan sağkalım süresi 11 ay (%95 güven aralığı 9.57-12.42) olarak elde edilmiştir. Temel hazard fonksiyonu loglojistik ve zayıflık dağılımı lognormal olan zayıflık modeli en iyi model olarak belirlenmiştir. Bu modele ait sonuçlar incelendiğinde, albumin değişkeninin akciğer kanseri hastaların ölüm riski üzerine etkisi istatistiksel olarak anlamlı bulunmuştur (p=0.018).

Tartışma ve Sonuç: Genellikle akciğer kanseri hastalarının yaşam süresini etkileyen değerlendirmeye alınamayan çevresel ve genetik faktörlerin etkisi vardır. Dolayısıyla, akciğer kanseri hastalarının yaşam süresini etkileyen faktörler değerlendirilirken bireyler arasındaki heterojenliklerden kaynaklanan zayıflık teriminin de dikkate alınması alınması gerekmektedir.

Anahtar Kelimeler: Akciğer kanseri, zayıflık, sağkalım analizi

INTRODUCTION

Lung cancer is the most common type of cancer in the world.[1] In the literature, approximately 12.8% of cancer cases and 17.8% of cancer deaths consist of lung cancer.[2] 1.8 million new cases are reported annually for lung cancer, the most common type of cancer in the world.[3]

It is reported that the second most common cause of death in our country is cancer. Turkey Statistical Institute have been reported that lung cancer is the most cause of cancer deaths and it accounts for 31% of cancer-related deaths according to data from 2017.[4] Lung cancer is the most common type of cancer in our country in total and in men. Annual incidence is 66.7 per 100,000 in men and 7.5 per 100,000 in women. Among all cancer types, it is in the first place with a rate of 27% in men, and it is in the fifth place in women with a rate of 5%.[5] Lung cancer is divided into two classes; non-small cell lung cancer (NSCLC) accounts for more than 80% of all cases and The other 20% constitutes small cell lung cancer (SCLC).[6]

The prognosis of lung cancer is bad. Despite all advances in diagnosis and treatment, 5-year survival in NSCLC is around 15%.[7] Since lung cancers usually do not show symptoms at an early stage, lung cancer is less diagnosed in the early stages. Distant organ metastasis is detected in 40% of patients with lung cancer when diagnosed. Metastasis of the brain (43%), adrenal glands (40%), liver (40%), bone (33%), kidneys (23%) and abdominal lymph nodes (30%) are frequently encountered.[8,9] The survival rate in lung cancer is very low, and determining and revealing prognostic factors with different statistical methods is very important in terms of supporting the clinic in patient management.

The method to be used in examining the time of death for lung cancer is survival analysis. When the literature is reviewed, classical survival methods are generally used in studies. The most used survival analysis is the Cox regression analysis. However, in order to use this model, there should be no heterogeneity between individuals. In cancer data, the risks of encountering with the event of interest (death) are generally different from each other due to the genetic structures or clinical characteristics of each individual. There is often heterogeneity between individuals due to unexplained variables or the effects of other variables that cannot be modeled. When the results are evaluated with survival models that take these heterogeneities into account, more unbiased evaluations will be made. There are different frailty models used in survival analysis. The aim of this study is to reveal the prognostic factors of lung cancer data by evaluating the most appropriate survival model with selection criteria to analyze lung cancer data.

Frailty Models

In survival analysis, individuals with similar independent factor characteristics for a particular disease are assumed to have the same risks. However, some individuals have a higher risk than other individuals, usually due to unknown factors.

These unknown heterogeneities between individuals are called frailty. The term frailty has a very important place in survival analysis and frailtyes among individuals should be investigated. Although the characteristics of the independent variables related to the outcome variable are similar in 28 individuals, each individual has a different risk of encountering the relevant outcome variable. Individuals show heterogeneity within themselves. Therefore, variable characteristics of individuals differ from individual to individual or observations may have different distribution characteristics.[10,11]

The frailty model is generally defined with the help of the following equation.

$$h_{ij}(t|u_i) = h_0(t)u_i \exp(x_{ij}\beta)$$

In the function $h_0(\cdot)$ is the basic hazard function, u_i is the frailty term of individuals in i.group, the x_{ij} term is the j.individual's covariances vector of i.group and the β coefficient refers to the regression coefficients. Different models can be created depending on the basic hazard function and the distribution of frailty. When analyzed according to the distribution of frailty, there are different distribution types.[12]

Gamma Frailty Model

In the Gamma frailty model, the term frailty is a random variable which probability density function is U ~ Gam (θ). The model is as follows.

$$f(u) = \frac{\theta^{-\frac{1}{\theta}} u^{\frac{1}{\theta}-1} \exp(-\frac{u}{\theta})}{\Gamma(\frac{1}{\theta})}$$

$\Gamma(\cdot)$ is a gamma function with a gamma (μ, θ) distribution. The term θ is the variance value of this distribution.[12]

Positive Stable Frailty Model

In the positive stable frailty model, there are two parameters α <1 and δ> 0. The model is obtained with the help of the function below.

$$f(u) = -\frac{1}{\pi u} \sum_{k=1}^{\infty} \frac{\Gamma(k(1-v)+1)}{k!} (-u^{v-1})^k \sin((1-v)k\pi)$$

The v parameter in the function takes a value between zero and one and is equal to 1-α. Since the mean and variance are undefined in the model, there is no frailty variance value corresponding to the heterogeneity parameter. Therefore, the v parameter is used in interpretations.[12]

Inverse GaussianFrailty Model

The inverse Gaussian frailty distribution IG (θ) is defined by the following function.

$$f(u) = \frac{1}{\sqrt{2\pi\theta}} u^{-\frac{3}{2}} \exp(-\frac{(u-1)^2}{2\theta})$$

θ parameter takes positive value.[12]

Lognormal Frailty Model

The lognormal frailty model is expressed in terms of LN (θ). The probability function of the model is expressed as follows.

$$f(u) = (2\pi\theta)^{-\frac{1}{2}} u^{-1} \exp\{-\frac{(logu)^2}{2\theta}\}$$

θ parameter takes positive value.[12]

Model selection criteria

AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) selection criteria will be used to evaluate the effectiveness of different frailty models.

AIC is defined as

$$AIC = -2logL(\hat{\theta}) + 2k$$

The θ parameter is model vector of parameters, $L(\hat{\theta})$ is the likelihood of the candidate model evaluated in the maximum likelihood estimate. The k is the estimated parameter numbers fort he candicate model.

Akaike (1978) and Schwarz (1978) designed two closely consistent model selection criteria from the bayes perspective. While Schwarz Koopman-Darmois derives the SIC (Schwarz Information Criteria) criterion for select models, Akaike derives the BIC (Bayesian Information Criterion) model selection criterion for selected model problems in linear regression.

BIC is defined as in the below.

$$BIC = -2logL(\hat{\theta}) + klogn$$

In model comparisons, the model with the lowest AIC and BIC value is preferred.[13,14]

MATERIAL AND METHOD

Data

In the study, the data of 185 patients diagnosed with lung cancer from the Medical Oncology Outpatient Clinic of Mustafa Kemal University Faculty of Medicine were retrospectively obtained from the patient files. In order to investigate the factors affecting the survival times of the patients, the end point was death and the time of diagnosis was taken as the time of onset. Our retrospective study approval was obtained from Mustafa Kemal University Faculty of Medicine Ethics Committee (2020/12).

Statistical Analysis

In the survival models, the end point was death. Survival rates were obtained by the Kaplan-Meier method. Frailty models have been established in order to evaluate the heterogeneities between individuals. For the frality models, loglogistic, exponential and lognormal distributions for the basic hazard function were examined, while the frailty distribution was investigated with gamma, inverse Gaussian, positive stable and lognormal distributions.

The statistical evaluations were made using SPSS 21.0 and R package program. The statistical significance level was taken as 0.05 and p <0.05 was considered statistically significant.

RESULTS

In the study, data of 185 patients from Mustafa Kemal University Medical Faculty Medical Oncology Outpatient Clinic were retrospectively obtained from patient files. The follow-up period of 185 lung cancer patients was varied between 1 and 52 months. 93.5% of the patients (173 people) encountered death. The average and median survival times of the data in the study are given in **Table 1**. When **Table 1** was examined, it was observed that the mean follow-up period was 14.876 and the median survival time was 11 months.

The survival curve according to the death status of the patients was given in **Figure 1**. The two-year survival rate of the patients is 46%. It was observed that most of the patients did not live after the 50th month.

Individuals with similar independent factor characteristics are assumed to have the same risks in Cox regression analysis. However, the risks of having the event of interest are different among patients due to unknown genetic characteristics or the effects of other independent variables that are not included in the model. These heterogeneities should be included in the model and analyzed with frailty models. **Table 2** contains the results of the frailty models we applied for our lung cancer data. As can be seen from the table, evaluations are made by taking the basic

Table 1. Descriptive statistics of survival time							
Mean				**Median**			
Coefficient	**Standard error**	**95% confidence interval**		**Coefficient**	**Standard error**	**95% confidence interval**	
		Lower Limit	**Upper Limit**			**Lower Limit**	**Upper Limit**
14.876	.901	13.109	16.642	11.000	.727	9.575	12.425

Table 2. Frailty model results for different distributions

		Coefficient	β	Standard error	p
FD: Gamma BHF: Loglogistic	Hemoglobin	0.330	1.3909	0.200	0.099
	Platelet	0.121	1.1286	0.202	0.551
	Albumin	0.462	1.5872	0.186	0.013
	CRP	0.299	1.3485	0.221	0.176
FD: Gamma BHF: Exponential	Hemoglobin	0.220	1.2460	0.165	0.183
	Platelet	0.105	1.1107	0.186	0.571
	Albumin	0.368	1.4448	0.157	0.019
	CRP	0.197	1.2177	0.192	0.306
FD: Gamma BHF: Lognormal	Hemoglobin	0.281	1.3244	0.174	0.106
	Platelet	0.135	1.1445	0.186	0.468
	Albumin	0.429	1.5357	0.163	0.008
	CRP	0.258	1.2943	0.198	0.193
FD: Inverse Gaussian BHF: Lognormal	Hemoglobin	0.280	1.3231	0.173	0.105
	Platelet	0.135	1.1445	0.186	0.468
	Albumin	0.429	1.5357	0.163	0.008
	CRP	0.258	1.2943	0.198	0.192
FD: Positive stable BHF: Lognormal	Hemoglobin	0.275	1.3165	0.189	0.144
	Platelet	0.134	1.1433	0.198	0.499
	Albumin	0.426	1.5311	0.202	0.035
	CRP	0.254	1.2891	0.208	0.221
FD: Lognormal BHF: Lognormal	Hemoglobin	0.293	1.3404	0.187	0.117
	Platelet	0.139	1.1491	0.193	0.471
	Albumin	0.437	1.5480	0.171	0.011
	CRP	0.267	1.3060	0.207	0.198
FD: Positive stable BHF: Üstel	Hemoglobin	0.220	1.2460	0.165	0.182
	Platelet	0.105	1.1107	0.185	0.569
	Albumin	0.368	1.4448	0.157	0.019
	CRP	0.197	1.2177	0.192	0.305
FD: Inverse Gaussian BHF: Loglogistic	Hemoglobin	0.337	1.4007	0.206	0.101
	Platelet	0.137	1.1468	0.217	0.527
	Albumin	0.470	1.5999	0.194	0.015
	CRP	0.303	1.3539	0.225	0.178
FD: Positive stable BHF: Loglogistic	Hemoglobin	0.316	1.3716	0.267	0.237
	Platelet	0.083	1.0865	0.222	0.709
	Albumin	0.487	1.6274	0.322	0.131
	CRP	0.312	1.3661	0.284	0.272
FD: Lognormal BHF: Loglogistic	Hemoglobin	0.451	1.5698	0.254	0.076
	Platelet	0.238	1.2687	0.249	0.339
	Albumin	0.570	1.7682	0.242	0.018
	CRP	0.363	1.4376	0.274	0.184

FD: Frailty Distribution, BHF: Basic Hazard Function

Table 3. AIC and BIC values for models

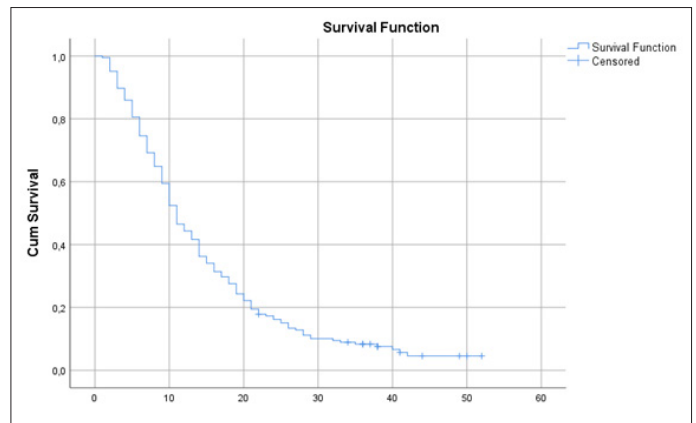| Basic Hazard Function		Frailty Distribution			
		Gamma	Invers Gaussian	Positive Stable	Lognormal
Loglogistic	AIC	1251.936	1251.877	1252.858	1249.970
	BIC	1274.479	1274.420	1275.400	1272.512
Exponential	AIC	1290.961		1290.961	
	BIC	1310.283		1310.283	
Lognormal	AIC	1251.168	1251.170	1251.184	1251.124
	BIC	1273.711	1273.712	1273.727	1273.666

Figure 1. Survival curve of patients with lung cancer

hazard function distribution and frailty distributions differently. **Table 3** contains AIC and BIC values of frailty models with different dispersion characteristics. When **Table 3** is examined, the lowest AIC and BIC values were obtained for the model with the basic hazard function loglogistic and the frailty distribution lognormal. (AIC: 1249,970; BIC: 1272,512). The models for frailty distributions with basic hazard functions exponential, inverse gaussian and lognormal distributions were not given in the table because they were not consistent. The frailty model with the highest AIC and BIC values was observed to belong to models in which the basic hazard function was exponential. Therefore, when the AIC and BIC criteria was evaluated, the frailty model with the lowest AIC and BIC values, loglogistic basic hazard function and lognormal frailty distribution constitutes the best model for our lung cancer data. When the results of this model were examined, the effects of hemoglobin, platelet and CRP variables on death risks were found to be statistically insignificant (each p>0.05). The effect of the albumin variable on the risk of death in lung cancer patients was found to be statistically significant (p=0.018). It was observed that the death risk of individuals with albumin less than 3.82 is 1.76 times higher than patients with albumin value greater than 3.82. In addition, when the models were examined, it was seen that the effect of no variable on survival time was not statistically significant for the model with a positive stable distribution of frailty and a loglogistic basic hazard function. In general, the effects of albumin on the risk of death in the models were found to be statistically significant.

CONCLUSION

In our study, the relationships of albumin, platelet, CRP and hemoglobin values of patients with lung cancer with survival time was evaluated. There are many factors on the prognosis of the disease. The effects of the factors that may be the most important in the study were basically evaluated. Degirmencioglu observed in his study that low hemoglobin had a negative effect on survival time. For CRP, similar to the results in our study, a significant relationship with survival time was not observed. Similar to the results in our study, low albumin levels were found to have a statistically significant effect on survival time.[15]

According to the study conducted by Yang et al.[16] In 2019, it was observed that the effects of albumin and CRP variables on survival times were statistically significant.

When the studies are evaluated, classical survival analysis methods are generally used in the analysis of data on lung cancer. However, there are heterogeneities among individuals due to environmental and genetic factors that are not taken into account when evaluating the factors affecting survival. It is recommended to include these heterogeneities in models. In this sense, we think that our study will make important contributions to the evaluation of the survival times of patients with lung cancer.

ETHICAL DECLARATIONS

Ethics Committee Approval: The study was carried out with the permission of Mustafa Kemal University Faculty of Medicine Ethics Committee (2020/12).

Informed Consent: Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process: Externally peer-reviewed.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

Author Contributions: All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893-917.

2. Savaş İ, Akkoçlu A, Göksel T, et al. Lung and Pleural Malignancies Working Group. Lung cancer diagnosis and treatment guideline. Turkish Thoracic Journal 2006; 7: 1-37)

3. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries - Bray - - CA: A Cancer Journal for Clinicians - Wiley Online Library- September 2018.)

4. (TÜİK. "Death Cause Statistics, 2017" News Bulletin, Number:27620, 26 April 2018.)

5. (T. C. Health MinistryCancer Incidence ın turkey in 2004-2006 www. kanser. gov. tr. Available: 26. 08. 2020)

6. Mountain CF. Revisions in the International System for Staging Lung Cancer. Chest 1997;111:1710-7.)

7. (Cancer Facts and Figures 2003, in American Cancer Society, Surveillance Research. Atlanta, American Cancer Society 2003)

8. Shields TW. Surgical Treatment of nonsmall cell lung cancer. In: Shields TW, Reed EW (eds). General Thoracic Surgery. 5th edition. Philadelphia: Lippincott Williams & Wilkins; 2000. 1311-41

9. Quint LE, Tummala S, Brisson LJ, et al. Distribution of distant metastases from newly diag-nosed non- smail cell lung cancer. Ann Thorac Surg 1996;62:246-50

10. Ozge P, Ahmet D, Handan A, Rian, D, Hasan K. Assessment of death risk of breast cancer patients with joint frailty models. Saudi Medical Journal, 2020;41(5):491-498.

11. Hanagal DD. (2011). Modelling Survival Data Using frailty Models. Boca Raton: Chapman & Hall/CRC.

12. Munda M, Rotolo F, Legrand C. Parfm: parametric frailty models in R. J Stat Soft 2012;51(11):1-20.

13. Ucal MŞ. A Brief Survey of Econometrics Model Selection Criteria 2006;7(2):41-46

14. The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications. Frank J. Fabozzi, Sergio M. Focardi, Svetlozar T. Rachev and Bala G. Arshanapalli. © 2014 John Wiley & Sons, Inc. Published 2014 by John Wiley & Sons, Inc. pp. 400-403

15. Değirmencioğlu S. Angiogenic and prognostic impacts ofvascular endothelial growth factor, endothelin-1 and alpha-calcitonin gene-related peptide serum levels in advanced non-smallcell lung cancer patients. Pamukkale Unıversity, medical school department of internal diseases medical oncology. Thesis 2011.

16. Yang JR, Xu JY, Chen GC, et al. Post-diagnostic C-reactive protein and albumin predict survival in Chinese patients with non-small cell lung cancer: a prospective cohort study. Sci Rep 2019;9:8143.