Comparison of Five Survival Models: Breast Cancer Registry Data from Ege University Cancer Research Center

Beş Sağkalım Modelinin Karşılaştırılması: Ege Üniversitesi Kanser Araştırma Merkezinden Elde Edilen Meme Kanseri Kayıtlarına Ait Veriler

Elvan AKTÜRK HAYAT,[a]
Aslı SUNER,[b]
Burak UYAR,[a]
Ömer DURSUN,[b]
Mehmet N. ORMAN,[c]
Gül KİTAPÇIOĞLU, MD[d]

[a]Department of Statistics,
Ege University Faculty of Science,
[b]Department of Statistics,
Dokuz Eylül University Faculty of Arts and Sciences,
[c]Department of Biostatistics,
Ege University Faculty of Medicine,
[d]Ege University Cancer Research Center,
İzmir

Geliş Tarihi/*Received:* 26.11.2009
Kabul Tarihi/*Accepted:* 05.08.2010

Summary of this article was presented as a oral presentation by the title of "Comparison of Five Survival Models on Breast Cancer Registry Data from Ege University Cancer Research Center (EUCRC)" at the 5th Conference of the Eastern Mediterranean Region of the International Biometric Society (EMR-IBS), 10-14 May, 2009, İstanbul, Turkey.

Yazışma Adresi/*Correspondence:*
Elvan AKTÜRK HAYAT
Ege University Faculty of Science,
Department of Statistics, İzmir,
TÜRKİYE/TURKEY
elvan.akturk@ege.edu.tr

ABSTRACT Objective: In this study, we aimed to compare the results of the survival analysis of the patients with breast cancer using Weibull, Gamma, Gompertz, Log-Logistic and Log-Normal parametric models. **Material and Methods:** The data obtained from 5457 patients with breast cancer from Ege University Cancer Research Centre between 1992 and 2007 was used in this study. The patients were divided into two groups with respect to their ages, they were divided into two groups as 49 and below and 50 and above. The Log rank test was applied to compare the survival curves of the two age groups obtained by Kaplan Meier method. A survival analysis was conducted by using Weibull, Gamma, Gompertz, Loglogistic and Lognormal distribution of parametric models. **Results:** Survival curves of two groups were compared by using a log-rank test and no statistical significant difference was found between the two groups. In the analysis of the survival periods using parametric models, the age variable is taken as the covariate. To determine the best model among parametric models, Akaike Information Criteria (AIC) was exploited. The results of the study revealed that the survival model found by the Gompertz distribution was the most appropriate one. **Conclusion:** By using AIC, the models obtained via Weibull, Loglogistic, Lognormal, Gamma and Gompertz were compared and the most suitable model for the obtained data distribution was determined. Although the AIC values for the five distributions in question were very close to each other, the Gompertz distribution, which had the lowest AIC value, was determined as the most suitable model.

Key Words: Survival analysis, breast neoplasms, Weibull distribution, Loglogistic distribution, Lognormal distribution, Gamma distribution, Gompertz distribution

ÖZET Amaç: Bu çalışmada meme kanserli hastalara ait sağkalım analizlerinin sonuçlarını Weibull, Gamma, Gompertz, Log-Logistic ve Log-Normal parametrik modelleri kullanarak karşılaştırmayı amaçladık. **Gereç ve Yöntemler:** Çalışmada Ege Üniversitesi Kanser Araştırma Merkezi' nde kayıtlı 1992 ile 2007 yılları arasındaki 5457 meme kanserli hastadan elde edilen veriler kullanılmıştır. Hastalar 50 yaş ve üstündekiler ile 50 yaş altındakiler olmak üzere iki gruba ayrıldılar. Kaplan Meier metodu ile elde edilen iki gruba ait sağkalım eğrileri karşılaştırılırken Log rank testi uygulanmıştır. Meme kanserli hasta verileri ile birlikte Weibull, Gamma, Gompertz, Loglogistic and Lognormal parametrik modelleri kullanılarak bir sağkalım analizi yapılmıştır. **Bulgular:** İki grubun sağkalım eğrileri long-rank testi kullanılarak karşılaştırılmış, istatistiksel olarak anlamlı fark bulunmamıştır. Parametrik testler kullanılarak yapılan sağkalım sürelerinin analizinde, yaş ortak değişken olarak ele alınmıştır. Parametrik modeller içerisinde en iyi olanı belirlemek üzere Akaike Bilgi Kriteri (ABK)'nden faydalanılmıştır. Çalışma sonuçları Gompertz dağılımı ile bulunan sağkalım modelinin en uygunu olduğunu ortaya koymuştur. **Sonuç:** ABK kullanılarak Weibull, Loglogistic, Lognormal, Gamma ve Gompertz ile elde edilen modeller karşılaştırılmış ve elde edilen veri dağılımı için en uygun model ele alınmıştır. Söz konusu beş dağılım hakkındaki ABK değerleri birbirine çok yakın bulunmuş ve ABK değeri en düşük olan Gompertz dağılımı, en uygun model olarak değerlendirilmiştir.

Anahtar Kelimeler: Sağkalım analizi, meme tümörleri, Weibull dağılımı, Loglojistik dağılım, Lognormal dağılım, Gamma dağılımı, Gompertz dağılımı

Turkiye Klinikleri J Med Sci 2010;30(5):1665-74

Breast cancer is the most frequent cancer in women while it is the second one in all humans. Although its prevalence changes in different societies, it is known that one out of 8-10 women in Western society would have breast cancer during her lifespan. The mortality rate of breast cancer was calculated as 3%. The prevalence of the disease is especially higher in North American and European countries than the rest of the world.[1]

Although 796 000 breast cancer cases and 314 000 deaths due to the disease are detected in a study conducted by the World Health Organisation (WHO) in 1990, 1 152 000 new cases and 411 000 deaths are detected in a study, conducted by the International Agency on Cancer for Research (IARC) under WHO administration in 2002. In this period of time, the incidence and mortality rates of breast cancer increased by 25%. The mortality rate of breast cancer is around 9.7%. Of all breast cancer cases, only 1% are men. According to all phases of the disease, the five-year survival rate has been reported as %73 in developed countries and %53 in developing countries. The significant difference between these two values can be explained by early diagnosis via mammography scans and better treatment in developed countries. Breast cancer mortality rate is 30% (190 000 deaths/636 000 cases) in developed countries; and 43% (221 000 / 514 000 cases) in under-developed countries.[2] According to the present data in our country, a prevalence of 20/100 000 for Eastern regions and 40-50/100 000 for Western regions may be determined.[3] The reason for this difference in the prevalence is the resemblance of the Western region's lifestyle to the Europeans. One of the four cancer types nests in breasts, and breast cancer is the main cause of death due to cancer.[4]

The factors that increase the risk of breast cancer are age over 50 years, presence of a close relative having breast cancer (if mother or sister has breast cancer the risk is multiplied by 2 or 3), another cancer being ascertained in the other breast, beginning of menstruation before age 12, absence of any pregnancies, and continuing to menstruate after age 50. In breast cancer where age is an important factor, 70% of the women diagnosed are over 50.[5]

This study aimed to compare the results of the survival analysis of the breast cancer patients by using Weibull, Gamma, Gompertz, Log-logistic and Log-Normal parametric models. Before conducting the parametric modelling, the patients were divided into two groups as 49-and-lower and 50-and-over, based on the critical age 50 for the breast cancer patients, and their survival rates were calculated using the Kaplan Meier Method.

MATERIAL AND METHODS

The data used in this study are obtained from Ege University Cancer Research Centre (EUCRC). The data used in the analyses belong to 5457 breast cancer patients who were diagnosed between 1992 and 2007. The patients were divided into two with respect to their ages as 49-and-lower and 50-and-over. The Log rank test was applied to compare the survival curves of the two age groups obtained by Kaplan Meier method. The survival analysis was conducted by using Weibull, Gamma, Gompertz, Loglogistic and Lognormal distribution of parametric models.

In order to determine the most suitable distribution to the survival periods, the accelerated failure time form of the Weibull, Lognormal, Loglogistic and Gamma distributions and the proportional bazard form of the Gompertz distribution were used. The models obtained via these five distributions were compared using the Akaike Information Criteria (AIC) and the most suitable model to the distribution of the data was determined.

LOGNORMAL DISTRIBUTION

The skewed distributions, where the average values are low, variances are high and the values are not negative such as the diversity of species, the distribution of the minerals in the crust of the earth, generally accord with Lognormal distribution. The lognormal distribution, related to the normal distribution but having the hypothesis that the random variable may only get positive values is usually used for economic data, the response data of the stimulant biologic materials, most types of survival data, distribution of the repair time of hardware, financial researches and studies on stock prices. When the examples of use of the lognormal distribution are examined, it is seen that among these examples there are geology, metallurgy, health, environment, atmospheric sciences, aerobiology, microbiology, plant physiology, ecology, food technology, linguistics, social sciences and economics.[6]

The theory of the lognormal distribution was characterised by McAlister in 1897. There is accordance to the lognormal distribution in many examples in the area of medicine. The studies on determining the survival in cancer, the studies of Sartwell on the length of the incubation time of the infectious diseases and the studies by Horner to determine the beginning age of the Alzheimer's disease may be among the examples of the medical studies.[6,7] Gaddum, with applications in biology, and Boag, with his studies on cancer have taken their place in literature.[8,9] In 1957, the history of the lognormal distribution, its features, estimation problems and its use in economics were examined in detail. By other researchers, the survival period distributions of Hodgkin's disease and many other diseases, like chronic leukemia, wereanalyzed via lognormal distribution, which is positively skewed and with survival period logarithms distributed normally.[10]

Tai et al. analyzed the limited-stage small-cell lung cancer patients with Kaplan-Meier curves, Cox proportional hazard model, Boag log-normal (cure-rate model) model and log-normal survival analysis methods.[11] Feinleib and MacMahon, in a study conducted on chronic lymphocytic and myeolocythic leukaemia patients, applied the lognormal distribution on chronic lymphocytic and myeolocythic leukaemia diagnosed in Caucasian patients lived in Brooklyn between 1943 and 1952.[12]

LOGLOGISTIC DISTRIBUTION

If the mortality ratio in a life analysis slowly decreases after it reaches to a maximum point over a finite period, it is suitable to use a non-monotonic failure rate distribution model on the life and lost time. For instance, in a study conducted on the curability of breast cancer, Lansgland found that the maximum point is reached in three years.[13]

In some of the data sets, it appeared that it is important to do analysis using suitable models such as lognormal, reverse gauss, loglogistic and Burr Type XII. However in cases where one comes across to censored data, using loglogistic distribution is mathematically more advantageous than other distributions. According to the study of Gupta et al, the loglogistic distribution is proved to be suitable in analysing survival data conducted by Cox, Cox and Oakes, Bennet, O'Quinley and Struthers. Gupta et al. used loglogistic distribution in survival analysis on lung cancer data in their studies. In their research, they estimated the point where the mortality ratio reached maximum level. They estimated the parameters of the distribution using the maximum likelihood estimate and bootstrap methods and they observed the proximity of the results.[14]

A study conducted by Byers in 1988 on the spreading ratio of HIV virus in San Fransisco between 1978 and 1986 indicated that loglogistic was the most suitable model among many distribution models to use with half censored data.[15] A study Zhou et al. conducted in 2007 emphasized that the maximum likelihood estimation was the most suitable method in estimating the parameters when performing analyses using loglogistic distribution on grouped data such as half-censored data.[16]

Loglogistic distribution is a continuous distribution for the random variable which is not negative in probability and statistics. Loglogistic analysis is used as a parametric model survival analysis where the ratio increases at the very beginning and later decreases.[10]

WEIBULL DISTRIBUTION

Weibull distribution which is a generalised version of the exponential distribution is widely used in modelling weather forecasts in meteorology, and defining the distribution of wind speed in radar modelling. Weibull distribution is preferred for performing survival data analysis in industrial engineering issues.[17] When the implementations in the discipline of medicine are examined, one may see that Weibull distribution is an important distribution model used in medicine since it is an flexible distribution that allows a monotonous increasing and decreasing of mortality ratio in patient groups. In a study carried out by Viscomi et al., the distribution of the survival period of childhood leukemia patients was analysed using the Weibull distribution.[18]

The correlation between the data and some distributions, including the Weibull distribution, is tested by Anderson-Darling test.[19] In a study conducted in Italy on the nationwide estimation of the cancer patients, some estimations were made for defining the parameters of the Weibull distribution.[20]

GAMMA DISTRIBUTION

Gamma distribution, for its compliance with the patient data, is a suitable distribution to use in survival distribution models. Additionally, it is used in queue systems, in raw material flow between manufacturing and deployment lines, in modelling the loads of the web service providers and risk management. When its use in the discipline of medicine is considered, it is seen that this distribution model is used in cirrhosis and chronic hepatitis patients. In a study conducted in 1983 on images obtained via hepatography in healthy adults by Galli et al., and in a thrombocyte survival study by Bolin et al. in 1986 this method was used.[21,22] Poon et al., used non parametric methods to assess the relationship between survival period (months) of the patients with nasopharyngeal tumor and different clinical characteristics in their study.[23]

The exponential distribution, which is a special version of the Gamma distribution, is used in cancer survival models.[24] In 1966 Zelen used the exponential distribution to model the survival periods in a study where he tried to develop a cure for cancer on a L1210 animal leukemia system.[25] In most of the clinical studies, the exponential model is used in defining the tumor development.[26] The exponential model is used in a study by Dewals and Bouckaert in 1985 on carrier bacteria and in a study by Sankrithi et al. in 1991 on nationwide infant deaths.[27,28]

GOMPERTZ DISTRIBUTION

Gompertz model, used frequently by medical researchers and biologists in modelling the mortality ratio data, was formulated by Benjamin Gompertz in 1825. Gompertz is a growth model and has been used in relation with tumor development. Ahuja and Nash showed that Gompertz distribution was, with a simple conversion, related to some distributions in the Pearson distributions family.[29]

According to Jaheen, Garg et al. obtained the maximum likelihood estimations of the Gompertz distribution parameters and Osman, in order to derive the composite Gompertz distribution which is assumed to accord with the Gamma distribution with one of the two parameters being random, used a Gompertz distribution with two parameters, worked on the features of the distribution and offered that it should be used in modelling the lifespan data analysing the survival ratio in heterogenic masses. Recently, some research and studies have been carried out on Gompertz distribution by Al-Hussaini et al.[30]

The hazard functions [h(x)], probability density functions [f(x)] and survival functions [S(x)] of the most frequently used parametric models, namely the Weibull, Gamma, Gompertz, Loglogistic and Lognormal distributions are given in Table 1.[31]

Parametric survival models are statistically more powerful than non-parametric or semi-parametric models. Two models used for correction of the covariate's effect on the survival function are accelerated failure time (AFT) and proportional hazard (PH) models.[32]

In AFT model, the natural logarithm of the survival period ln(t) is explained by a linear function of the covariates as $\ln t_j = x_j \beta + z_j$. Here, X is a vector of the covariates, β is the regression coefficient and z_j is the error which has the density of f( ). The distribution of the error term defines the regression model. For instance, if f( ) is a normal density function, the lognormal regression model, if f( ) is a loglogistic density, the loglogistic regression model is obtained. The effect of the AFT model is the change in the time scale by the $\exp(-x_j \beta)$ factor. With relation to these factors being greater

TABLE 1: The hazard functions, probability density functions and survival functions for the parametric distributions.

Distribution	Parameter	h(x)	f(x)	S(x)
Weibull	$\lambda > 0$ $\alpha > 0$	$\lambda \alpha x^{\alpha-1}$	$\lambda \alpha x^{\alpha-1} \exp(-\lambda x^{\alpha})$	$\exp(-\lambda x^{\alpha})$
Loglogistic	$\alpha, \lambda > 0$	$\frac{\alpha x^{\alpha-1}\lambda}{1+\lambda x^{\alpha}}$	$\frac{\alpha x^{\alpha-1}\lambda}{\left[1+\lambda x^{\alpha}\right]^2}$	$\frac{1}{1+\lambda x^{\alpha}}$
Lognormal	$\alpha > 0$	$\frac{f(x)}{S(x)}$	$\frac{\exp\left[-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right]}{x(2\pi)^{1/2}\sigma}$	$1-\Phi\left[\frac{\ln x-\mu}{\sigma}\right]$
Gompertz	$\lambda > 0$ $\alpha \in (-\infty, \infty)$	$\lambda \exp(\alpha x)$	$\lambda \exp(\alpha x) \exp\left(\frac{\lambda}{\alpha}(1-\exp(\alpha x))\right)$	$\exp\left[\frac{\lambda}{\alpha}\left(1-e^{\alpha x}\right)\right]$
Gamma	$\alpha, \beta, \lambda > 0$	$\frac{f(x)}{S(x)}$	$\frac{\alpha \lambda^{\beta} x^{\alpha\beta-1} \exp\left(-\lambda x^{\alpha}\right)}{\Gamma(\beta)}$	$1-\mathrm{I}\left[\lambda x^{\alpha}, \beta\right]$

TABLE 2: The survival functions related to the AFT and PH models for the five distributions used.

Distribution	Metric	Survival Function	Parametric Expression	Ancillary Parameter
Weibull	*PH*	$\exp\left(-\lambda_j t_j^P\right)$	$\lambda_j = \exp\left(x_j\beta\right)$	*p*
	AFT	$\exp\left(-\lambda_j t_j^P\right)$	$\lambda_j = \exp\left(-p x_j \beta\right)$	*p*
Gompertz	*PH*	$\exp\left\{-\lambda_j \gamma^{-1}\left(e^{\gamma t_j}-1\right)\right\}$	$\lambda_j = \exp\left(x_j\beta\right)$	γ
Lognormal	*AFT*	$1-\Phi\left\{\frac{\ln\left(t_j\right)-\mu_j}{\sigma}\right\}$	$\mu_j = x_j\beta$	σ
Loglogistic	*AFT*	$\left\{1+\left(\lambda_j t_j\right)^{1/\gamma}\right\}^{-1}$	$\lambda_j = \exp\left(-x_j\beta\right)$	γ
Gamma $\kappa > 0$	*AFT*	$1-I\left(\lambda, u\right)$	$\mu_j = x_j\beta$	σ, κ
Gamma $\kappa = 0$	*AFT*	$1-\Phi\left(z\right)$	$\mu_j = x_j\beta$	σ, κ
Gamma $\kappa < 0$	*AFT*	$I\left(\gamma, u\right)$	$\mu_j = x_j\beta$	σ, κ

or smaller than 1, the time becomes accelerated or decelerated.

In PH model, the covariates have a proportional effect on the hazard function. In the function, $h(t_j) = h_0(t) g(x_j)$, $h_0(t)$ is the basic hazard function, $g(x_j)$ is a non-negative function of the covariates and expressed as $g(x_j) = \exp(x_j\beta)$. The survival functions related to the AFT and PH models for the five distributions used in the analyses are summarised (Table 2).[33]

When comparing the models related to the breast cancer patient data and determining the most suitable model to the distribution of the data, the Akaike information criteria (AIC) is used.[33] AIC suggested by Akaike in 1974, is an information assessment criteria for comparing the models.[34] The AIC value is calculated as $AIC = -2(\log likelihood) + 2(c + p + 1)$. In this formula, c is the number of the covariates in the model and p is the parameter number.

The analyses were conducted using the STATA (version 10.0) statistical software.

RESULTS

Analysis of gender distribution of the 5698 breast cancer patients revealed that 95 (1.7%) of them were men and 5568 (98.3%) of them were women. When the survival periods were examined, it was seen that there were 146 missing observations. Missing observations and the data of the male patients were excluded from the analysis. Data revealed that 149 of the patients (2.7%) had died. The average age of the 5457 female patients diagnosed as breast cancer was calculated as 51.70 ± 12.48 years. When the diagnosis methods of the patients were examined, it was seen that 5279 of them (96.7%) diagnosed by biopsy of primary, 68 of them (1.2%) by metastasis, three of them (0.1%) were diagnosed at autopsy, 46 of them (0.8%) were diagnosed by cytology/haematology, and 45 of them (0.8%) were diagnosed in a clinical research before 2006. The diagnosis method of 16 patients was registered as unknown. The tumor was localized in mamilia/areola region in 117 of them (2.1%), 128 of them (2.3%) had tumor in the middle part of the breast, 381 of them (7%) had tumor in the upper inner breast quadrant, 212 of them (3.9%) had tumor in the lower inner breast quadrant, 1814 of them (33.2%) had tumor in the upper outer breast quadrant, 291 of them (5.3%) had tumor in the lower outer breast quadrant, 14 of them (0.3%) had tumor in the breast's axillary node, 718 of them (13.2%) had an extensive lesion in the breast, 1767 of them (32.4%) had tumor in an unknown position in the breast, 13 of them (0.2%) had tumor in labium majus, one of them had tumor in endocervix and one of them in endometrium. Five thousand forty seven of the patients (96.1%) received treatment, 96 of them (1.8%) did not receive any treatment, and one of them received symptomatic treatment. The treatment conditions of 114 patient (2.1%) were unknown. In the studies conducted, it is set forth that breast cancer increases during the active productive phase until menopause and shows a proportional decrease after ages 50-55.[35] Similar studies show that, in men the prevalence of breast cancer is approximately 1.0%.[36] This research group has findings pertinent to the literature.

The patients were divided into two groups; 49-and-lower (age1) and 50-and-over (age2). Sixty seven of 2589 (2.6%) patients ≤49 years of age and 82 of 2868 (2.9%) patients ≥50 years of age of died.

In the survival function graph obtained by Kaplan-Meier which is a non-parametric technique, it is seen that the estimated survival probability of the 49-and-lower group patients was higher than the one estimated in the 50-and-over group (Figure 1). As a result of the log rank test used in comparison of the survival curves belonging to different age groups, it is observed that there was no any statistically significant difference between the survival curves of the breast cancer patients in 49-and-lower and 50-and-over groups ($\chi^2 = 2.11$, $p = 0{,}1459$).

The hazard function graphs for age groups, created in accordance with Lognormal, Loglogistic, Weibull, Gamma and Gompertz distributions are

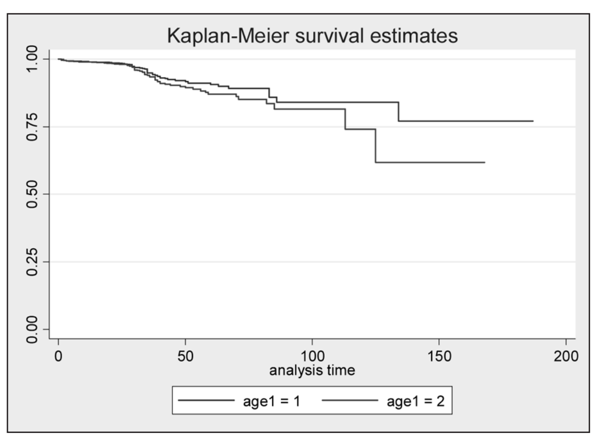

FIGURE 1: Kaplan –Meier survival estimates.

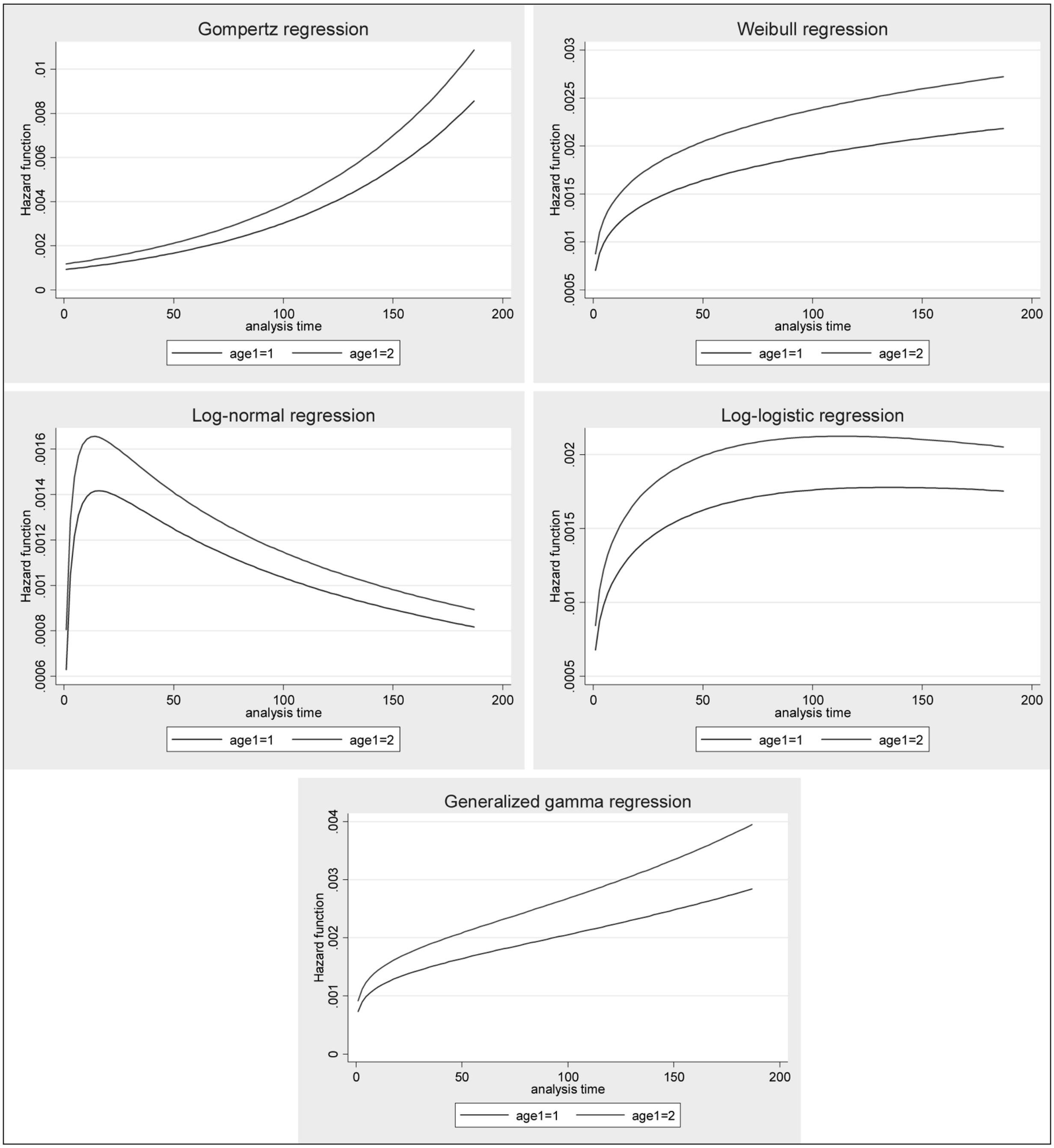

FIGURE 2: The hazard functions related to five distributions used in the study.

given in Figure 2. When the hazard function graphs obtained as a result of the analyses conducted are examined, it is seen that the mortality risk of the 49-and-lower group patients are lower than the ones in the 50-and-over group (Figure 2).

The survival function graphs for age groups, created in accordance with Lognormal, Loglogistic, Weibull, Gamma and Gompertz distributions are given in Figure 3.

When the survival function graphs are examined, it is observed that the survival probability of the patients in the 49-and-lower group is higher than the ones in the 50-and-over group. The graph of the Gompertz model exhibits more clearly the

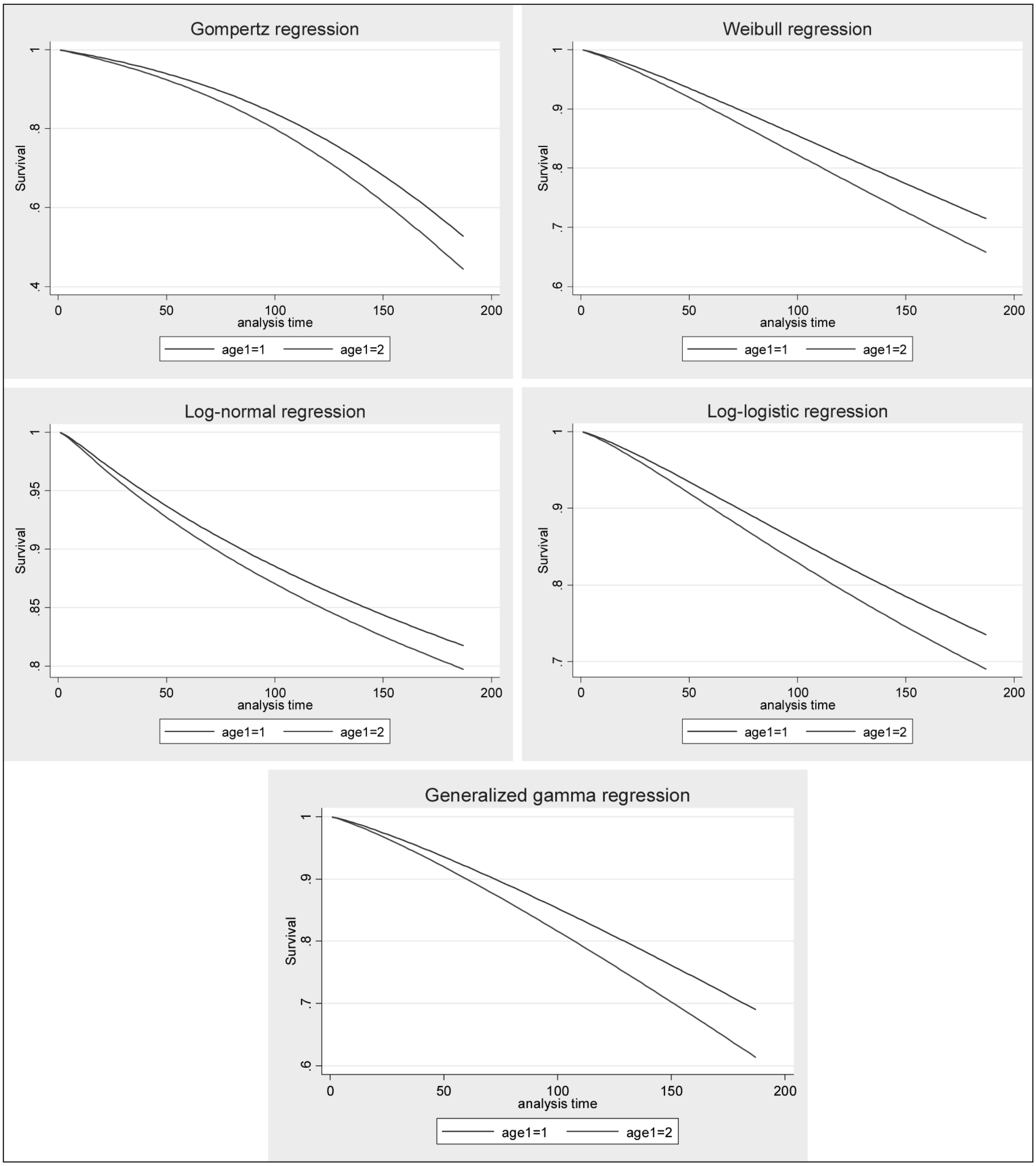

FIGURE 3: The hazard functions related to five distributions used in the study.

curvilinearity observed in the Kaplan-Meier curve (Figure3).

When analysing the survival periods using parametric models, the age variable is taken as the covariate. The parameter obtained as a result of the analysis, the loglikelihood and the AIC values are presented in Table 3. As lower values of AIC suggest a better model, the model obtained by using Gompertz model is determined as the most suitable model for the 49-and-lower and 50-and-over groups among five models.

The time ratio (TR) values were calculated as a result of the AFT function and the hazard ratio (HR) values were calculated as a result of the PH form and their confidence intervals were given. The time ratio and hazard ratio values obtained for

TABLE 3: Parameter, loglikelihood and AIC values.

Parameter	Weibull	Lognormal	Log-logistic	Gompertz	Gamma
Age	-0,182	-0,156	-0,178	0,239	-0,187
Constant	6,311	7,305	6,233	-7,237	6,193
Ancillary	0,822	2,117	0,806	0,012	0,423
Loglikelihood	-722,009	-734,432	-722,853	-719,377	-721,294
AIC	1450,018	1474,863	1451,706	1444,753	1450,588

all distributions related to the age variable were not found statistically significant ($p>0.05$). This shows that the age variable does not pose a risk factor for breast cancer patients (Table 4).

TABLE 4: TR and HR values for age in five distributions.

Distribution	TR(95%CI)	p-value
Weibull AFT	.83 (.64 1.09)	.180
Lognormal AFT	.86 (.63 1.17)	.324
Loglogistic AFT	.84 (.64 1.10)	.196
Gamma AFT	.83 (.64 1.08)	.162
Distribution	HR (95%CI)	p-value
Gompertz PH	1.27 (.92 1.76)	.148

CONCLUSION

In this study, we aimed to compare the results of the survival analysis of breast cancer patients using Weibull, Gamma, Gompertz, Loglogistic and Lognormal models. The data obtained from EUCRC between 1992 and 2007 were used. The data used in this study comprises of only female patients' data.

In conclusion, as a result of the analyses depending on the parameter estimates, age variable was not found as a risk factor. By using AIC, the models obtained via Weibull, Loglogistic, Lognormal, Gamma and Gompertz models were compared and the most suitable model for the obtained data distribution was determined. Although the AIC values of the five distributions in question were very close to each other, the Gompertz distribution, which had the lowest AIC value, was determined as the most suitable model. It was concluded that the Gompertz distribution model was more suitable for these survival data obtained from EUCRC.

REFERENCES

1. Stewart BW, Kleihues P. The Global Burden of Cancer. World Cancer Report. 1st ed. Lyon: IARC Press; 2003. p.12-7.
2. Özmen V. [Screening and registering programs for breast cancer in Turkey and in the world]. Journal of Breast Health 2006;2(2):55-8.
3. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics 2002. CA Cancer J Clin 2005;55(2):74-108.
4. Shyyan R, Masood S, Badwe RA, Errico KM, Liberman L, Ozmen V, et al. Breast cancer in limited-resource countries: diagnosis and pathology. Breast J 2006;12(Suppl 1):S27-37.
5. Robb C, Haley WE, Balducci L, Extermann M, Perkins EA, Small BJ, et al. Impact of breast cancer survivorship on quality of life in older women. Crit Rev Oncol Hematol 2007;62(1):84-91.
6. Limpert E, Stahel WA, Abbt M. Log-normal distributions across the sciences: Keys and clues. BioScience 2001;51(5):341-52.
7. Horner RD. Age at onset of Alzheimer's disease: clue to the relative importance of etiologic factors? Am J Epidemiol 1987;126(3):409-14.
8. Gaddum JH. Log normal distributions. Nature London 1945;156(3964):463-6.
9. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of the Royal Statistical Society Series B 1949;11(1):15-53.
10. Lee ET, Wang JW. Some well-known parametric survival distributions and their applications. Statistical Methods for Survival Data Analysis. 3rd ed. Hoboken, New Jersey: John Wiley&Sons, Inc; 2003. p.134-62.
11. Tai P, Chapman JA, Yu E, Jones D, Yu C, Yuan F, et al. Disease-specific survival for limited-stage small-cell lung cancer affected by statistical method of assessment. BMC Cancer 2007;7:31.
12. Feinleib M and Macmahon B. Variation in the duration of survival of patients with the chronic leukemias. Blood 1960; 15(3): 332-49.
13. Langlands AO, Pocock SJ, Kerr GR, Gore SM. Long term survival of patients with breast cancer: a study of curability of the disease. Br Med J 1979; 2(6200):1247-51.
14. Gupta RC, Akman O, Lvin S. A study of log-logistic model in survival analysis. Biom J 1999;41(4):431-43.
15. Byers RH Jr., Morgan WM, Darrow WW,Doll L, Jaffe HW, Rutherford G, et al.Estimating AIDS infection rates in the San Francisco cohort. AIDS 1988; 2(3):207-10.

16. Zhou YY, Mi J, Guo S. Estimation of parameters in logistic and log-logistic distribution with grouped data. Lifetime Data Anal 2007; 13(3):421-9.
17. Weibull W. A statistical distribution functions of wide applicability. J Appl Mech 1951; 18(2):293-7.
18. Viscomi S, Pastore G, Dama E, Zuccolo L, Pearce N, Merletti F, et al. Life expectancy as an indicator of outcome in follow-up of population-based cancer registries: the example of childhood leukemia. Ann Oncol 2006;17(1): 167-71.
19. Akram M, Aman Ullah M, Taj R. [Survival analysis of cancer patients using parametric and non-parametric approaches]. Pakistan Vet Journal 2007;27(4):194-8.
20. Inghelmann R, Grande E, Francisci S, De Angelis R, Micheli A, Verdecchia A, et al. National estimates of cancer patients survival in Italy: A model-based method. Tumori 2005;91(2)109-15.
21. Galli G, Maini CL, Salvatori M, Andreasi F. A practical approach to the hepatobiliary kinetics of 99mTc-HIDA: clinical validation of the method and a preliminary report on its use for parametric imaging. Eur J Nucl Med 1983; 8(7):292-8.
22. Bolin RB, Greene JR. Stored platelet survival data analysis by a gamma model. Transfusion 1986;26(1):28-30.
23. Poon D, Yap SP, Wong ZW, Cheung YB, Leong SS, Wee J, et al. Concurrent chemoradiotherapy in locoregionally recurrent nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys 2004;59(5):1312-8.
24. Lee ET, Go OT. Survival analysis in public health research. Annu Rev Public Health 1997;18:105-34.
25. Zelen M. Application of exponential models to problems in cancer research. J Roy Statist Soc Series A 1966;129(3):368-98.
26. Spratt JA, Von Fournier D, Spratt JS, Weber EE. Decelerating growth and human breast cancer. Cancer 1993;71(6):2013-9.
27. Dewals P, Bouckaert A. Methods for estimating the duration of bacterial carriage. Int J Epidemiol 1985;14(4):628-34.
28. Sankrithi U, Emanuel I, Van Belle G. Comparison of linear and exponential multivariate models for explaining national infant and child mortality. Int J Epidemiology 1991; 20(2):565-70.
29. Ahuja JC, Nash SW. The generalized Gompertz-Verhulst family of distributions. Sankhya Ser A 1967;29(2):141-56.
30. Jaheen ZF. Prediction of progressive censored data from the Gompertz model. Communications in Statistics-Simulation and Computation 2003; 32(3):663-76.
31. Ata N, Karasoy D, Sözer MT. [Parametric and nonparametric survival models for nonproportional hazards]. İstatistikciler Dergisi 2008;1(3):125-34.
32. Kleinbaum DG, Klein M. Parametric survival models. In: Gail M, Krickeberg K, Samet J, Tsiatis A, Wong W, eds. Survival Analysis-A Self-Learning Text. 2nd ed. New York: Springer; 2005. p.260-312.
33. Ata N, Karasoy D, Sözer MT. [Stratified cox regression model for non-proportional hazards and an application on breast cancer patients]. Turkiye Klinikleri J Med Sci 2008; 28(3):327-32.
34. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974;19(6):716-23.
35. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. Lancet Oncol 2001; 2(3):133-40.
36. Weiss JR, Moysich KB, Swede H. Epidemiology of male breast cancer. Cancer Epidemiol Biomarkers Prev 2005;14(1):20-6.