

Comparison of Survival Models for Analyzing Prognostic Factors in Gastric Cancer Patients

Danial Habibi¹, Mohammad Rafiei¹, Ali Chehrei², Zahra Shayan^{3*}, Soheil Tafaqodi¹

Abstract

Objective: There are a number of models for determining risk factors for survival of patients with gastric cancer. This study was conducted to select the model showing the best fit with available data. **Methods:** Cox regression and parametric models (Exponential, Weibull, Gompertz, Log normal, Log logistic and Generalized Gamma) were utilized in unadjusted and adjusted forms to detect factors influencing mortality of patients. Comparisons were made with Akaike Information Criterion (AIC) by using STATA 13 and R 3.1.3 softwares. **Results:** The results of this study indicated that all parametric models outperform the Cox regression model. The Log normal, Log logistic and Generalized Gamma provided the best performance in terms of AIC values (179.2, 179.4 and 181.1, respectively). On unadjusted analysis, the results of the Cox regression and parametric models indicated stage, grade, largest diameter of metastatic nest, largest diameter of LM, number of involved lymph nodes and the largest ratio of metastatic nests to lymph nodes, to be variables influencing the survival of patients with gastric cancer. On adjusted analysis, according to the best model (log normal), grade was found as the significant variable. **Conclusion:** The results suggested that all parametric models outperform the Cox model. The log normal model provides the best fit and is a good substitute for Cox regression.

Keywords: Cox regression- Parametric models- AIC-Gastric cancer

Asian Pac J Cancer Prev, **19** (3), 749-753

Introduction

One of the most common cancers throughout the world is gastric cancer (GC). The incidence of GC varies worldwide. The highest rates are in Eastern Asia and Central and Eastern Europe and South America. Moreover, the lowest rates are in Northern America and most parts of Africa (Torre et al., 2015).

GC is the fourth most common cancer and the second leading cause of cancer-related deaths all over the world. (Li et al., 2017). The highest mortality rate of GC in South-Central and Western Asia belongs to Iran (Rahman et al., 2014). Northern and northwestern regions are at high risks of GC in Iran, especially Ardebil, a northwestern province, which has the highest incidence (Malekzadeh et al., 2009).

Lauren classified GC as diffuse type and intestinal type of adenocarcinoma. Intestinal GC is more prevalence among males and older age groups and is connected probably to environmental factors. The diffuse is more common in younger age groups, which is frequent over females and has a worse prognosis than the intestinal (Ma et al., 2016).

Adenocarcinomas, divided histologically into

intestinal and diffuse histotypes, includes more than 95% of gastric cancers. 50-70% of the cases are the intestinal types which are the most common variants of gastric cancers (Rugge et al., 2015).

Gastric cancer is usually treated by surgery, radiotherapy and chemotherapy. Surgery is the prime treatment for gastric cancer at the early stages. The supplementary treatments will be radiotherapy, chemotherapy and surgery. Radiotherapy and chemotherapy and surgical procedures are also used at advanced stages (Zare et al., 2015). In Iran, at least 80 percent of GC's are diagnosed at advanced levels (Malekzadeh et al., 2009).

Due to the censored data and skewness time, the common statistical methods cannot be used in analyzing these data (Collett, 2015). Frequently, the goal of medical studies on GC is computing the survival time and influential factors on patient's survival time based on the demographic and clinical information. For this purpose, generally, Cox proportional hazard model is used as a semi-parametric model. The proportional hazard condition is an important assumption in Cox regression.

Researchers in medical sciences often prefer semi-parametric models to the parametric ones because of their minimal assumptions. In contrast to semi-parametric

¹Department of Biostatistics, Faculty of Medicine, Arak University of Medical Sciences, ²Pars Clinicopathology Clinicopathology Laboratory, Arak, ³Trauma Research Center, Department of Community Medicine, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran. *For Correspondence: shayan@sums.ac.ir

model, a parametric survival model assumes that the survival time follows a known distribution and the effect of covariates on survival time is determined directly. The parametric models give more precise estimates of the quantities of interest than semi-parametric or non-parametric models. Of course, an inappropriate parametric model provides a wrong estimate of interest quantity (Klein and Moeschberger, 2003; Kleinbaum and Klein, 2005).

Parametric models have been extensively used in fitting survival data for GC patients. For example, Pourhoseingholi et al., (2011) showed that all parametric models outperformed the Cox model. Pourhoseingholi et al., (2009) indicated that log normal model has a better performance compared to Cox model. Ghorbani Gholiabad et al., (2014) concluded that Weibull, Log normal and Log logistic models were better than Cox model, and among all parametric models, the Log normal was the best one. Zhu et al., (2011) proved that Weibull model can perform better than Cox model. Zare et al., (2015) represented that the parametric model outperformed in comparison to Cox regression model and among all parametric models, the Exponential and Weibull models were the best ones. Wang et al., (2011) indicated that the Log normal model performed better than Cox regression model. As an example, this scenario was utilized for other types of diseases (Hashemian et al., 2013; Adelian et al., 2015; Shayan et al., 2014).

Therefore, our aim in this study, were comparing the survival time of gastric cancer patients based on parametric models (Exponential, Weibull, Gompertz, Log normal, Log logistic and Generalized Gamma (GG)) and the Cox regression model to determine the best-fitted model. Also, we examined the effect of several predictors on patient's survival time.

Materials and methods

Participants study and data collection

We collected a data set of gastric cancer patients at Khansari, Valiasr, Ghods and Emam Khomeini Hospital in Arak, in addition to Alzahra and Omid Hospital in Isfahan from 2000 to 2010. According to the completely recorded information of all hospital files, 104 files were finally eligible for conducting this study. Sampling method was convenience. The measured variables of data included age, sex, the location of the tumor, the tumor size, the presence or absence of ulcers in the stomach area, the type of surgery, the type of stomach cancer, the largest diameter of the metastatic nest, the largest nest of the metastatic lymph nodes, the largest diameter of LM, number of involved LN and the largest ratio of metastatic nest to lymph node.

Statistical Analysis

The death and survival of the patients were considered as the desired event and censor respectively. The event time was defined as the interval between diagnosis and a death or censoring (in months).

We used Kaplan-Meier estimator to estimate the survival function and log rank test to compare the

survival distributions, as well as to investigate the relationship between one or more covariates and time using proportional hazard regression. On the other hand, we have considered parametric models such as Exponential, Weibull, Gompertz, Log normal, Log logistic and Generalized Gamma.

To study life time data, one of the flexible parametric models is Weibull distribution. Exponential distribution is a special case of the Weibull distribution. These distributions included both PH and AFT models. The Weibull distribution is proper for modeling data with monotone hazard rates, whereas the Exponential distribution is suitable for data with constant hazard rates.

The Gompertz model is parameterized only in the PH form, and is appropriate for modeling data with monotone hazard rates. The log normal and Log logistic models are suitable for modeling data with non-monotone hazard rates and are implemented only as AFT models. In many cases, the results of regression models were similar to each other.

The Generalized Gamma model is commonly used as an appropriate model because of its extremely flexible hazard function. This model is implemented only in the AFT form. Weibull, Exponential and log normal distributions are special cases of the Generalized Gamma model.

To compare the semi-parametric and parametric models we used Akaike Information Criterion (AIC), a measure of the goodness of fit of an estimated statistical model, to choose the best model knowing that a lower AIC indicates the better likelihood.

Schoenfeld residuals (ph test) and log(-log s(t)) against log(t) plots were drawn for each variable to examine a proportional hazard model of the Cox regression model. Calculations were carried out by STATA 13 and R 3.2.5 statistical softwares.

Results

Patient demographics, clinical and pathology studies

Table 1 and 2 indicate the clinical, demographical and pathological characteristics of patients. A total number of 104 patients with gastric cancer from 2002 to 2010 entered this study. Of the 104 patients, 75 (71.2%) were male and 29 (27.9%) were female. The average age of participants was 62.9 (SD=12.6) years. Of 104 patients, 67 (67.7%) experienced death, and the others (32.3%)

Table 1. Clinical, Demographic and Pathologic Observations of Patients with Gastric Cancer

Variable	Mean	Standard Deviation
Age	62.91	12.62
Tumor size	5.84	2.72
Number of LN	6.82	4.89
Number of involved LN	2.76	3.23
Largest diameter of LM	7.15	5.91
Largest diameter of metastatic nest	5.75	9.63
Largest ratio of metastatic nest to lymph node	0.51	0.39

Table 2. Clinical and Demographic Characteristics of Patients with Gastric Cancer

Variables	N (%)	Median	Log-Rank test p-value
Sex			
Male	75 (72.1)	23	0.451
Female	29 (27.9)	24	
The location of tumor			
Body	45 (44.6)	23	0.519
Cardia	19 (18.8)	9	
Antrum	37 (36.6)	25	
Ulcers			
Present	78 (78)	24	0.828
Absent	22 (22)	20	
Type of gastrectomy			
Total	39 (37.5)	20	0.491
Partial	65 (62.5)	23	
Type of adenocarcinoma			
Intestinal	75 (72.1)	23	0.353
Signet	23 (22.1)	18	
Mucinous	6 (5.8)	27	
Stage			
Mild	59 (56.7)	29	0.004*
Severe	45 (43.3)	13	
Grade			
Badly	21 (22.3)	12	0.007*
Well	73 (77.7)	24	

* Significance level 5%

were censored (right censored). The minimum, maximum, mean and median of the survival time were 3, 78, 21.7 and 23 months respectively. Table 2 indicates median of survival time according to clinical and demographical characteristics. Survival times were compared by Log Rank test. The result of the Log Rank test indicated that the stage and grade were significant. Patients with well grade and mild stage had more survival time than others.

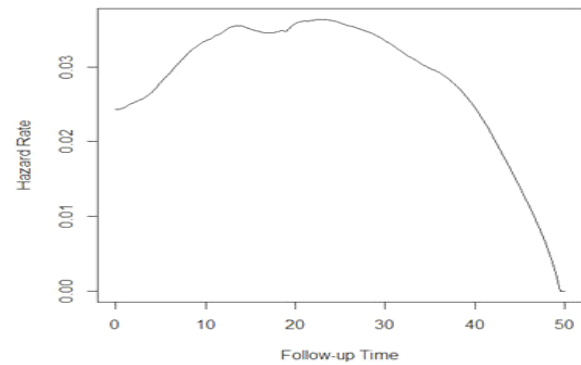


Figure 1. The Death Hazard Function of Gastric Cancer Patients (Follow up Time Based on Month)

Survival

The survival rates from 1 to 5 years were 0.49, 0.29, 0.21, 0.21 and 0.14 respectively. The hazard function is plotted in Figure 1. Up to 20 months after the diagnosis of the disease, the death hazard increases, then decreases. Therefore, the figure has a non-monotone hazard rate that suggests the appropriateness of the Log logistic, Log normal, and Generalized Gamma models.

After adjusting by covariates, Cox regression model and parametric models were used to determine patients' survival time. Being interested in investigating that whether the final results have been changed when all covariates were included in models, adjusted and unadjusted models were performed.

Table 3 shows unadjusted analysis of Cox regression model and parametric models. Based on table 3, stage, grade, the largest diameter of the metastatic nest, the largest diameter of LM and the number of involved LN were found as significant variables by all models.

Table 4 indicates the adjusted model of Cox regression and parametric models. Variables were entered into multiple Cox regression models if a variable with $p < 0.2$ was in a Cox model.

Table 3. Unadjusted Cox Regression and Parametric Models with Prognostic Factors

Model	Independent Variables	Stage	Grade	Tumor size	Dnodes ^a	Dnodes ^b	Inodes ^c	Nestnode ^d
Cox	HR ^f (SE ^g)	1.99 (0.490)	2.22 (0.137)	1.08 (0.500)	1.05 (0.010)	1.11 (0.022)	1.16 (0.038)	4.34 (1.49)
	P-VALUE	0.005	0.009	0.118	0.0001	0.0001	0.0001	0.0001
Exponential	HR (SE)	1.99 (0.486)	2.08 (0.145)	1.08 (0.049)	1.04 (0.009)	1.10 (0.020)	1.15 (0.035)	3.81 (1.25)
	P-VALUE	0.005	0.015	0.104	0.0001	0.0001	0.0001	0.0001
Weibull	HR (SE)	2.07 (0.51)	2.17 (0.139)	1.09 (0.051)	1.05 (0.010)	1.12 (0.022)	1.18 (0.038)	4.69 (1.57)
	P-VALUE	0.003	0.01	0.07	0.0001	0.0001	0.0001	0.0001
Gomperts	HR (SE)	1.98 (0.487)	2.04 (0.146)	1.08 (0.049)	1.04 (0.009)	1.11 (0.022)	1.16 (0.037)	4.05 (1.36)
	P-VALUE	0.005	0.016	0.109	0.0001	0.0001	0.0001	0.0001
Log logistic	TR ^h (SE)	0.51 (0.110)	0.44 (0.610)	0.93 (0.034)	0.96 (0.011)	0.92 (0.014)	0.88 (0.028)	0.26 (0.065)
	P-VALUE	0.002	0.003	0.056	0.0001	0.0001	0.0001	0.0001
Log normal	TR (SE)	0.55 (0.118)	0.46 (0.581)	0.93 (0.035)	0.96 (0.009)	0.92 (0.014)	0.88 (0.028)	0.28 (0.069)
	P-VALUE	0.005	0.004	0.06	0.0001	0.0001	0.0001	0.0001
GG ^e	TR (SE)	0.58 (0.134)	0.495 (0.542)	0.93 (0.035)	0.96 (0.011)	0.92 (0.015)	0.05 (0.022)	0.28 (0.070)
	P-VALUE	0.019	0.006	0.056	0.0001	0.0001	0.033	0.0001

a, The largest diameter of metastatic nest; b, The largest diameter of LM; c, The number of involved LN; d, The largest ratio of metastatic nest to lymph nodes; e, Generalized Gamma; f, Hazard Ratio; g, Standard Error; h, Time Ratio

Table 4. Adjusted Cox Regression and Parametric Models with Prognostic Factors

Model	Independent Variables	Stage	Grade	Tumor size	Dnests ^a	Dnodes ^b	Inodes ^c	Nestnode ^d
Cox	HR ^f (SE ^g)	0.89 (0.297)	2.22 (0.171)	1.04 (0.054)	1.04 (0.016)	1.05 (0.033)	1.05 (0.067)	1.35 (0.901)
AIC=378.40	P-VALUE	0.721	0.036	0.4	0.017	0.139	0.426	0.647
Exponential	HR (SE)	0.95 (0.313)	1.72 (0.213)	1.03 (0.051)	1.02 (0.015)	1.04 (0.034)	1.05 (0.064)	1.22 (0.809)
AIC=194.66	P-VALUE	0.866	0.137	0.574	0.088	0.177	0.411	0.766
Weibull	HR (SE)	0.83 (0.278)	2.04 (0.184)	1.05 (0.054)	1.04 (0.016)	1.06 (0.034)	1.08 (0.070)	1.21 (0.810)
AIC=184.18	P-VALUE	0.501	0.059	0.352	0.02	0.052	0.242	0.773
Gomperts	HR (SE)	0.87 (0.296)	1.72 (0.212)	1.03 (0.052)	1.03 (0.015)	1.06 (0.034)	1.07 (0.067)	1.19 (0.798)
AIC=194.02	P-VALUE	0.692	0.134	0.499	0.068	0.097	0.283	0.791
Log logistic	TR ^h (SE)	0.93 (0.210)	0.51 (0.497)	0.98 (0.031)	0.98 (0.008)	0.98 (0.023)	0.99 (0.037)	0.65 (0.290)
AIC=179.43	P-VALUE	0.759	0.009	0.508	0.023	0.324	0.724	0.331
Log normal	TR (SE)	0.93 (0.219)	0.56 (0.462)	0.97 (0.032)	0.98 (0.010)	0.97 (0.025)	0.98 (0.041)	0.74 (0.256)
AIC=179.16	P-VALUE	0.764	0.027	0.327	0.07	0.222	0.645	0.536
GG ^e	TR (SE)	0.92 (0.232)	0.56 (0.465)	0.97 (0.033)	0.98 (0.011)	0.97 (0.025)	0.98 (0.042)	0.74 (0.360)
AIC=181.12	P-VALUE	0.731	0.027	0.319	0.085	0.237	0.692	0.538

a, The largest diameter of metastatic nest; b, The largest diameter of LM; c, Number of involved LN, d, The largest ratio of metastatic nest to lymph node; e, Generalized Gamma; f, Hazard Ratio; g, Standard Error; h, Time Ratio

First, we used AIC to select best models. The lowest AIC leads us to identify the best one. According to the results, all parametric models outperform in comparison to Cox regression model and among parametric models, Log normal, Log logistic and Generalized Gamma (AIC=179.16, 179.43 and 181.12) were the best models respectively.

After entering significant variables in adjusted survival, the grade was significant by Log normal ($p=0.027$), Log-logistic ($p=0.009$), GG ($p=0.027$) and Cox ($p=0.036$). Also, the largest diameter of the metastatic nest was significant by Cox ($p=0.017$).

Based on the best model (log normal), patients with severe grade compared to those with a good grade (TR=0.56, $p=0.027$) experienced death faster and on the basis of Cox model, the hazard rate of death of patients with severe grade is 2.22 times more than those with good grade (HR=2.22, $p=0.036$). In Cox model, one unit increase in the largest diameter of the metastatic nest, increases the hazard rate of death of patients by 4% (HR=1.04, $p=0.017$).

Based on the Log logistic model, the grade and the largest diameter of the metastatic nest were significant variables. In this case, one unit increase in the largest diameter of metastatic nest decreases the time of death after diagnosis by 2% (TR=0.98).

Discussion

Frequently, the medical science researchers are interested in using Cox proportional model instead of parametric models. The parametric model has a more accurate estimate if the condition of a parametric model meets. Few studies have been carried out on patients with gastric cancer, especially in Iran.

Our aim in this study was to identify prognostic factors using parametric models and semi-parametric method and to determine the best fitting model. Hence, we used common parametric models such as Exponential, Weibull,

Gompertz, Log normal, Log logistic and Generalized Gamma. Finally to identify the best model, we utilized AIC index knowing that the lowest AIC is the best model for fitting data.

Our results showed that the survival rates from 1 to 5 years were 0.49, 0.29, 0.21, 0.21 and 0.14, respectively. This means that more than half of deaths occurred within the first year after diagnosis and 30% of deaths occurred during the second year. Also, 5-years survival rate was 14% that was lower than other studies (Ansari et al., 2011; Zare et al., 2014). Veisani and Delpisheh (2016) have done a systematic review and a meta-analysis study to obtain one-year and five-years survival rates in Iranian gastric cancer patients. They reported the overall survival rates of 0.52, 0.31, 0.24, 0.22 and 0.15, respectively. Our results were consistent to their report.

We found lognormal, log-logistic and Generalized Gamma (the AIC in the lognormal is very close to log-logistic model) as the best parametric models respectively. It is consistent to the studies of Pourhoseingholi et al., (2011), Pourhoseingholi et al., (2009) and Ghorbani Gholiabad et al., (2014). Also, our results were different from the studies of Zhu et al., (2011), Zare et al., (2015) and Nardi and Schemper (2003).

In unadjusted analysis, the results indicated that the stage, the grade, the largest diameter of the metastatic nest, the largest diameter of LM, the number of involved LN and the largest ratio of the metastatic nest to lymph node were variables influencing the survival of patients with GC.

Zhu et al., (2011) showed that the results of Weibull model can be more precise than Cox model. They found that the Histologic grade was significant. Pourhoseingholi et al., (2011) demonstrated that the parametric model was better than the Cox regression for the fitting model. They found that the extent of wall penetration and presence of pathological distance metastasis were potential risk factors for death. Pourhoseingholi et al., (2009) found that log normal performed better than the Cox model. Distance

of metastasis, the extent of wall penetration, the tumor size and the histology type were significant prognostic factors based on log normal model. Wang et al., (2011) indicated that according to AIC, the log normal model performed better than Cox regression model. Age at diagnosis time, past history, surgical curative degree, Borrmann type, Lauren's classification, pT stage, and pN stage were significant prognostic factors in both log normal and Cox models. Moreover, cancer location, distant metastasis status, and histological types were found to be significant prognostic factors in log normal results alone.

Another study by Ghorbani Gholiabad et al., (2014) showed that the log normal was the best model in gastric cancer and primary progress of the disease, type of treatment and metastasis were significant variables. Another study by Zare et al., (2015) has been conducted. They indicated that Exponential and Gompertz were the best parametric models compared to Cox's regression. The result of some studies also showed the Weibull model as the best parametric model (Nardi and Schemper, 2003; Zhu et al., 2011).

Our results indicated that the parametric survival model was a more suitable analysis due to some reasons. The hazard function shape revealed that parametric models were more appropriate than semi-parametric ones. Table 4 showed that AIC for parametric models was lower than that for semi-parametric models.

Although most of researchers are interested in using the Cox regression model, the parametric model is also useful in this condition if the states of parametric model exist. In Cox regression model, the proportional hazard condition is an important assumption. The parametric model does not require this assumption and will be a credible alternative for Cox regression model.

Acknowledgments

This article was extracted from the research of student Mr. Soheyl Tafaghodi for the master degree in Biostatistics at Arak University of Medical Sciences.

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 Organ Transplant Med*, **6**, 119.
- Ansari J, Chehrei A, Amini M, et al (2011). The prognostic significance of Her2-Neu over expression in gastric carcinomas. *Iran J Cancer Prev*, **4**, 170.
- Collett D (2015). *Modelling survival data in medical research*, CRC press.1-1.
- Ghorbani Gholiabad S, Yazdani Charati J, Jan Babaie G (2014). Evaluation of parametric and semi-parametric models in survival analysis of patients with gastric cancer. *J Mazandaran Uni Med Sci*, **24**, 11-8.
- Hashemian AH, Beiranvand B, Rezaei M, et al (2013). A comparison between cox regression and parametric methods in analyzing kidney transplant survival. *World App Sci J*, **26**, 502-7.
- Klein J, Moeschberger M (2003). *Survival analysis: statistical methods for censored and truncated data*. Springer, New

York, 393.

- Kleinbaum D, Klein M (2005). *Survival analysis: A self-learning approach*. Nova Iorque: Springer, 290.
- Li J, Li J, Cai G, et al (2017). Proapoptotic and growth-inhibitory effects of plumbagin on human gastric cancer cells via suppression of signal transducer and activator of transcription 3 and protein kinase B. *Altern Ther Health Med*, **23**, 42-8.
- Ma J, Shen H, Kapesa L, et al (2016). Lauren classification and individualized chemotherapy in gastric cancer. *Oncol lett*, **11**, 2959-64.
- Malekzadeh R, Derakhshan MH, Malekzadeh Z (2009). Gastric cancer in Iran: epidemiology and risk factors. *Arch Iran Med*, **12**, 576-83.
- Nardi A, Schemper M (2003). Comparing cox and parametric models in clinical studies. *Stat Med*, **22**, 3597-610.
- Pourhoseingholi M, Moghimi-Dehkordi B, Safaee A, et al (2009). Prognostic factors in gastric cancer using log-normal censored regression model. *Indian J Med Res*, **129**, 262.
- Pourhoseingholi M, Pourhoseingholi A, Vahedi M, et al (2011). Alternative for the cox regression model: using parametric models to analyze the survival of cancer patients. *Iran J Cancer Prev*, **4**, 1-9.
- Rahman R, Asombang AW, Ibdah JA (2014). Characteristics of gastric cancer in Asia. *World J Gastroenterol*, **20**, 4483.
- Rugge M, Fassan M, Graham DY (2015). Epidemiology of gastric cancer. in 'gastric cancer', Eds Springer, pp 23-34.
- Shayan Z, Ayatollahi SMT, Zare N, Moradi F (2014). Prognostic factors of first birth interval using the parametric survival models. *Iran J Reprod Med*, **12**, 125-30.
- Torre LA, Bray F, Siegel RL, et al (2015). Global cancer statistics, 2012. *CA Cancer J Clin*, **65**, 87-108.
- Veisani Y, Delpisheh A (2016). Survival rate of gastric cancer in Iran; a systematic review and meta-analysis. *Gastroenterol Hepatol Bed Bench*, **9**, 78.
- Wang B-B, Liu C-G, Lu P, et al (2011). Log-normal censored regression model detecting prognostic factors in gastric cancer: a study of 3018 cases. *World J Gastroenterol*, **17**, 2867.
- Zare A, Hosseini M, Mahmoodi M, et al (2015). A Comparison between accelerated failure-time and cox proportional hazard models in analyzing the survival of gastric cancer patients. *Iran J Public Health*, **44**, 1095-102.
- Zare A, Mahmoodi M, Mohammad K, et al (2014). A comparison between kaplan-meier and weighted kaplan-meier methods of five-year survival estimation of patients with gastric cancer. *Acta Med Iran*, **52**, 764.
- Zhu HP, Xia X, Chuan HY, et al (2011). Application of Weibull model for survival of patients with gastric cancer. *BMC Gastroenterol*, **11**, 1.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.