

## RESEARCH ARTICLE

# Comparison between Parametric and Semi-parametric Cox Models in Modeling Transition Rates of a Multi-state Model: Application in Patients with Gastric Cancer Undergoing Surgery at the Iran Cancer Institute

Ali Zare<sup>1</sup>, Mahmood Mahmoodi<sup>2\*</sup>, Kazem Mohammad<sup>3</sup>, Hojjat Zeraati<sup>4</sup>, Mostafa Hosseini<sup>5</sup>, Kouros Holakouie Naieni<sup>6</sup>

## Abstract

**Background:** Research on cancers with a high rate of mortality such as those occurring in the stomach requires using models which can provide a closer examination of disease processes and provide researchers with more accurate data. Various models have been designed based on this issue and the present study aimed at evaluating such models. **Materials and Methods:** Data from 330 patients with gastric cancer undergoing surgery at Iran Cancer Institute from 1995 to 1999 were analyzed. Cox-Snell Residuals and Akaike Information Criterion were used to compare parametric and semi-parametric Cox models in modeling transition rates among different states of a multi-state model. R 2.15.1 software was used for all data analyses. **Results:** Analysis of Cox-Snell Residuals and Akaike Information Criterion for all probable transitions among different states revealed that parametric models represented a better fitness. Log-logistic, Gompertz and Log-normal models were good choices for modeling transition rate for relapse hazard (state 1→state 2), death hazard without a relapse (state 1→state 3) and death hazard with a relapse (state 2→state 3), respectively. **Conclusions:** Although the semi-parametric Cox model is often used by most cancer researchers in modeling transition rates of multi-state models, parametric models in similar situations- as they do not need proportional hazards assumption and consider a specific statistical distribution for time to occurrence of next state in case this assumption is not made - are more credible alternatives.

**Keywords:** Gastric cancer - multi-state model - parametric model - proportional hazards model - transition rate

*Asian Pac J Cancer Prev*, 14 (11), 6751-6755

## Introduction

Gastric cancer is one of the most common causes of cancer deaths all over the world. Every year, more than 870 thousand new cases are reported throughout the world and more than 650 thousand people die from this type of cancer (Parkin, 1998). According to the latest statistics of Iran Cancer Research Center, gastric cancer is the most common cancer among Iranian men and the third most common cancer among Iranian women after breast cancer (Mohagheghi et al., 1998; Mohagheghi, 2004). One of the most important objectives specified after the right diagnosis and prompt treatment for the patients with gastric cancer is the survival rate increase especially the 5-year survival rate. Unfortunately, more than 80% of patients with gastric cancer are diagnosed at a stage when common treatments such as gastrectomy, chemotherapy, or radiation therapy are not effective in increasing the patients' survival (Gunderson and Sosin, 1982; Wisbeck et al., 1986; Sadighi et al., 2005; Samadi et al., 2007; Sadighi

et al., 2008; Association, 2011). For this reason, the 5-year survival rate is low in patients with gastric cancer after surgery (Thong-Ngam et al., 2001; Triboulet et al., 2001; Schwarz and Zagala-Nevarez, 2002; Adachi et al., 2003; Ding et al., 2004). The increase in these patients' survival after surgery requires using models which could provide a closer examination on the behavior of variables so that it will better describe the natural process of the disease and will provide the researchers with more accurate data.

One of the statistical models designed to accomplish this purpose is the multi-state model. According to this model, patients experience different states (save for death event) during the study from the beginning to death event. The time reaching each state and factors affecting its occurrence play a fundamental role in patients' survival. Considering these states (often called intermediate events) has developed a novel approach in survival studies, because the natural process of disease in such cases can be considered as a stochastic process in which patients can be placed in various states throughout the study.

Department of Epidemiology and Biostatistics, University of Medical Sciences, Tehran, Iran \*For correspondence: mahmoodim@tums.ac.ir

Most studies on cancer regard the death event as the only probable occurrence for patients, but in many cases, events other than death event occur in patients during the study period which may affect the final results. Relapse is an obvious example for these events. Not only does it affect patients' survival as a variable but its occurrence is also influenced by different factors.

Standard models of survival are the simplest of multi-state model. In these models, the patient is "Alive" at the beginning of the study and then his/her state may change to "Death". This is the only transition which is considered for patients during the study. This transition can be illustrated as show in Figure 1.

Models more complex than two-state, provide the patients with the probability of more transition during the study. These models are used when the initial state of the patient, i.e. "being alive", is itself divided to two or more other states. The number of division depends on the type of disease. In the process of gastric cancer, the survival time of patient is recorded since the patient has undergone surgery. After surgery the patient enters the study and is subjected to death hazard. In these studies, the occurrence of death and relapse are considered as the end point of the study and the intermediate even respectively. This modeling is schematically shown in Figure 2.

The most important practical application of multi-state models is transition rate modeling among states in each transition. There are various statistical methods for this purpose including parametric and semi-parametric models. These models are divided into two main categories: Proportional Hazard model and Accelerated Failure-time model. In the proportional hazard model, modeling is done based on hazard function. In this case, if baseline hazard is considered parametric, one of the exponential, Weibull or Gompertz models will be obtained. If the baseline hazard is considered indefinite, the semi-parametric Cox model will be achieved. In the accelerated failure-time model, modeling is done on time logarithm to the occurrence of next state. The obtained models in this case include: Exponential, Weibull, Log-normal, Log-logistic, Gompertz and Generalized gamma. It should be noted that the Weibull and exponential models are the only ones that have, both, the PH and AFT features.

The semi-parametric Cox model does not need to

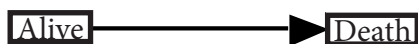


Figure 1. Standard Models

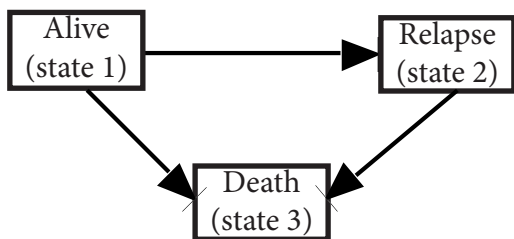


Figure 2. Three Transitions for Patients During the Study in the Above Model. Death hazard without a relapse (state 1→state 3); Relapse hazard (state 1→state 2); Death hazard with a relapse (state 2→state 3)

consider a specific probability distribution for time to the occurrence of next state; therefore, it is the most useful model in modeling transition rates of multi-state models (Hougaard, 1999; Yagi et al., 2000; Andersen and Keiding, 2002; Adachi et al., 2003; Buonadonna et al., 2003; Chau et al., 2004; Zeraati et al., 2005; Dehkordi and Tabatabaee, 2007; Biglarian et al., 2009). But this model is severely affected by proportional hazards assumption and, for this reason, is often called Cox proportional hazard model. In cases where the proportional hazard model is not tenable, inferences derived from this model will be flawed and will have a bias (Hougaard, 1999; Collett, 2003; Klein and Moeschberger, 2003; Hosmer Jr et al., 2011). Accelerated failure-time models are particularly important in such circumstances. These models—due to having a specific probability distribution for time to the occurrence of next state —make statistical inference more accurate and cause the standard errors of the estimations to be smaller in ratio with the time when there were no such assumptions. Modeling of transition rates in multi-state models is often identified by Cox proportional hazard model (Klein et al., 1984; Kay, 1986; Andersen, 1988; Hougaard, 1999; Andersen and Keiding, 2002; Klein and Moeschberger, 2003; Putter et al., 2007; De Wreede et al., 2010; Jackson, 2011). Neither have these studies generally used proportional hazards assumption nor did they attempt to identify a suitable parametric model as an alternative to Cox proportional hazard model. The main problem in the application of multi-state models is the need to determine the most appropriate model in modeling transition rates. So, in addition to comparing parametric and Cox semi-parametric models in modeling transition rates among different states, Akaike Information Criterion and Cox-Snell Residuals have been also used in this study to assess these models.

**Materials and Methods**

In this study, 330 patients with gastric cancer with the following data were studied: *i*) the patients had been hospitalized and had undergone surgery from 1995 to 1999 in surgical wards of Cancer Institute of Iran; *ii*) they had records in the archives of the hospital, and in their files their addresses and phone numbers were available for subsequent follow-up. The survival time of patients was determined after surgery and those patients who were still alive at the end of study period or the ones whose data were not available after a specific time-period were considered right censored.

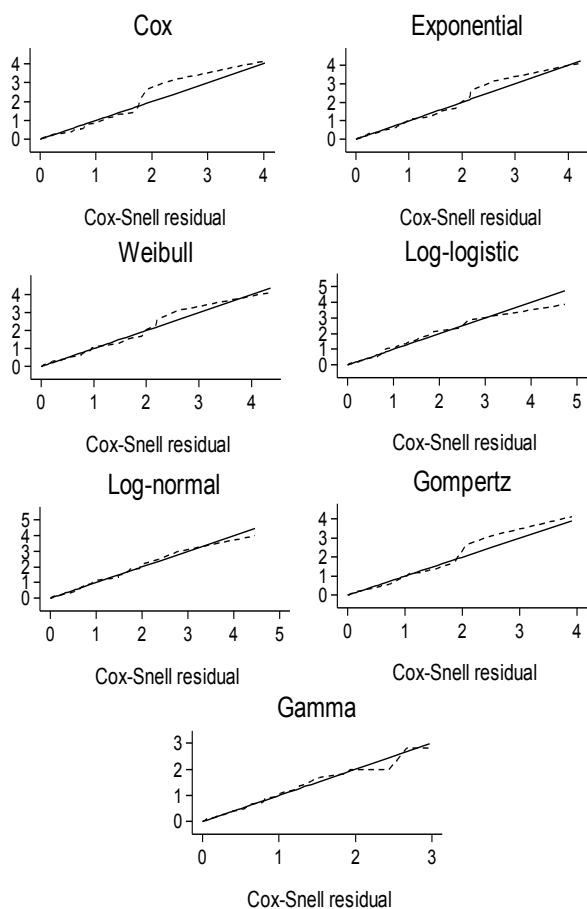
Since it is common to use the Cox proportional hazards model in modeling transition rates of multi-state models, there is a risk that if the assumption of proportional hazards is not fulfilled, the results will not be reliable enough. Although some researchers tend to turn a blind eye to this defect in their researches due to the ease of Cox's model's application and its interpretations, it is essential to use alternative models with a higher degree of reliability such as parametric models for more precise investigations in such cases. Therefore, in this study to compare parametric models including Exponential, Weibull, Log-normal, Log-logistic, Gompertz, and Generalized gamma as well as

Cox semi-parametric model in modeling transition rates among different states, a multi-state model with three states of patient's being alive without a relapse (state 1), relapse (state 2) and death (state 3) was considered. Moreover, Cox-Snell Residuals and Akaike Information Criterion were used to assess these models properly. Cox-Snell Residuals is a graphical criterion for assessing the fitness of parametric and semi-parametric models; the less deviation of residuals from the bisector, the more appropriate fitness of model (Weissfeld and Schneider, 1990; Escobar and Meeker Jr, 1992; Collett, 2003). Graphical methods are often associated with optical illusion. For a better judgment, thus, Akaike information criterion can be used along with Cox-Snell residuals. Akaike information criterion is used to measure the goodness of statistical models' fitness, and the smaller it is, the better it is (Collett, 2003; Klein and Moeschberger, 2003). AIC for the models used in this study has been calculated according to the following equation:  $AIC = -2 \times \log(\text{likelihood}) + 2(p+k)$

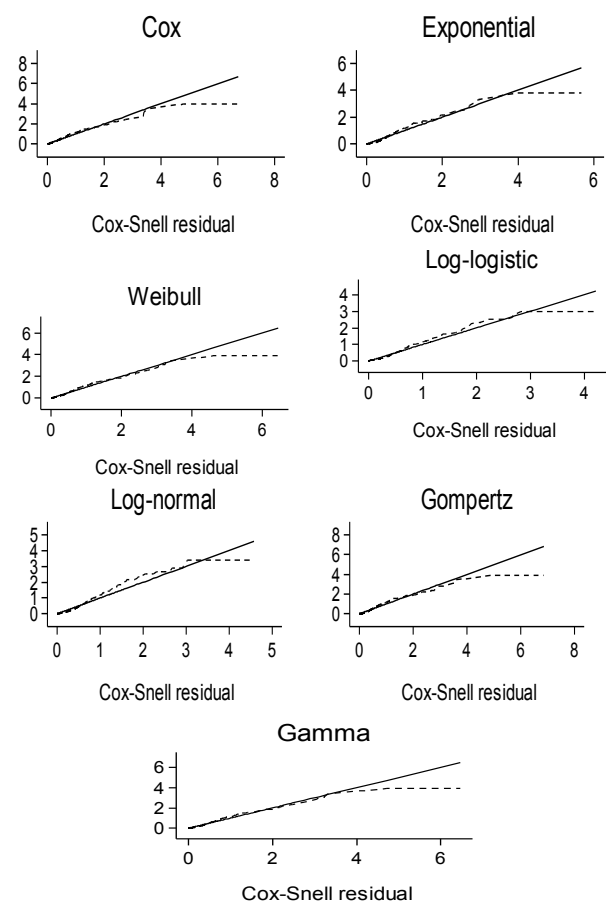
In which  $p$  is the number of parameters in the model and  $k$  is a constant coefficient which has been used depending on the type of model. For example,  $k$  is for the exponential model and  $k=2$  for Weibull model (Klein and Moeschberger, 2003). The smaller the AIC is, the more efficacious the model will be. All data were analyzed using R 2.15.1 software.

## Results

Figure analysis of Cox-Snell residuals for parametric models and Cox semi-parametric model in modeling transition rates in three transitions; relapse hazard (*state 1*→*state 2*), death hazard without a relapse (*state 1*→*state 3*) and death hazard with a relapse (*state 2*→*state 3*) represents better fitness of parametric models compared with Cox semi parametric model. The figure of Cox-Snell residuals for relapse hazard (*state 1*→*state 2*) shows that among parametric models, Log-logistic model has a better fitness to data (Figure 3). Furthermore, analysis results of these residuals to compare parametric and Cox semi-parametric models in modeling transition rates for death hazard without a relapse (*state 1*→*state 3*) also revealed that among parametric models Gompertz proved better fitness to the data (Figure 4). Besides, the analysis of these residuals in modeling transition rates for death hazard with a relapse (*state 2*→*state 3*) showed that Log-normal model is a suitable alternative for Cox semi-parametric model in modeling this transition rate (Figure 5). Akaike information criterion confirms these results too. Based on this criterion, Log-logistic model in modeling transition rate *state 1*→*state 2*, Gompertz in modeling transition rate *state 1*→*state 3* and Log-normal in modeling transition rate *state 2*→*state 3* are the best models (Table 1).



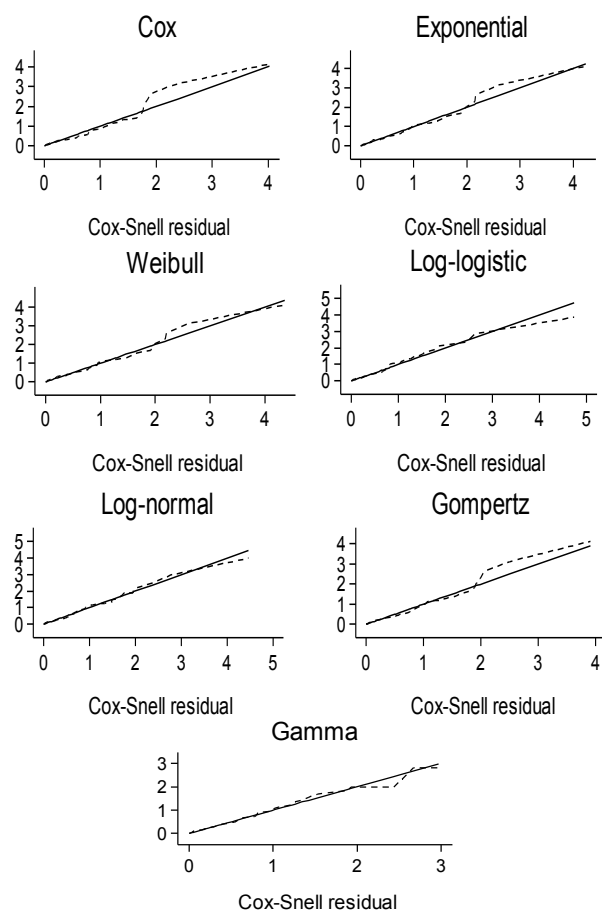
**Figure 3. The Cox-Snell Residuals in the Considered Cox Proportional Hazard and Parametric Models in Modeling Transition Rate *state 1*→*state 2***



**Figure 4. The Cox-Snell Residuals in the Considered Cox Proportional Hazard and Parametric Models in Modeling Transition Rate *state 1*→*state 3***

**Table 1. The Comparison Results of the Akaike Information Criterion (AIC) between the Cox Proportional Hazard and Parametric Models**

Models	state 1→state 2	state 1→state 3	state 2→state 3
Cox	414.2	1884.5	231
Exponential	293.9	809.7	163.3
Weibull	290.4	807.8	162.4
Log-logistic	285.5	808	161.8
Log-normal	290	863.9	160.5
Gompertz	294	805.5	163.8
Gamma	288.8	809.7	161.3



**Figure 5. The Cox-Snell Residuals in the Considered Cox Proportional Hazard and Parametric Models in Modeling Transition Rate state 2→state 3**

## Discussion

Most cancer researchers tend to use Cox semi-parametric model rather than parametric models in modeling transition rates among different states in a multi-state model. A systematic review on cancer journals indicates that only 5% of studies in which Cox model has been used for modeling transition rates among different states, investigated the required assumptions for this model (Altman et al., 1995). The absence of proportional hazards assumption causes the estimations of transition rates among different states to be unreliable and biased. Moreover, studies conducted in this scope demonstrate that either proportional hazards assumption is made or not, parametric models are more efficient (Orbe et al., 2002; Patel et al., 2006). Therefore, parametric models

such as exponential, Weibull, log-normal, log-logistic, Gompertz and gamma can be better choices in such situations. Considering a particular statistical distribution for time to the occurrence of next state and requiring no assumption of proportional hazards (PH), these models provide fitness for data.

A major objective of this paper is to investigate the comparative performance of Cox semi-parametric and parametric survival models in modeling transition rates of multi-state models. So, in this study the results of Cox semi-parametric model and parametric models were compared in modeling transition rates of a multi-state model with three states of patient's being alive without a relapse (state 1), relapse (state 2) and death (state 3). To assess these models, Akaike information criterion (AIC) and Cox-Snell residuals were used. The analysis of Cox-Snell residuals for all probable transitions among states revealed that parametric models had better fitness. This finding is consistent with the findings obtained from most studies carried out on patients with gastric cancer (Orbe et al., 2002; Nardi and Schemper, 2003; Dehkordi, 2007; Pourhoseingholi et al., 2007). In the meanwhile, Log-logistic, Gompertz, and Log-normal were suitable choices for modeling transition rate state 1→state 2, modeling transition rate state 1→state 3, and modeling transition rate state 2→state 3, respectively. In addition, the analysis of models based on Akaike information criterion also confirmed the results obtained from Cox-Snell residuals (Table 1).

Although most researchers in medical and cancer fields have made use of Cox semi-parametric models in modeling transition rates of a multi-state model into account, results of parametric models have often been more reliable and have had less bias. As parametric models do not need proportional hazards assumption (PH) in similar situations and they consider a specific statistical distribution for time to the occurrence of next state, they have a better fitness. Parametric models will also be credible alternatives to Cox semi-parametric model where proportional hazard assumption is not made. In addition, fully parametric models may offer some advantages. Based on asymptotic results, parametric models lead to more efficient parameter estimates than Cox model. With a decrease in sample sizes, relative efficiencies may further change in favor of parametric models. When empirical information is sufficient, parametric models can provide some insights into the shape of the baseline hazard.

## References

Adachi Y, Tsuchihashi J, Shiraishi N, et al (2003). AFP-producing gastric carcinoma: multivariate analysis of prognostic factors in 270 patients. *Oncology*, **65**, 95-101.

Altman D, De Stavola B, Love S, Stepniowska K (1995). Review of survival analyses published in cancer journals. *Bri J Cancer*, **72**, 511.

Andersen PK (1988). Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Stat Med*, **7**, 661-70.

Andersen PK, Keiding N (2002). Multi-state models for event history analysis. *Stat Methods Med Res*, **11**, 91-115.

Association JGC (2011). Japanese gastric cancer treatment



- Comparison between Parametric and Semi-parametric Cox Models in Modeling Transition Rates with Gastric Cancer* guidelines 2010 (ver. 3). *Gastric Cancer*, **14**, 113-23.
- Biglarian A, Hajizadeh E, Kazemnejad A, Zali M (2009). Survival analysis of gastric cancer patients using Cox model: a five year study. *TUMJ*, **67**, 5.
- Buonadonna A, Lombardi D, De Paoli A, Bidoli E, Frustaci S (2003). Adenocarcinoma of the stomach: univariate and multivariate analyses of factors associated with survival. *Suppl Tumori*, **2**, 31-4.
- Chau I, Norman AR, Cunningham D, et al (2004). Multivariate prognostic factor analysis in locally advanced and metastatic esophago-gastric cancer-pooled analysis from three multicenter, randomized, controlled trials using individual patient data. *J Clin Oncol*, **22**, 2395-403.
- Collett D (2003). Modelling survival data in medical research: CRC press.
- De Wreede LC, Fiocco M, Putter H (2010). The mstate package for estimation and prediction in non-and semi-parametric multi-state and competing risks models. *Comput Methods Programs Biomed*, **99**, 261-74.
- Dehkordi B (2007). Comparing cox regression and parametric models for survival analysis of patients with gastric cancer. *Iranian J Epidemiol*, **3**.
- Dehkordi BM, Tabatabaee H (2007). Modeling survival analysis in gastric cancer patients using the proportional hazards model of Cox. *Iran J Epidemiol*, **3**, 1-2.
- Ding YB, Chen GY, Xia JG, et al (2004). Correlation of tumor-positive ratio and number of perigastric lymph nodes with prognosis of patients with surgically-removed gastric carcinoma. *World J Gastroenterol*, **10**, 182-5.
- Escobar LA, Meeker Jr WQ (1992). Assessing influence in regression analysis with censored data. *Biometrics*, **48**, 507-28.
- Gunderson LL, Sosin H (1982). Adenocarcinoma of the stomach: areas of failure in a re-operation series (second or symptomatic look) clinicopathologic correlation and implications for adjuvant therapy. *Int J Radiation Oncol & Biology & Physics*, **8**, 1-11.
- Hosmer Jr DW, Lemeshow S, May S (2011). Applied survival analysis: regression modeling of time to event data: Wiley-Interscience.
- Hougaard P (1999). Multi-state models: a review. *Lifetime Data Anal*, **5**, 239-64.
- Jackson CH (2011). Multi-state models for panel data: the msm package for R. *J Stat Software*, **38**, 1-29.
- Kay R (1986). A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*, **42**, 855-65.
- Klein JP, Klotz JH, Grever MR (1984). A biological marker model for predicting disease transitions. *Biometrics*, **40**, 927-36.
- Klein JP, Moeschberger ML (2003). Survival analysis: techniques for censored and truncated data: Springer.
- Mohagheghi MA (2004). Annual report of Tehran cancer registry 1999: The cancer institute publication.
- Mohagheghi MA, Mosavi-Jarrahi A, Zeraati H (1998). Annual report of Tehran University of medical sciences district cancer registry 1997. Tehran: The cancer institute publication.
- Nardi A, Schemper M (2003). Comparing Cox and parametric models in clinical studies. *Stat Med*, **22**, 3597-610.
- Orbe J, Ferreira E, Núñez Antón V (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Stat Med*, **21**, 3493-510.
- Parkin D (1998). Epidemiology of cancer: global patterns and trends. *Toxicol Lett*, **102**, 227-34.
- Patel K, Kay R, Rowell L (2006). Comparing proportional hazards and accelerated failure time models: an application in influenza. *Pharmaceutical Stat*, **5**, 213-224.
- Pourhoseingholi MA, Hajizadeh E, Moghimi Dehkordi B, et al (2007). Comparing cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pac J Cancer Prev*, **8**, 412.
- Putter H, Fiocco M, Geskus R (2007). Tutorial in biostatistics: competing risks and multi state models. *Stat Med*, **26**, 2389-430.
- Sadighi S, Mohagheghi M, Haddad P, et al (2008). Life expectancy with perioperative chemotherapy and chemoradiotherapy for locally advanced gastric adenocarcinoma. *TUMJ*, **66**, 664-9.
- Sadighi S, Raafat J, Mohagheghi M, Meemary F (2005). Gastric carcinoma: 5 year experience of a single institute. *Asian Pac J Cancer Prev*, **6**, 195-6.
- Samadi F, Babaei M, Yazdanbod A, et al (2007). Survival rate of gastric and esophageal cancers in Ardabil province, North-West of Iran. *Arch Iran Med*, **10**, 32-7.
- Schwarz RE, Zagala-Nevarez K (2002). Recurrence patterns after radical gastrectomy for gastric cancer: prognostic factors and implications for postoperative adjuvant therapy. *Ann Surg Oncol*, **9**, 394-400.
- Thong-Ngam D, Tangkijvanich P, Mahachai V, Kullavanijaya P (2001). Current status of gastric cancer in Thai patients. *J Med Assoc Thailand*, **84**, 475-82.
- Triboulet J, Fabre S, Castel B, Tournel H (2001). Adenocarcinomas of the distal esophagus and cardia: surgical management. *Cancer Radither*, **5**, 90-7.
- Weissfeld LA, Schneider H (1990). Influence diagnostics for the Weibull model fit to censored data. *Stat Probabil Lett*, **9**, 67-73.
- Wisbeck WA, Becker EM, Russell AH (1986). Adenocarcinoma of the stomach: autopsy observations with therapeutic implications for the radiation oncologist. *Radiother Oncol*, **7**, 13-8.
- Yagi Y, Seshimo A, Kameoka S (2000). Prognostic factors in stage IV gastric cancer: univariate and multivariate analyses. *Gastric Cancer*, **3**, 71-80.
- Zeraati H, Mahmoudi M, Kazemnejad A, Mohammad K (2005). Postoperative survival in gastric cancer patients and its associated factors: a method based on a non-homogenous semi-markovian process. *Int J Cancer Res*, **1**, 87-93.