RESEARCH COMMUNICATION

Statistical Comparison of Survival Models for Analysis of Cancer Data

Bijan Moghimi-Dehkordi[1]*, Azadeh Safaee[1], Mohamad Amin Pourhoseingholi[1], Reza Fatemi[1] , Ziaoddin Tabeie[2], Mohammad Reza Zali[1]

Abstract

Background: The Cox Proportional Hazard model is the most popular technique to analysis the effects of covariates on survival time but under certain circumstances parametric models may offer advantages over Cox's model. In this study we use Cox regression and alternative parametric models such as: Weibull, Exponential and Lognormal models to evaluate prognostic factors affecting survival of patients with stomach cancer. Comparisons were made to find the best model. Methods: To determine independent prognostic factors reducing survival time for stomach cancer, we compared parametric and semi-parametric methods applied to patients who registered in one cancer registry center located in southern Iran using the Akaike Information Criterion. Results: Of a total of 442 patients, 266 (60.2%) died. The results of data analysis using Cox and parametric models were approximately similar. Patients with ages 60-75 and >75 years at diagnosis had an increased risk for death followed by those with poor differentiated grade and presence of distant metastasis (P<0.05). Conclusion: Although the Hazard Ratio in Cox model and parametric ones are approximately similar, according to Akaike Information Criterion, the Weibull and Exponential models are the most favorable for survival analysis.

Key Words: Stomach cancer - prognosis - Cox regression - parametric models - AIC

Asian Pacific J Cancer Prev, **9**, 417-420

Introduction

The development of methods to analyze survival data is one of the areas in statistics that have increased the most in the last few years. This is because, in many practical situations, researchers are interested in the survival time until the occurrence of an event, such as the failure time of a component or the time until the death of a subject. However, it is not always possible to observe the exact time when an event occurs, but only its interval is identified (Corrente et al., 2003).

The survival time is used to evaluate the effects of treatments for stomach cancer (Jun et al., 2004) and survival rates are calculated in the analysis for the survival time (Lee, 1992). Many studies have been conducted on the survival analysis of stomach cancer in recent several decades (Lundegardh et al., 1986; Jimeno-Aranda et al., 1996; Pinheiro et al., 1999; Msika et al., 2000a; 2000b; Barchielli et al., 2001; Janer et al., 2002). Since stomach cancer often is not detected until an advanced stage, survival rate is rather low. In one series, only a few patients diagnosed with stomach cancer survived five years or more after diagnosis (Sánchez-Bueno et al., 1998).

Up-to-date information on cancer survival is important as reference material for clinicians, oncologists, epidemiologists and scientists involved in clinical work, medical auditing or research (Talbäck et al., 2004).

The objective of many studies is to characterize the different survival distributions that correspond to different subgroups within a heterogeneous population. A descriptive summary of such a comparison could consist of parametric or semi-parametric methods. There are two major regression models used for right censored data: Cox proportional hazards model as a semi-parametric method (Cox, 1972) and accelerated failure time model as a parametric model. Many of the standard parametric models such as Weibull, Exponential and Lognormal are accelerated failure time models. However Cox regression is the most widely employed model in survival analysis, parametric models (Lawless, 1998).

Researchers in medical sciences often tend to prefer semi-parametric instead of parametric models because of fewer assumptions but some comments recommended that under certain circumstances, parametric models estimate the parameter more efficient than Cox (Efron, 1977; Oakes, 1977). In parametric models we often use maximum likelihood procedures to estimate the unknown parameters and this technique and its interpretation are familiar for researchers. Also accelerated failure time can be used as relative risk with similar interpretation in Cox

[1]Research Center for Gastroenterology and Liver Diseases, Shaheed Beheshti University (M.C.), Tehran, [2]Fars Cancer Registry, Pathology Department, Faculty of Medicine, Fars University of Medical Science, Shiraz, Iran
**For correspondence: b_ moghimi_ de@ yahoo.com*

regression.

In this study we used Cox regression and alternative parametric models such as: Weibull, Exponential and Lognormal models to evaluate the prognostic factor that affect on survival of patients with stomach cancer and comparisons were made to find the best model.

Materials and Methods

The data represent a historical cohort study of all patients registered from March 2001 until March 2006, in cancer registry center of Fars province in southern Iran with a diagnosis of stomach cancer and entered into the study. Vital status of these patients was followed until the end of March 2006 by the various sources of information. Multivariate analysis of prognostic factors was carried out by two methods: (_) Cox proportional hazard model (as semi-parametric method) and (__) Weibull, Exponential and Lognormal models (as parametric methods). Cox's model has become the most used procedure for modeling the relationship of covariates to a survival or other censored outcome (Therneau T, Grambsch, 2000). However, it has some restrictions. One of the restrictions to using the Cox model with time-fixed covariates is its proportional hazards assumption; it means the hazard ratio between two sets of covariates is constant over time. This is due to the common baseline hazard function canceling out in the ratio of the two hazards.

Although the parametric models might be somewhat more efficient, they have more assumptions but if the assumptions are met, the analysis is more powerful. We have considered Weibull and Exponential models with respect to the assumptions of constant and monotone baseline hazard respectively and lognormal model because its baseline hazard has value 0 at t=0, increases to maximum and then decreases, approaching 0 as becomes large. The likelihood value estimates were employed to comparison among models.

Evaluation Criteria

To set a comparison among models we used Akaike Information Criterion (AIC) proposed in Akaike (1974). AIC is a measure of the goodness of fit of an estimated statistical model (Akaike, 1977). The AIC is an operational way of trading off the complexity of an estimated model against how well the model fits the data. For our models discussed, the AIC is given by,

$$AIC = -2*\log(likelihood) + 2(p+k)$$

Where p is the number of parameter, k=1 for the exponential model, k=2 for the Weibull and log normal models (Klein and Moeschberger, 1997). Lower AIC indicates better likelihood. All calculations were carried out by STATA (version8.0) statistical software.

Results

There were 442 patients (68.6% males). The mean age at time of diagnosis was 58.4 ±14.5 years and mean survival time was 26.5 months. Of 442 patients, 266 (60.2%) were dead and the others were censored (right censored). We compared Parametric and semi parametric models by using AIC. According to the graphical test (not shown here) the proportional hazard assumption holds. Table 1 shows the results of multivariate analysis using Cox and alternative parametric models. According to Cox regression analysis, among variables that entered to model (age at diagnosis, sex, ethnicity, marital status, male occupation, female occupation, family history of cancer, history of gastrointestinal diseases, body mass index, smoking, type of first treatment, grade of tumor, distant metastasis and time interval between beginning first symptoms date to diagnostic date) the factors influencing on survival of patients were: age at diagnosis (60-75 and >75 vs. others), distant metastasis (have vs. not have) and grade of tumor(poor differentiated vs. other subgroups) ($p<0.05$).

The results of data analysis using parametric models are, similar to the Cox regression. Patients aged 60-75 and >75 years at diagnosis had an increased risk for death, followed by those with poor differentiated grade and presence of distant metastasis ($P<0.05$). Although the Hazard Ratio in Cox and parametric models are approximately similar, from the AIC, Weibull and Exponential are the most favorable for survival analysis.

Table1. Prognostic Factors of Stomach Cancer using Cox and Parametric Models

Prognostic factors	Cox regression HR† (CI§ 95%)	Exponential HR (CI 95%)	Log-normal HR (CI 95%)	Weibull HR (CI 95%)
Metastasis				
Negative‡	1	1	1	1
Positive	1.53 (1.16-2.02)	1.67 (1.27-2.02)	1.84 (1.07-3.12)	1.64 (1.24-2.16)
Grade of tumor				
Well differentiated‡	1	1	1	1
Moderately differentiated	NS*	NS	NS	NS
Poorly differentiated	1.45(1.10-1.91)	1.49 (1.13-1.97)	1.43 (1.03-1.85)	1.56 (1.18-2.06)
Age at diagnosis				
<45	‡	1	1	1
45-60	NS	NS	NS	NS
60-75	1.36 (1.01-1.89)	1.54 (1.10-1.96)	1.67 (1.04-2.75)	1.03 (1.03-2.81)
>75	1.71 (1.13-2.59)	1.93 (1.29-2.87)	2.47 (1.64-3.70)	2.29 (1.49-3.49)
AIC#	2,157	850	1,159	848

† Hazard Ratio § Confidence interval * Not significant ‡ Reference group # Akaike Information Criterion

Discussion

Cox regression model is the most common way of analyzing prognostic factors in clinical research. This is probably due to the fact that this model allows us to estimate and make inference about the parameters without assuming any distribution for the lifetime, whose distribution is often unknown. However, it does have the requirement of proportional hazards, which is not always satisfied by the data. In a review of survival analyses in cancer journals (Altman et al., 1985), it was found that only 5% of all studies using the Cox regression model with respect to checking the underlying assumptions. In these situations, parametric models (such as Lognormal, Weibull and Exponential) provide an alternative framework to fit the data. Moreover, under these models we measure the direct effect of the explanatory variables on the survival time and not on a conditional probability, as we do in the Cox regression model. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the mean lifetime.

This present study aimed at investigation the comparative performance of Cox and parametric models in a survival analysis of patients with stomach carcinoma. We used Akaike Information Criterion (AIC) to evaluate among models. In our example the proportional hazard assumptions were hold and the all parametric model residual (not shown here) indicated a perfect fit.

In present study, three variables were as independent prognostic factors on survival of patients with stomach cancer in both multivariate methods, but the coefficients have a little difference in two models.

In our study, age at diagnosis was an independent prognostic factor by the both two methods (parametric and semi-parametric) in stomach cancer patients. This finding was similar to the result of many studies (Matley et al., 1988; Mitsudomi et al., 1989; Harrison and Fielding, 1995; Maehara et al., 1995; Pacelli et al., 1999; Tuech et al., 1999; Basili et al., 2003; Faycal et al., 2005; Saito et al., 2006) but was differing from some studies (Salvon-Harman et al., 1994; Sánchez-Bueno et al., 1998; Zhang et al., 2004).

With respect to the degree of cellular differentiation, the best prognosis has been found in well differentiated tumors (Cady et al., 1989; Shiu et al., 1989; Arveux et al., 1992; Carriaga and Henson, 1995; Kitamura et al., 1996; Yokota et al., 2000; Park et al., 2006). Patients with low-grade tumors had a greater survival rate than those with high-grade tumors, the difference being statistically significant. These results were incorporate with our findings. In a study, grade of tumor was not related to survival of stomach cancer patients (Oertli et al., 1994).

In many issues, metastasis of tumor mentioned as an independent prognostic factor with survival of patients (Massacesi et al., 2000; Zhang et al., 2004; Yoshida et al., 2004; Park et al., 2006). Also, our result represented that patients who have metastasis at time of diagnosis had a poorer prognosis of survival.

AIC criteria indicated Weibull and Exponential model are similarly the best models in multivariate analysis. Our data strongly supported that these two models among all models that we used for survival analysis could be lead to more precise results as an alternative for Cox. There were also some studies using parametric model to analysis the survival of gastric cancer. The group of Ferreira and Nunez-Anton conducted a simulation study to comparing Cox and accelerated failure time models they also presented this comparison in a gastric cancer data set that the proportional hazard assumption did not hold. The findings showed a perfect fitting for lognormal (Orbe et al., 2002). Pourhoseingholi et al used parametric model in compared to Cox for survival of gastric carcinoma finding Lognormal and Exponential regression with same parameter estimations (Pourhoseingholi et al., 2007).

In our data, the percent of censoring was 39.8%. A good discrimination among parametric models requires the censoring percentage not to exceed 40-50% (Nardi and Schemper, 2003). In addition, Oakes (1977) discussed that; asymptotically well fitted parametric models should be more efficient than Cox if parameter values are far from zero.

In Conclusion, lthough regression coefficients are not all the same, age at diagnosis, grade of tumor and metastasis should be considered as most important prognostic factors that affect life expectancy of patients with stomach cancer.

However the Cox parameter estimations are familiar for researchers in the field of medical sciences, the results in accelerated failure times is not unknown for medical scientists. So these parameters can be interpreted as factor accelerating or decelerating similarly in the interpretation of Cox' Hazard Ratio. These parametric models can easily conducted by maximum likelihood estimators and let the researchers to explore the data through the different relationships consist of leaner trend, nonlinear ones or interactions and when the proportional hazard assumption dose not hold these methods lead to acceptable conclusions. In spite of this advantage further study should be carried out to evaluate the effects of practical cases such as small sample size, large censoring and changing in proportional hazard assumption or duration time's distribution.

Acknowledgements

We would like to thank Fars Province Cancer Registry staff for their valuable collaboration, and Fars University of Medical Sciences for the preparing the grant in this study.

References

Akaike H (1977). A new look at the statistical model identification. *IEEE Trans Automatic Control*, **19**, 716-23.

Altman DG, De Stavola BL, Love SB, Stepniewska KA (1985). Review of survival analyses published in cancer journals. *Br J Cancer*, **72**, 511-8.

Arveux P, Faivre J, Boutron MC, et al (1992). Prognosis of gastric carcinoma after curative surgery. A population based study using multivariate crude and relative survival analysis. *Dig Dis Sci*, **37**, 757-63.

Barchielli A, Amorosi A, Balzi D, Crocetti E, Nesi G (2001). Long-term prognosis of gastric cancer in a European country: a population-based study in Florence. 10-year survival of cases diagnosed in 1985-1987. *Eur J Cancer*, **37**, 1674-80.

Basili G, Nesi G, Barchielli A, Manetti A, Biliotti G (2003). Pathologic features and long-term results in early gastric cancer: report of 116 cases 8-13 years after surgery. *World J Surg*, **27**, 149-52.

Cady B, Rossi RL, Silverman ML, Piccione W, Heck TA (1989). Gastric adenocarcinoma, A disease in transition. *Arch Surg*, **124**, 303-8.

Carriaga MT, Henson DE (1995). The histologic grade of cancer. *Cancer*, **75 Suppl**, 406-21.

Corrente JE, Chalita LVAS, Moreira JA (2003). Choosing between Cox proportional hazards and logistic models for interval-censored data via bootstrap. *J Applied Stats*, **30**, 37–47.

Cox DR (1972). Regression models and life-table. *J Royal Stat Soc B*, **34**, 187-220.

Efron B (1977). The efficiency of Cox's likelihood function for censored data. *J Am Stat Assoc*, **72**, 557-65.

Faycal J, Bessaguet C, Nousbaum JB, et al (2005). Epidemiology and long term survival of gastric carcinoma in the French district of Finistere between 1984 and 1995. *Gastroenterol Clin Biol*, **29**, 23-32.

Harrison JD, Fielding JW (1995). Prognostic factors for gastric cancer influencing clinical practice. *World J Surg*, **19**, 496-500.

Janer G, Sala M, Kogevinas M (2002). Health promotion trials at worksites and risk factors for cancer. *Scand J Work Environ Health*, **28**, 141-57.

Jimeno-Aranda A, Sainz Samitier R, Aragues GM (1996). Gastric cancer in the province of Zaragoza: a survival study. *Neoplasma*, **43**, 199-203.

Jun T, Xiao-Dong W, Zhen–Chun CH (2004). Survival of patient with gastric cancer in changle city of China. *World J Gastroenterol*, **10**, 1543-46.

Kitamura K, Yamaguchi T, Taniguchi H, et al (1996). Clinicopathological characteristics of gastric cancer in the elderly. *Br J Cancer*, **73**, 798-802.

Klein J, Moeschberger M. Survival Analysis: Techniques for Censored and Truncated Data, New York: Springer-Verlag, 1997.

Lawless JF (1998). Parametric models in survival analysis. In Encyclopedia of Biostatistics. Armitage P. Colton T. Wiley: New York, 3254-64.

Lee ET (1992). Statistical Methods for Survival Data Analysis. 2nd Ed. New York: John Wiley Sons Inc.: 9-117.

Lundegardh G, Adami HO, Malker B (1986). Gastric cancer survival in Sweden: lack of improvement in 19 years. *Ann Surg*, **204**, 546-51.

Maehara Y, Oshiro T, Oiwa H, et al (1995). Gastric carcinoma in patients over 70 years of age. *Br J Surg*, **82**,102-5.

Massacesi C, Norman A, Price T, et al (2000). A clinical nomogram for predicting long-term survival in advanced colorectal cancer. *Eur J Cancer*, **36**, 2044-52.

Matley PJ, Dent DM, Madden MV, Price SK (1988). Gastric carcinoma in young adults. *Ann Surg*, **208**, 593–6.

Mitsudomi T, Matsusaka T, Wakasugi K, et al (1989). A clinicopathological study of gastric cancer with special reference to age of the patients; an analysis of 1630 cases. *World J Surg*, **13**, 225–31.

Msika S, Benhamiche AM, Jouve JL, Rat P, Faivre J (2000a). Prognostic factors after curative resection for gastric cancer. A population-based study. *Eur J Cancer*, **36**, 390-6.

Msika S, Benhamiche AM, Rat P, Faivre J (2000b). Long-term prognosis of gastric cancer in the population of Cote-d'Or. *Gastroenterol Clin Biol*, **24**, 649-55.

Nardi A. Schemper M (2003). Comparing Cox and parametric models in clinical studies. *Statist Med*, **22**, 3597-610.

Oakes D (1977). The asymptotic information in censored survival data. *Biometrika*, **64**, 441-8.

Oertli D, Herzog U, Tondelli P (1994). Prognostic factors in resected stomach carcinoma. *Schweiz Med Wochenschr*, **124**, 945-52.

Orbe J, Ferreira E, Nunez-Anton V (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statist Med*, **21**, 3493-510.

Pacelli F, Doglietto GB, Alfieri S, et al (1999). Survival in early gastric cancer: multivariate analysis on 72 consecutive cases. *Hepatogastroenterology*, **46**, 1223-8.

Park YK, Kim JC, Koh YS, et al (2006). Early gastric carcinoma in young patients. *Int Surg*, **91**, 316-9.

Pinheiro PS, van der Heijden LH, Coebergh JW (1999). Unchanged survival of gastric cancer in the southeastern Netherlands since 1982: result of differential trends in incidence according to Lauren type and subsite. *Int J Cancer*, **84**, 28-32.

Pourhoseingholi MA, Hajizadeh E, Moghimi Dehkordi B, et al (2007). Comparing Cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pacific J Cancer Prev*, **8**, 24-6.

Saito H, Osaki T, Murakami D, et al (2006). Effect of age on prognosis in patients with gastric cancer. *Anz J Surg*, **76**, 458-61.

Salvon-Harman JC, Cady B, Nikulasson S, et al (1994). Shifting proportions of gastric adenocarcinomas. *Arch Surg*, **129**, 381-9.

Sánchez-Bueno F, Garcia-Marcilla JA, Perez-Flores D, et al (1998). Prognostic factors in a series of 297 patients with gastric adenocarcinoma undergoing surgical resection. *Brit J Surg*, **85**, 255-60.

Shiu MH, Perrotti M, Brennan MF (1989). Adenocarcinoma of the stomach: a multivariate analysis of clinical, pathologic and treatment factors. *Hepatogastroenterology*, **36**, 7-12.

Talbäck M, Stenbeck M, Rosén M (2004). Up-to-date long-term survival of cancer patients: an evaluation of period analysis on Swedish Cancer Registry data. *Eur J Cancer*, **40**, 1361-72.

Therneau T, Grambsch P (2000). Modeling Survival Data: Extending the Cox Model, New York: Springer-Verlag.

Tuech JJ, Cervi C, Pessaux P, et al (1999). Early gastric cancer: univariate and multivariate analysis for survival. *Hepatogastroenterology*, **46**, 3276-80.

Yokota T, Kunii Y, Teshima S, et al (2000). Significant prognostic factors in patients with early gastric cancer. *Int Surg*, **85**, 286-90.

Yoshida M, Ohtsu A, Boku N, et al (2004). Long-term survival and prognostic factors in patients with metastatic gastric cancers treated with chemotherapy in the Japan Clinical Oncology Group (JCOG) study. *Jpn J Clin Oncol*, **34**, 654-9.

Zhang XF, Huang CM, Lu HS, et al (2004). Surgical treatment and prognosis of gastric cancer in 2,613 patients. *World J Gastroenterol*, **10**, 3405-8.