RESEARCH COMMUNICATION

Survival Analysis of Patients with Esophageal Cancer using Parametric Cure Model

Mahboube Rasouli, Mahmood Reza Ghadimi, Mahmood Mahmoodi*, Kazem Mohammad, Hojjat Zeraati, Mostafa Hosseini

Abstract

Background & Objectives: Esophageal cancer is a major cause of mortality and morbidity in the Caspian littoral north-eastern part of Iran. The aim of this study was to calculate cure function as well as to identify the factors that are related to this function among patients with esophageal cancer in this geographical area. Methods: Three hundred fifty nine cases of esophageal cancer registered in the Babol cancer registry during the period of 1990 to 1991 (inclusive) were followed up for 15 years up to 2006. Parametric cure model was used to calculate cure fraction and investigate the factors responsible for probability of cure among patients. Results: Sample of subjects encompassed 62.7% men and 37.3% women, with mean ages ofdiagnosis was 60.0 and 55.3 years, respectively. The median survival time reached about 9 months and estimated survival rates in 1, 3, and 5 years following diagnosis were 23%, 15% and 13%, respectively. Results show the family history affects the cured fraction independently of its effect on early outcome and has a significant effect on the probability of uncured. The average cure fraction was estimated to be 0.10. Conclusion: As the proportionality assumption of Cox model does not meet in certain circumstances, a parametric cure model can provide a better fit and a better description of survival related outcome.

Key words: Esophageal cancer - long term survivor - parametric cure model - Iran

Asian Pacific J Cancer Prev, **12**, 2359-2363

Introduction

Cancer is one of the important causes of death and disabilities in the world (Parkin et al., 1999; Yazdanbod et al., 2004). It has been received a striking amount of health care resources (Zali, 2005). In 1984, over half of the new cancer cases were reported from developing countries (Parkin et al., 1984). Cancer has been estimated to become the important cause of death in many developed and developing countries including Iran (Ferlay et al., 2004; Yazdanbod et al., 2004).

Esophageal cancer is the sixth common cause of cancer mortality in the world. The incidence of this disease shows a considerable geographic variation in the world (American Cancer Society, 2007). Esophageal cancer has a high incidence in such regions as China, Iran, South Africa, Uruguay, France and Italy (Lu et al., 1999). Most of the esophageal cancer cases in Iran have been reported from the north and northeast regions of the country. Result from a survey by the Iran Cancer Institute, reported that 9% of all cancers and 27% of gastrointestinal cancers were esophageal carcinoma. The male to female ratio was 1.7 to 1 (Ghavamzadeh et al., 2001).

A recent report from the Ministry of Health in Iran shows that more than 70% of deaths are caused by cardiovascular diseases, injuries and cancers, so studying the burden of cancer as one of the three important causes of death in the country is essential (Naghavi, 2000). Esophageal cancer is one of the ten most common diseases worldwide, the five-year survival rate being 3% to 10% (Whelan et al., 1993; Dušek et al., 2005). Results from several epidemiological studies show that hot drinks, alcohol and tobacco are the main risk factors for esophageal cancer (Medvec, 1988; Glade et al., 1999; Bollschweiler et al., 2002; Eloubeidi et al., 2002; Enzinger and Mayer, 2003; Tsottles et al., 2005). Also, geographical distribution is effective in esophageal cancer (Corley and Buffler, 2001; Nyren et al., 2002; Stein et al., 2005; Mohebbi et al., 2008). Highest incidence of esophageal cancer occurs in the age group 50-70 years. Also, the frequency of the disease is higher in men (Nyren et al., 2002; Ferlay et al., 2004; Zendehdel et al., 2007). Theoretically, esophageal cancer may be treatable in its early stages; therefore, early detection is desirable. Survival data are often modeled

Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
**For Correspondence: mahmoodim@tums.ac.ir*

using a Cox proportional hazards model, which is one of the most popular methods to analyze survival data (Cox, 1972). In short follow-up studies, the assumption of a constant risk ratio is very reasonable. In long follow-up studies, however, it is more appropriate to assume that time somehow influences the hazard ratios. When the assumption of proportionality is violated, then the results from a Cox model are not reliable and other modeling approaches should be considered instead.

One approach to model long-term survival studies is through the use of mixture models, known as cure models (Perperoglou et al., 2007). In traditional survival analysis, all subjects in the population are assumed experience the interest event but in some studies a considerable proportion of subjects may be long-term survivor and never experience the event of interest, if the follow-up period is long enough, thus they can be considered cured. A cure model is a mixed model composed of the cure fraction model and the survival model of non-cured subjects which estimates both the cure fraction and the survival function for the uncured. Cure model analysis, introduced 50 years ago, is approximately better suited to the analytic requirements of clinical research in survival data where cure is achieved (Sposto, 2002). In this study a considerable fraction of patients are long term survivor or cured and naive use of Cox regression analysis can be misleading in these circumstances; therefore, parametric cure models (PCMs) are used to analyze our data set. The aim of this paper is to analyze data from prospective study on cases with esophageal cancer, as well as to investigate the proportion of cures among patients and to assess the factors influencing the survival of patients with parametric cure models

Materials and Methods

This survey was a prospective study in which a total of 359 patients were registered at the Babol cancer registration with esophageal cancer between the years 1990-1991, who were followed up for a period of 15 years by the year 2006, and entered into the study. Patients enrolled into the study were at the early stage of the disease as proved by the pathologist diagnosis. The socio-demographic status was obtained through a structured questionnaire.

In this paper parametric cure models (PCMs) are used to analyze data. PCMs can be classified either as mixture (Boag, 1949; Berkson and Gage, 1952; Farewell et al., 1982; Kuk and Chen, 1992) or non-mixture (Haybittle, 1959; Yakovlev, 1996; Gieser et al., 1998; Tsodikov, 1998) types. In mixture models the overall survival of the patients consists of two parts, a survival function $S_U(t/X)$ which models the survival of not-cured patients, denoted by the subscript $_U$, and a probability of a patient been cured $\pi(X^*)$ which depends upon some covariates $X^*$and takes the logistic form log $[\pi(X^*)/\{1-\pi(X^*)\}]$. On the whole, the survival function at time for patients with covariates X and X^* is given as follows:

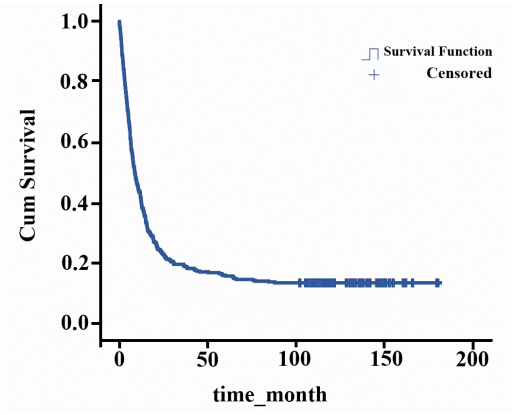

Figure 1. Plot of Kaplan Meier Estimates

$S(t/X, X^*)= [1-\pi(X^*)]\ S_U(T/X)+\pi(X^*)$

Now $S(t/X, X^*)$ is the unconditional survival function of the entire population, $\pi(X^*)=P(C=1/X^*)$ where X^* is a covariate vector that may include exactly the same covariates as X (such as our study) and C is an indicator of cured patients, i.e. C=1 if the patient is cured and C=0 otherwise. The non-mixture cure fraction model was originally developed in the modeling of tumor recurrence (Tsodikov, 2003). This model takes the form, $S(t)=\pi F(t)$.

The hazard function for the non-mixture model is $h(t)=-ln(\pi)f(t)$ so, the model simply employs the fact for any survival function with a cured fraction. In this study we used Weibull mixture cure model with the logistic link. There are theoretical reasons for the suitability of the Weibull as a distribution for survival times.

The covariates to be included in the models are: sex, age, current job, education, province, ethnicity, place of residence, migration status, family history, cigarette smoking. The study was confirmed by the Ethics Committee of Tehran University of Medical Sciences.

Results

Of the 359 patients with esophageal cancer included in this study, 225 (62.7%) were men and 134 (37.3%) were women. The mean age at diagnosis was 55.23±11.01 years. Estimated survival rates in 1, 3, and 5 years following diagnosis were 23%, 15% and 13%, respectively, and estimated percentiles for survival time in 25%, 50% and 75% were 21.8, 9 and 4.1 months, respectively. During the follow up 310 (86.3%) deaths were observed, where 63.2% were men and 36.8% were women, and 49 (13.6%) were still alive or exact details of their survival status were available.

We checked proportionality assumption in SAS, the test of proportionality was significant (p-value=0.017) and the PH assumption was inappropriate in these data. Also, we used a graphical approach (this graphical approach used a plot of Kaplan Meier estimates) to show that our data are a sample from a population containing cured individuals (Figure 1).

As Figure 1 shows, a striking fraction of patients are

Table 1. Estimates of the Weibull Mixture Cure Model for Long-term and Short-term Survivors

Parameter		P-value	OR	CI: 95%	
Long-term	Age (years)	0.3617	1.01427	0.9838	1.0457
	Sex	0.9950	1.0029	0.4032	2.4944
	Province	0.4710	0.74137	0.3280	1.6758
	Place	0.1112	1.71198	0.8829	3.3197
	Positive Family history	0.0226	2.60774	1.1450	5.9393
	Education	0.1999	0.53242	0.2028	1.3978
	Cigarette smoking	0.0565	2.09501	0.9796	4.4803
	Migration	0.5367	0.70627	0.2337	2.1345
	Job (Farmer)	0.9436	0.97011	0.4172	2.2555
	Job (Employee)	0.5930	0.49504	0.0373	6.5644
	Ethnicity (Aryan)	0.6051	0.73668	0.2307	2.3529
	Ethnicity (Gilak)	0.7337	1.53602	0.1288	18.324
	Ethnicity (Torkaman)	0.7106	1.29536	0.3291	5.0990
Short term	Age (years)	0.6554	0.99759	0.9870	1.0082
	Sex	0.5568	1.10036	0.7992	1.5150
	Province	0.0055	1.57995	1.1450	2.1801
	Place	0.3355	0.88826	0.6976	1.1311
	Family history	0.0057	1.42340	1.1088	1.8273
	Education	0.1331	0.71938	0.4678	1.1062
	Cigarette smoking	0.8536	0.97527	0.7469	1.2734
	Migration	0.1889	1.33705	0.8663	2.0636
	Job (Farmer)	0.5291	1.09708	0.8215	1.4650
	Job (Employee)	0.9812	0.98295	0.2333	4.1408
	Ethnicity (Aryan)	0.1299	0.73076	0.4867	1.0972
	Ethnicity (Gilak)	0.0084	0.35860	0.1675	0.7679
	Ethnicity (Torkaman)	0.7918	1.06109	0.6823	1.6501

long term survivor or cured. Also the proportional hazard is not appropriate, so a Cox model is not reliable, thus we analyzed these data with parametric cure model, the model used below is a Weibull model with a logistic link.

Table 1 shows the results obtained from Weibull mixture cure model with the logistic link, the average cure fraction being estimated to be 0.10. Odds ratios and confidence intervals for long-term survivors are given at the top of the table and the hazard ratios and confidence intervals for short-term survivors are given in the bottom of the Table. Among the long-term survivors, odds ratio for a positive family history of cancer is reported to be 2. This means that the risk of being not cured (at the risk to undergo the event of interest) is significantly increased for those people having positive family history (vs. reference group).

Odds ratio is reported to be 2.1 for a cigarette smoking so smoking group has the higher risk associated with being not cured compared to non-smoking group, but P_value for this variable, at 0.05 level of significance, is borderline. There is a little difference between the ethnicity groups among long-term survivors, it is worth mentioning that the risk of being not cured for Gilak group has the highest odds ratio compared to other groups but there is no evidence of significant effect of this variable.

Among short-term survivors, hazard ratio reported 1.4 for a positive family history of cancer so it indicates that desired event (death) is happening faster for the group with positive family history of cancer than for the group without positive family history of cancer. Both parts of the model select this variable as having a significant effect at 0.05 level of significance. So positive family history increases the probability of being uncured and reduces the survival of those people who are uncured. Hazard ratio of 0.97 for a cigarette smoking among short term survivors means that there is no difference between the survival of smoking and non-smoking groups, so there is no evidence of significant effect of this variable.

Discussion

Most studies have shown that family history for esophageal has a strong risk associated with the disease (Bagheri, 1997). Result of an early study in Iran showed that 47% of 427 Turkmen people with esophageal cancer had positive family history for esophageal cancer. The age of onset for 40% of those with positive family history was younger than 50 years (Ghadirian, 1985; Pour and Ghadirian, 1974). In several case-control studies in Iran carried out in the high-risk region, odds ratios has been reported 1.8 to 7 for a positive family history among patients (Bagheri, 1997; Shafieizadeh et al., 2005; Akbari et al., 2006). Two recent studies of familial risk in the high-risk area, one based on a case parent study and the other based on a cohort study, have estimated a more than two-fold increase in the risk of esophageal cancer among first degree relatives (Shafieizadeh et al., 2005; Akbari et al., 2006). Compatible with the findings in Iran, studies addressing the familial aggregation in the other areas of the Asian esophageal cancer belt have reported a higher frequency of a positive family history of esophageal cancer among patients living in high-risk regions compared to low- risk regions (Ghadirian, 1985;

Wang et al., 1992).

Several earlier case–control studies have shown that estimates of the association between smoking and adenocarcinomas of the esophagus and gastric cardia have varied (Li et al., 1989; Levi et al., 1990; Wu-Williams et al., 1990; Palli et al., 1992; Jedrychowski et al., 1993; Kabat et al., 1993; Brown et al., 1994; Gao, 1994; Gonzalez et al., 1994; Vaughan et al., 1995; Zhang et al., 1996). In four case–control studies of a combination of adenocarcinomas of the esophagus and gastric cardia, all identified a statistically significant association between cigarette smoking and disease (Kabat et al., 1993; Brown et al., 1994; Vaughan et al., 1995; Zhang et al., 1996).

In the literature, there are many studies on the field of cancer, through which researchers tend to examine the effects of covariates on patients survival using Cox regression model. The Cox model is a very powerful and useful tool for survival analysis. However, if the assumption of proportionality does not hold, the results might be misleading. In a data set with large follow-up period as the present one, it is natural to assume that the assumption of proportionality will not hold. A systematic study on Cancer Journals showed that only in 5% of studies of cancer in which Cox regression model is used, the assumptions of the model have been investigated (Altman et al., 1985). If presumptions are not met, results of Cox model are seriously under question. As an alternative, parametric cure model can be employed.

The PCMs described in this paper are in many ways suited to analysis of survival data where a significant proportion of patients are cured. In situations where the PH assumption can reasonably be assumed to apply, PCMs do not provide better or more efficient analyses compared to Cox analysis. In these situations there may be little to choose between the two approaches, and one could argue that the Cox model is preferable because it does not rely on parametric assumptions on the underlying failure process but when the PH assumption are not appropriate, PCMs provide better or more efficient analyses compared to Cox analysis. One advantage of PCMs is that they provide a coherent statistical approach to investigate the effects of covariates on the time of failure separately from their effects on ultimate outcome.

One limitation of this study is the absence of clinical variables (including type of esophageal cancer (adenocarcinoma, squamous), stage of disease, tumor size, metastatic status because clinical variables in the Babol cancer registry were not recorded.

Acknowledgements

The authors thank Iranian National Institute of Health Research (NIHR) Tehran University of Medical Science for data gathering, financial support and collaboration in this study.

References

Akbari M, Malekzadeh R, Nasrollahzadeh D, et al (2006). Familial risks of esophageal cancer among the Turkmen population of the Caspian littoral of Iran. *Int J Cancer*, **119**, 1047-51.

Altman DG, Destavola BL, Love SB, et al (1985). Review of survival analyses published in cancer journals. *Br J Cancer*, **72**, 511-8.

American Cancer Society (2007). Cancer Facts and Figures 2007. Atlanta, Georgia. American Cancer Society.

Bagheri M (1997). Study of risk factors of GI cancer in Mazanderan Provience. PhD Thesis.

Berkson J, Gage RP (1952). Survival curves for cancer patients following treatment. *J Am Statist Assoc*, **47**, 501-15.

Boag JW (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J R Statist Soc B*, **11**, 15-44.

Bollschweiler E, Wolfgarten E, Nowroth T, et al (2002). Vitamin intake and risk of subtypes of esophageal cancer in Germany. *J Cancer Res Clin Oncol*, **128**, 575-80.

Brown L, Silverman DT, Pottern LM, et al (1994). Adenocarcinoma of the esophagus and esophagogastric junction in white men in the United States: alcohol, tobacco, and socioeconomic factors. *Cancer Causes Control*, **5**, 333-40.

Corley D A, Buffler PA (2001). Oesophageal and gastric cardia adenocarcinomas:analysis of regional variation using the Cancer Incidence in Five Continents database. *Int J Epidemiol*, **30**, 1415-25.

Cox DR (1972). Regression models and life tables(with Discussion). *J R Statist Soc B*, **34**, 187-220.

Dusek L, Muzik J, Kubasek M,et al (2005). The Czech national web portal of cancer epidemiology [online]. Masaryk University, 26.3.2007 [cit. 2007-3-26]. Incidence and mortality C15 – malignant neoplasm of esophageal, time trend 1977–2002. Available from: URL:http://www.svod.cz.

Eloubeidi MA, Desmond R, Arguedas MR, et al (2002). Prognostic factors for the survival of patients with esophageal carcinoma in the U.S.: the importance of tumor length and lymph node status. *Cancer*, **95**, 1434-43.

Enzinger PC, Mayer RJ (2003). Esophageal cancer. *N Engl J Med*, **349**, 2241-52.

Farewell VT (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041-6.

Ferlay J, Bray F, Pisani P, et al (2004). Globocan 2002: cancer incidence.Mortality and prevalence worldwide. IARC cancer base no. 5,version 2.0. Lyon (France): IARC Press.

Gao Y, Mclaughlin JK, Blot WJ, et al (1994). Risk factors for esophageal cancer in Shanghai, China. I. Role of cigarette smoking and alcohol drinking. *Int J Cancer*, **58**, 192–6.

Ghadirian P (1985). Familial history of esophageal cancer. *Cancer*, **56**, 2112-6.

Ghadirian P, Stein GF, Gorodetzky C, et al (1985). Oesophageal cancer studies in the Caspian littoral of Iran: some residual results, including opium use as a risk factor. *Int J Cancer*, **35**, 593-7.

Ghavamzadeh A, Moussavi A, Jahani M, et al (2001). Esophageal cancer in Iran. *Semin Oncol*, **28**, 153-7.

Gieser P, Chang MN, Rao PV, et al (1998). Modelling cure rates using the Gompertz model with covariate

information. *Statist Med*, **17**, 831-9.
Glade MJ (1997). Nutrition 1999. Food, nutrition, and the prevention of cancer: a global perspective. American Institute for Cancer Research / World Cancer Research Fund. American Institute for Cancer Research, 15, 523-6.
Gonzalez C, Agudo A, Montes J, et al (1994). Tobacco and alcohol intake in relation to adenocarcinoma of the gastric cardia in Spain. *Cancer Causes Control*, **5**, 88-9.
Haybittle JL (1959). The estimation of the proportion of patients cured after treatment for cancer of the breast. *Br J Radiology*, **32**, 725-33.
Jedrychowski W, Boeing H, Wahrendorf J, et al (1993).Vodka consumption, tobacco smoking and risk of gastric cancer in Poland. *Int J Epidemiol*, **22**, 606–13.
Kabat G, Ng SK, Wynder EL, (1993). Tobacco, alcohol intake, and diet in relation to adenocarcinoma of the esophagus and gastric cardia. *Cancer Causes Contro*l, **4**, 123-32.
Kuk A, Chen CH (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79**, 531-41.
Levi F, Ollyo JB, Vecchia C, et al (1990). The consumption of tobacco, alcohol and the risk of adenocarcinoma in Barrett's oesophagus. *Int J Cancer*, **45**, 852–4.
Li J, Ershow AG, Chen ZJ, et al (1989). A case-control study of cancer of the esophagus and gastric cardia in Linxian. *Int J Cancer*, **43**, 755–61.
Lu S, Lin P, Wang G, et al (1999). Comprehensive prevention and treatment for esophageal cancer. *Chin Med J*, **112**, 918-23.
Medvec BR, (1988). Esophageal cancer: treatment and nursing interventions. *Semin Oncol Nurs*, **4**, 246-56.
Mohebbi M. Mahmoodi M, Wolfer R, et al (2008). Geographical spread of gastrointestinal tract cancer incidence in the Caspian Sea region of Iran: spatial analysis of cancer registry data. *BMC Cancer*, **8**, 137.
Naghavi M, (2000). Death report from 10 provinces in Iran. Tehran Ministry of Health.
Nyren O, Adami HO, Hunter D, et al (2002). Esophageal Cancer:Textbook of Cancer Epidemiology, New York, Oxford University Press.137-61.
Palli D, Bianchi S, Decarli A, et al (1992). A case-control study of cancers of the gastric cardia in Italy. *Br J Cancer*, **65**, 263–6.
Parkin DM, Stjernsward J, Muir CS (1984). Estimates of the worldwide frequency of 12 major cancers. *Bull WHO*, **62**, 163-82.
Parkin DM, Pisani P, Ferlay J (1999). Estimates of the worldwide incidence of 25 major cancers in 1990. *Int J Cancer*, **80**, 827-41.
Perperoglou A, Keramopoullos A, Houwelingen H (2007). Approaches in modelling long-term survival: An application to breast cancer. *Statist Med*, **26**, 2666-85.
Pour P, Ghadirian P (1974). Familial cancer of the esophagus in Iran. *Cancer*, **33**, 1649-52.
Shafieizadeh T, Holakoiee K, Fotouhi A, et al (2005). Familial esophageal cancer in Babaol M.S. School of Public Health, Tehran University of Medical Sciences.
Sposto R (2002). Cure model analysis in cancer: an application to data from the Children's Cancer Group. *Statist Med*, **21**, 293-312.
Stein HJ, Vonrahden BH, Siewert JR (2005). Survival after oesophagectomy for cancer of the oesophagus. *Langenbecks Arch Surg*, **390**, 280-5.
Tsodikov A (1998). A proportional hazards model taking account of long-term survivors. Biometrics, 54, 1508–16.
Tsodikov A., Ibrahim JG., Takovlev A (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. J Am Statist Ass, 98, 1063-78.
Tsottles ND, Reedy AM (2005). Esophageal cancer. In Cancer nursing : principles and practice (CH Yarbro, MH Frogge & M Goodman, eds). Jones and Bartlett,Boston, Mass.
Vaughan T, Davis S, Kristal A, Thomas DB (1995). Obesity, alcohol, and tobacco as risk factors for cancers of the esophagus and gastric cardia: adenocarcinoma versus squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev*, **4**, 85-92.
Wang Y, Han XY, Su W, et al (1992). Esophageal cancer in Shanxi Province, People's Republic of China: a case-control study in high and moderate risk areas. *Cancer Causes Control*, **3**, 107-13.
Whelan SL, Parkin DM, Masuyer E (1993). Trends in cancer incidence and mortality. Lyon: IARC Scientific Publications 102.
Wu WA, Yu MC, Mack TM (1990). Life-style, workplace, and stomach cancer by subsite in young men of Los Angeles county. *Cancer Res*, **50**, 2569-76.
Yakovlev AY (1996). Threshold models of tumor recurrence. *Math Computer Model*, **23**, 153-64.
Yazdanbod A, Nasseri S, Malekzadeh R (2004). Upper gastrointestinal cancer in Ardabil, North West of Iran: A review. *Arch Iranian Med*, **7**, 173-7.
Zali M (2005). Indices related to gastric cancer in Tehran and seven city provinces in the years 1999 to 2002. *J Islamic Azad University Med*, **15**, 15-8.
Zendehdel K, Nyren O, Edberg A, et al (2007). Risk Indicators for Esophageal Cancer Some Medical Conditions and Tobacco-Related Factors. Ph.D.
Zhang Z, Kurtz RC, Sun M, et al (1996). Adenocarcinomas of the esophagus and gastric cardia: medical conditions, tobacco, alcohol, and socioeconomic factors. *Cancer Epidemiol Biomarkers Prev*, **5**, 761-8.