RESEARCH COMMUNICATION

Determining of Prognostic Factors in Gastric Cancer Patients Using Artificial Neural Networks

Akbar Biglarian[1]*, Ebrahim Hajizadeh[2], Anoshirvan Kazemnejad[2], Farid Zayeri[3]

Abstract

    <u>Background & Objectives</u>: The aim of this study is to determine diagnostic factors for Iranian gastric cancer patients and their importance using artificial neural network and Weibull regression models. <u>Methods</u>: This study was a historical cohort study with data gathered from 436 registered gastric cancer patients who underwent surgery between 2002 and 2007 at the Taleghani Hospital (a referral center for gastrointestinal cancers), Tehran, Iran. In order to determine risk factors and their importance, neural network and Weibull regression models were used. <u>Results</u>: The Weibull regression analysis showed that lymph node metastasis and histopathology of tumor were selected as important variables. Based on the neural network model, staging, lymph node metastasis, histopathology of tumor, metastasis, and age at diagnosis were selected as important variables. The true prediction of neural network was 82.6%, and for the Weibull regression model, 75.7%. <u>Conclusion</u>: The present study showed that the neural network model is a more powerful tool in determining the important variables for gastric cancer patients compared to Weibull regression model. Therefore, this model is recommended for determining of risk factors of such patients.

Keywords: Gastric cancer - prediction - survival analysis - Weibull regression - artifitial neural network

Introduction

    Gastric cancer (GC) is the cancerous growth of stomach cells (CancerFacts, 2005-2008) which can develop in any part of the stomach and may spread to other organs; particularly the esophagus, lungs and the liver (WHO, 2006). GC is the fourth most commonly occurring cancer throughout the world (Inoue et al., 2005). Nearly one million people die from this cancer worldwide per year (WHO, 2006). Generally, GC has reported the second most common cancer related cause of death in the world (Inoue et al., 2005; Crew et al., 2006). The reports show that nearly 65-70% of the incident cases and deaths resulting from GC in 2002 have been occurred in less developed countries (Ferlay et al., 2004). Most countries have reported a high mortality rate (70-90%), except in Japan (about 40%) (Terry et al., 2002).

    In the past two decades, in spite of the promotion of hygiene in Iran, death rates due to cancers have remained as a major health problem in Iranian people (Mohagheghi et al., 2004; Sajadi et al., 2005). Epidemiologic studies in Iran show that the incidence of GC for men estimated to be 26.1 and in women is about 11.1 per one hundred thousand people (Sajadi et al., 2005), also GC is the second most prevalent cancer in men and the fourth in total population. This cancer is very lethal as a fact that the diagnosis of most of patients occurs in the advanced stages (Mohagheghi et al., 2004; Sajadi et al., 2005; Hajiani et al., 2006).

    In cancer research, determining of prognostic factors and survival rate of patients is very important aim. In the previous decades, the data analysts have used a various survival methods for this aim. In this context, the traditional survival methods such as Cox regression modeling or parametric regression modeling are used for analyzing of survival data sets. However, when we use this model some underlying assumptions should be considered (Chia et al., 2002; Lee & Wenyu 2003). When these assumptions are not satisfactory, we should utilize other free assumption modeling approaches such as artificial neural network (ANN).

    In this study, we have developed the ANN model to determine the risk factors of gastric cancer patients.

Materials and Methods

Patients and factors

    In this study, we analyzed the data from 436 patients with stomach cancer who underwent surgery between 2002 and 2007 at the Taleghani Hospital (a referral center for gastrointestinal cancers), Tehran, Iran. Before surgery, all patients underwent endosonography, biopsy and histopathologic examination. This study was conducted by the Research Center for Gastroenterology and Liver

[1]Department of Biostatistics, University of Social Welfare and Rehabilitation (USWR), [2]Department of Biostatistics, Faculty of Medical Science, Tarbiat Modares University, [3]Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran *For Correspondence : abiglarian@gmail.com

Akbar Biglarian et al

Disease of Shahid Beheshti University of Medical Sciences, Tehran, Iran. In this historical cohort study, the required information for each patient including age at diagnosis, sex, familial history, histology type, lymph nodes metastasis, histopathology of tumor, tumor stage, metastasis was gathered from his/her registered documents in the Cancer Archive of Taleghani Hospital. We also registered the survival time of each patient (in month) after total or subtotal gastrectomy.

Statistical and Neural Network analysis

In this study, at first, to determine of risk factors and to predict the survival rate of the patients, we utilized a Wiebul parametric regression model.

At the second, an ANN model was used. In the ANN modeling process, we randomly divided the data into two subsets: 263 patients (nearly 60%) for constructing the models (training subset) and the remaining (nearly 40%) for testing the model (as the validation subset). After evaluating the model, we applied multiple layer perceptron (MLP) networks to determine important risk factors. In this context, we use independent variable importance analysis and by using normalized importance, the risk factors were determined. In addition, the area of under receiver operation characteristic (AUROC) curve was used for comparing the prediction power of the described models. Note that, in fitting ANN model we used a three-layer MLP network with 8 variables (17 nodes) in input layer, 8 nodes in middle layer and one node in output layer. Because of patient's status was bi-state (dead or censored), the logistic transfer function was utilized as the activation function in middle and output layers. In addition, we have utilized back-propagation learning algorithm with learning rate of 0.3 and momentum of 0.9 for learning net.

For the data analysis, we used the SPSS version 16.0 and the S-PLUS 8.0 softwares.

Results

Of the 436 GC patients 161 (36.9%) were died and others were censored; 315 patients (72.2%) were men and others were women. The mean±SE and median±SE of survival times were 32.13±1.48 and 27.50±2.72 respectively.

In the first step of the modeling process, the data was divided in training (nearly 60% of patients) and testing (nearly 40% of patients) subsets. The Mantel-Cox test showed that the estimated survival curves using the training and testing subsets have no significant difference (P=0.583). Afterward 468 model of MLP (three-layer) was fitted for 13 topology based on 6 to 18 nodes in middle layer, and with 0.8 to 0.95 momentum, and learning rate 0.01 to 0.40. At concluding, the model with eight hidden node, learning rate of 0.3, and momentum of 0.90 was preferred as suitable model. For this model, maximum of sum of square of error prediction was 21.5, AUROC was 0.826 and correct predictions rate was 81.50%.

In the next step, based on validation set, the Weibull regression and NN models were used to determine the risk factors. The results were presented in Table 1.

For comparing the accuracy of the models' prediction,

Table 1. Weibull and ANN Modeling Results for Determining the Effect of Prognostic Factors on GC Patients' Survival

Weibull Regression		ANN Model	
Ordered Variables	P_value	Ordered Variables	Normalized Importance
Lymph node metastasis	0.010	Staging	0.217
Histopathology of tumor	0.084	Lymph node metastasis	0.178
Familial history	0.121	Histopathology of tumor	0.153
Age	0.158	Metastasis	0.143
Metastasis	0.194	Age	0.105
Histology type	0.259	Histology type	0.074
Sex	0.615	Familial history	0.073
Staging	0.910	Sex	0.057

Table 2. Classification Accuracy of ANN and Weibull Models in Testing Subset

Status	Observed (No)	True prediction by ANN No (%)	True prediction by Weibull No (%)
Dead	62	51 (82.2)	49 (79.0)
Survived	111	92 (82.9)	82 (73.9)
Total	173	143 (82.6)	131 (75.7)

we used true classification (the proportion of patients that classified correctly in dead and survived groups) of the patients in the testing subset. The obtained results were presented in Table 2. As we can see, the ANN model lead to more accurate predictions compared to the Weibull model (true prediction of 82.6% vs. 75.7%). The area under ROC curve, calculated from testing data, for ANN model was 0.86, and for Weibul model, 0.77.

Discussion

Gastric cancer is the most prevalent malignancy in Iran and in the world. The main aims of this study were identifying some of the most prognostic factors of GC and comparing the ability of Weibull regression and neural network models in predicting the survival of the GC patients.

In this study, the Weibul regression analysis showed that the patients' survival time was related to lymph node metastasis, histopathology of tumor, familial history, age at diagnosis, metastasis, histology type, sex, and staging respectively. In this analysis, only lymph node metastasis has had significant relationship with survival time and this factor with histopathology of tumor have had more importance. In ANN strategy, staging, lymph node metastasis, histopathology of tumor, metastasis, age, histology type, familial history, and sex was respectively as importance variables. Of these variables, staging, lymph node metastasis, histopathology of tumor, metastasis, and age at diagnosis have had more importance. This result may be throughout interaction terms between variables that the ANN model is considers it.

Published studies have reported lymph node metastasis is the most important prognostic factors in patients with GC (Siewert et al., 1993; Sendler et al., 1995; De Manzoni

et al., 1996; Siewert et al., 1998; Bonenkamp et al., 1999; Schröder et al., 2001).

Moreover, other studies have referred to other risk indicators such as gender, number of involved lymph nodes, histological type, and type of complementary treatment, as the significant effective factors for survival of the GC patients (Ghiandoni et al., 1998; Borie et al., 2004; Toneri et al., 2006; Liu et al., 2007; Peyre et al., 2008; Ozgüç et al., 2008).

Lai et al. conducted an ANN-based study for predicting tumor staging in GC patients. They reported an accuracy of 81.8% for predicting the tumor stage in primary GC patients (Lai et al., 2008). In another study in Taiwan, Chien et al. used the ordinary logistic regression, ANN and decision tree methods for predicting post-operative complications of GC patients. Results of their study indicated that the ANN was better technique for predicting the post-operative complications compared to logistic regression and decision tree methods (Chien et al., 2008). In the present study, we compared the results of Weibull regression and ANN model in determining important risk factors and true prediction of GC patients. Our findings indicate that the ANN is a proper technique for this aim.

In conclusion, use of the ANN model for determining prognostic factors of GC patients is more efficient than Weibull regression model. Therefore, this strategy is suggested to determining the important risk factors of survival of GC patients.

Acknowledgement

References

Bonenkamp JJ, Hermans J, Sasako M, et al (1999). Dutch Gastric Cancer Group: extended lymph-node dissection for gastric cancer. *N Engl J Med*, **340**, 908-12.

Borie F, Rigau V, Fingerhut A, et al (2004). French Association for Surgical Research. Prognostic factors for early gastric cancer in France: Cox regression analysis of 332 cases. *World J Surg*, **28**, 686-91.

CancerFactsMD.com, (2005-2008). Stomach (Gastric) Cancer. Available at: http://www.cancerfactsmd.com/stomach-gastric-cancer/

Chien CW, Lee YC, Ma T, et al (2008) The application of artificial neural networks and decision tree model in predicting post-operative complication for gastric cancer patients. *Hepatogastroenterology*, **55**,1140-5.

Chih-Lin Chia, W Nick Street b, William H Wolbergc (2002). Application of Artificial Neural Network-Based Survival Analysis on Two Breast Cancer Datasets. American Medical Informatics Association Annual Symposium (AMIA 2007), 130-134, Chicago, IL, November 2007.

Crew KD, Neugut AI (2006). Epidemiology of gastric cancer. *World J Gastroenterol*, **12**, 354-62.

De Manzoni G, Verlato G, Guglielmi A, et al (1996). Prognostic significance of lymph node dissection in gastric cancer. *Br J Surg*, **83**, 1604-8.

Ferlay J, Bray F, Pisani P, et al (2004). GLOBOCAN 2002: cancer incidence, mortality and prevalence worldwide, Version 2.0.IARC CancerBase no 5. Lon: IARC Press, 2004, (http://www-depdb.iarc.fr/globocan2002.htm).

Ghiandoni G, Rocchi MB, Signoretti P, et al (1998). Prognostic factors in gastric cancer evaluated by using Cox regression model. *Minerva Chir*, **53**, 497-504.

Hajiani E, Sarmast Soshtari MH, Masjedizadeh R, et al (2006). Clinical profile of gastric cancer in Khuzestan, southwest of Iran. *World J Gastroenterol*, **12**, 4832-5.

Inoue M, Tsugane S (2005). Epidemiology of gastric cancer in Japan. *Postgrad Med J*, **81**, 419-24.

Lai KC, Chiang HC, Chen WC, et al (2008). Artificial neural network-based study can predict gastric cancer staging. *Hepatogastroenterology*, **55**, 1859-63.

Lee ET, Wenyu WJ (2003). Statistical methods for survival data analysis, John Wiley & Sons, Inc, pages, 8-15.

Liu C, Lu P, Lu Y, et al (2007). Clinical implications of metastatic lymph node ratio in gastric cancer. *BMC Cancer*, **7**, 200.

Mohagheghi MA (2004). Annual Report of: Tehran Cancer Registry 1999. The Cancer Institute Publication.

Ozgüç H, Sönmez Y, Yerci O (2008). Metastatic/resected lymph nodes ratio-based classification in gastric cancer. *Turk J Gastroenterol*, **19**, 2-7.

Peyre CG, Hagen JA, DeMeester SR, et al (2008). The number of lymph nodes removed predicts survival in esophageal cancer: an international study on the impact of extent of surgical resection. *Ann Surg*, **248**, 549-56.

Sajadi A, Nouraei M, Mohagheghi MA, et al, (2005). Cancer Occurrence in Iran in 2002, an International Perspective: *Asian Pac J Cancer Prev*, **6**, 359-63.

Sajadi A, Raafat J, Mohagheghi MA, et al (2005). Gastric carcinoma, 5 years experience of a single institute, *Asian Pac J Cancer Prev*, **6**, 195-6.

Schröder W, Hölscher AH (2001). Aktuelle therapiestrategien beim Magenfrühkarzinom. *Onkologe*, **7**, 610-22.

Sendler A, Dittler AJ, Feussner H, et al (1995). Preoperative staging of gastric cancer as precondition for multimodal treatment. *World J Surg*, **19**, 501-8.

Siewert JR, Böttcher K, Roder JD, et al (1993). Prognostic relevance of systematic lymph node dissection in gastric carcinoma. *Br J Surg*, **80**, 1015-8.

Siewert JR, Böttcher K, Stein HJ, et al (1998). Relevant prognostic factors in gastric cancer: ten-year results of the German gastric cancer study. *Ann Surg*, **288**, 449-61.

Terry MB, Gaudet MM, Gammon MD (2002). The

epidemiology of gastric cancer. *Gastric Cancer*, **12**, 109-6.

Toneri M, Mitsumoto Y, Shimabukuro M, et al (2006). Prognosis of patients with gastric carcinoma treated in remote island hospitals in Japan: a multivariable analysis. *Anticancer Res*. **26**, 3621-5.

WHO (February 2006),Cancer. Retrieved on 2007-05-24.