Ann Colorectal Res. 2014 June; 2(2): e22329. DOI: 10.17795/acr-22329
Published online 2014 June 30.

Research Article

Effective Attributes in Colorectal Cancer Relapse Using Artificial Neural Network and Cox Proportional Hazards Regression

Saeedeh Pourahmad [1,2]; Bahareh Khosravi [2]; Mohammad Mohamadianpanah [1,*]

[1]Colorectal Research Center, Faghihi Hospital, Shiraz University of Medical Sciences, Shiraz, IR Iran
[2]Department of Biostatistics, Medical School, Shiraz University of Medical Sciences, Shiraz, IR Iran

*Corresponding author: Mohammad Mohamadianpanah, Colorectal Research Center, Faghihi Hospital, Shiraz University of Medical Sciences, P. O. Box: 71936, Shiraz, IR Iran. Tel: +98-7116125168, Fax: +987116474320, E-mail: mohpanah@gmail.commohpanah@sums.ac.ir

Received: July 26, 2014**; Revised:** August 17, 2014**; Accepted:** August 19, 2014

Background: The use of statistical methods to analyze data, regardless of their theoretical assumptions, leads to misinterpretation of the results.
Objectives: Effective attributes in colorectal cancer relapse were investigated through survival analysis in the present study. Comparison between the results of artificial neural network (ANN) method and Cox proportional hazards (Cox PH) model was the main purpose of this research.
Patients and Methods: A total of 184 patients with locoregional colorectal cancer, referred to Shahid Faghihi Hospital (Shiraz, Iran) for surgery, were followed in a five-year period for possible relapse during 2003-2011. Disease-free survival was then modeled based on the patients' attributes, using Cox PH regression and ANN methods. All the attributes effective on disease relapse were investigated by these two methods.
Results: A total of 114 (62%) males and 70 (38%) females with a median age of 54 (range: 23-84) years old participated in the study. Among them, there were 95 (51.6%) patients with colon cancer and 89 (48.4%) with rectum cancer. In addition, 53 patients relapsed and 131 patients did not present any relapse or missed the follow up (censored data). The results showed that the accuracy rate in prediction was higher for the ANN method than the Cox PH model (78.2% versus 72.7%). In addition, the area under the receiver operating curve (ROC) was also more for the ANN method (0.86 versus 0.74). Five attributes of the patients, including neoadjuvant treatment, perforation and/or obstruction, perineural invasion, stage, and tumor grade, were significant through the Cox HP model. The first five attributes by the ANN method were surgeon, primary tumor site, perforation and/or obstruction, age, and adjuvant treatments. In this study, the order of attributes determined by the ANN method was rather confirmed by the physicians.
Conclusions: The results showed superiority of the ANN method over the Cox PH model with respect to the area under the ROC and the accuracy rate in prediction. However, this method requires a large data set to learn the relations and cannot distinguish the confounding attributes.

*Keywords:*Colorectal Cancer; Artificial Neural Network Method; Cox Proportional Hazards Model; Relapse

1. Background

Colorectal cancer is the third most common cancer in western countries and the fourth most common cancer in the world. For industrialized countries, it is known as the second cause of cancer mortality after lung cancer. In Iran, this cancer is the fourth common cancer after skin, breast, and stomach cancers (1). Fortunately, although being progressive and fatal, it is preventable. According to the Iranian Annual National Cancer report, the disease affects males and females equally and it occurs commonly after the age of 50; however, it may occur earlier in hereditary and familial cases. Statistics in Iran reveal that half of the patients are less than 50 years old and it is estimated that 6% of people after the age of 50 surely get the disease (2). Therefore, it seriously affects emotional issues, social and economic statuses of individuals, families, and ultimately the society. Furthermore, relapse of the disease may have more undesirable consequences. Therefore, determining the effective attributes in relapse occurrence may be important. In this issue, duration of the disease-free period after the surgery can be modeled based on patients' attributes through survival analysis.

Among the statistical methods appropriate for survival analysis, Cox proportional hazards (PH) regression model is frequently applied in clinical studies (3, 4). Although this method is a famous approach for modeling the survival data, it has an assumption of stability of the hazards ratio and/or independence of the event time, which must be considered when using the method or interpreting the results (5). In this regard, some modifications have been made to overcome the limitations of Cox PH regression model or replace it by more appropriate methods. For instance, applying weighted estimation in Cox PH regression (6) and using parametric models as an alternative for the Cox PH regression model (7) were two recent attempt in this context. However, methods with less theoretical or statistical concepts and the ones which do not take any assumptions for data have been of the most interest; especially in clinical researches (8). Artificial neural network (ANN) is such a method, frequently used for modeling complex relations without any underlying assumptions for data structure. Analysis of huge datasets with a large number of attributes is another characteristic of the ANN method (9). This method has recently attracted more attention in modeling various relations, including survival data (10-12). Some advantages and disadvantages of the ANN method have been mentioned in clinical studies compared with Cox PH regression analysis (13-15). The survival rate of patients with colorectal cancer has also been modeled in different aspects by Kaplan Meier (16), Cox PH regression (17, 18) and ANN methods (19, 20).

2. Objectives

In this paper, the effectiveness of attributes on the time of relapse as well as on disease-free survivals of colorectal cancer was investigated in an Iranian population. Cox PH regression and ANN methods were applied on a real dataset and their results were compared using the correct prediction accuracy percentage, area under the receiver operating curve (ROC), and order of effective attributes.

3. Patients and Methods

3.1. Study Population

A total of 184 patients with histologically proven resected locoregional invasive colorectal adenocarcinoma were enrolled in this study. The patients were referred to Namazi Hospital, Shiraz University of Medical Sciences, during 2003-2011; we did not involve patients presenting in situ or metastatic disease, with pathologies other than adenocarcinoma, and with unresectable or inoperable disease. In addition, patients who achieved complete pathological responses following neoadjuvant chemoradiation were excluded. We also excluded those with missing or incomplete medical records or lacking complete pathological reports. All the patients underwent standard curative surgical resection for their locoregional colorectal cancers. Tumors were pathologically restaged according to the American Joint Committee on Cancer (AJCC) Tumor Node Metastasis (TNM) staging system, 7th edition (21). The initial evaluation included comprehensive history and physical examination, colonoscopy, and chest, abdominal, and pelvic computed tomography (CT) scans. Pelvic magnetic resonance imaging (MRI) and/or transrectal ultrasonography was considered for the rectal primary site.

3.2. Statistical Analysis

Survival analyses including Cox PH regression and ANN modeling method were done using Matlab software. Disease-free survival rate was defined as the percentage of patients free of colorectal cancer after five years. The disease-free survival durations were measured from the date of initial treatment till any type of treatment failure or the last follow-up. All the potential tumors and patients' characteristics were analyzed for their impact on the disease-free survival rates. Cox PH regression model is a mathematical model for analysis of survival data with two theoretical assumptions, including the stability of hazards ratio and independence of the event times. Cox PH model is applied when predicting attributes do not depend on the time and the hazards ratio are stable over time. In addition, the time of event occurrence must be independent for individuals (5). Validity of the results depends on the presence of these assumptions. In the present study, to determine the effective attributes on the disease-free survival period in colorectal cancer, Cox PH model along with a three-layer ANN were applied. All the available information of the patients including 18 attributes were applied as the predictor variables in the models. To estimate the coefficients in Cox PH regression model, the maximum likelihood ratio approach was applied with backward conditional method for variables selection. For the designed ANN, one input, one hidden, and one output layer were considered. There were 18 nodes (neurons) in the first layer (the number of inputs); 5-20 nodes in the hidden layer (to choose the best design), and one node in the output layer. Other characteristics considered in designing ANN were: back-propagation forward learning algorithm, sigmoid activation function, learning rate from 0.01 to 0.4, and momentum from 0.8 to 0.95. The best design of ANN was chosen according to the prediction accuracy rate and the area under the ROC. In both methods, data sets were randomly divided to two sets; 70% as the training set for learning and 30% as the testing set forv alidation. The results were reported on the validation set for the best design of ANN and Cox PH regression. The percentage of correct prediction accuracy, the area under the ROC, and the order of attributes input in relapse were compared between the two mentioned methods.

4. Results

A total of 184 patients participated in this study, including 114 (62%) males and 70 (38%) females, with a median age of 54 (range: 23-84) years old. Among them, there were 95 (51.6%) patients with colon cancer and 89 (48.8%) with rectum cancer. Patients were followed up for a median of five years after the surgery to observe any possible relapse. Accordingly, the dataset in this research included 53 patients with relapse and 131 patients without any relapse or missing the follow-up (censored data). Table 1 describes the patients' attributes used in the modeling process as inputs and output variables. The results were reported on the validation sets (55 patients) for both models, except for important attributes which were determined from the training sets (129 patients). Based on the results of Cox PH model, five of 18 inputs were effective attributes for prediction of the disease relapse, including neoadjuvant treatment, perforation or obstruction, perineural invasion, stage, and tumor grade, respectively. In fact, the coefficients of these variables were significant at 0.1 levels. The area under the ROC for this model was 0.74, which was statistically significant (P = 0.007) (Table 2). In addition, the percentage of correct prediction for Cox PH model was 72.7% on the validation set (Table 3). In the ANN model, an initial weight was randomly assigned to each input and these weights were updated during the training process to achieve the prediction minimum error (between the actual outputs and the predicted ones). Therefore, all the inputs were entered to the modeling process and assigned with weights. In this study, the absolute value of the ultimate weight for each input was considered as a criterion to select the important attributes in prediction ofrelapse. Table 4 shows the order of inputs in ANN model compared with Cox PH model. The area under the ROC for the ANN model was 0.86 (P<0.001) (Table 2); the accuracy rate of prediction for this method was 78.2%.

5. Discussion

Comparison between the results of ANN and Cox PH methods in determining the effectiveness of attributes in disease relapse was the main purpose of the present study. The results showed that the accuracy rate in prediction was higher for the ANN method compared to Cox PH model (78.2% versus 72.7%). In addition, the area under the ROC was more for the ANN method compared with Cox PH (0.86 versus 0.74). However, both of them were high enough to be statistically significant (P < 0.01). Many studies have compared these methods for survival analysis in various diseases (13-15). For instance, they were compared in a study to determine the prognostic factors and predict the survival probability of gastric cancer patients. The results confirmed the superiority of ANN model in determining the significant prognostic variables for these patients compared with the Cox PH model (14). In the case of colorectal cancer, different methods have been used for survival analysis in previous researches such as Kaplan Meier method (16), Cox PH regression model (17, 18), and ANN method (19), as well as their comparisons (20). However, they were different with the present study in terms of the events definition and survival time, and also the comparative method of the models. The results of this study confirmed that Cox PH model applied a subset of attributes in the final model (the significant ones), whereas the ANN method used all patients' attributes in the modeling process and the absolute value of weights indicated their importance. Overall, Cox PH model needs to admit some theoretical assumptions on its data structure; however, its results are easy to interpret and the odds ratio and related confidence intervals can be calculated.

Table 1. Attributes Description of 184 Patients Applied in the Modeling Process

	No. (%)
Event (output)	
Not relapsed	131 (71.2)
Relapsed	53 (28.8)
Predictors variables (inputs)	
Gender	
Male	114 (62)
Female	70 (38)
Cancer site	
Colon	95 (51.6)
Rectum	89 (48.4)
T stage	
T0-T2	44 (23.9)
T3	140 (76.1)
Stage	
0-2	119 (64.7)
3	65 (35.3)
Grade	
Well differentiated	125 (67.9)
Moderately or poorly differentiated	59 (32.1)
Lymphatic-vascular invasion	
Yes	110 (59.8)
No	74 (40.2)
Perineural invasion	
Yes	158 (85.9)
No	26 (14.1)
Perforation or obstruction	
Yes	147 (79.9)
No	37 (20.1)
Surgeon	
Colorectal	45 (24.5)
Non-Colorectal	139 (75.5)
Laboratory	
Academic	58 (31.5)
Private	126 (68.5)
Neoadjuvant treatment [a]	
Not received	164 (89.1)
Received	20 (10.9)
Adjuvant[a] **treatment**	
Radiotherapy + chemotherapy	111 (60.3)
Chemotherapy alone	73 (39.7)
Adjuvant chemotherapy regimen	
5-Fu + LV	67 (36.4)
FOLFOX	77 (41.8)
Others	40 (21.8)
Age, y, Median (range)	53.5 (23-84)
Total lymph nodes	6 (0-48)
Positive lymph nodes	0 (0-35)
Tumor size	5 (0-116)
Time (disease-free survival)	21 (1-124)

[a] Neoadjuvant or adjuvant chemoradiation included conventional external beam radiotherapy using mega voltage linear accelerator photons.

Table 2. Results of the Receiver Operating Curve on the Validation Set [a]

Model	Area	Stand. Error	P Value
Cox proportional hazards	0.74	0.07	0.007
ANN	0.86	0.05	0.0003

[a] Abbreviation: ANN, artificial neural network.

Table 3. The Accuracy Rate in Prediction for Both Models on the Validation Set [a]

	Observed Number	True Prediction by ANN, No. (%)	True Prediction by Cox PH, No. (%)
Not relapsed	4	33 (60)	34 (61.8)
Relapsed	15	10 (18.2)	6 (10.9)
Total	55	43 (78.2)	40 (72.7)

[a] Abbreviations: ANN, artificial neural network; Cox PH, Cox proportional hazards.

Table 4. The Importance of Patients' Attributes in Disease Relapse Prediction According to Their Orders in the Training Set for Each Model [a]

ANN Model		Cox PH Model	
Inputs' Attributes	**Absolute Values of Final Weight**	**Significant Attributes in Final Step of** the **Model**	**Absolute Values of Coefficients**
Surgeon	0.42	Perforation or Obstruction	1.18
Cancer site	0.37	Neoadjuvant treatment	1.05
Perforation or obstruction	0.34	Perineural invasion	0.98
Age, y	0.28	Stage	0.67
Adjuvant	0.25	Grade	0.66
Stage	0.25		
Positive lymph nodes	0.24		
Neoadjuvanttreatment	0.24		
Tumor size	0.2		
Total lymph nodes	0.19		
Lymphatic-vascular invasion	0.18		
Adjuvant chemotherapy Regimen	0.18		
Grade	0.16		
Perineural invasion	0.1		
Gender	0.09		
Laboratory	0.08		
T stage	0.07		
Time	0.05		

[a] Abbreviations: ANN, artificial neural network; Cox PH, Cox proportional hazards.

In comparison, the ANN method is a powerful tool to model complex relations without any limitations for data structure. However, it requires a large data set to learn the relations and validate them. In addition, it applies all the attributes in the modeling process and cannot distinguish the confounding ones. The attempt on larger data sets is suggested for future studies to compare these methods precisely in details.

Acknowledgements

The authors are thankful to the Research Improvement Center of Shiraz University of Medical Sciences, Shiraz, Iran for their help in data gathering and Mr. J. Zomorodian for his valuable comments in data analysis.

Authors' Contributions

Saeedeh Pourahmad: design, writing and revising the manuscript, and approval of the data analysis. Bahareh Khosravi: design and data analysis. Mohammad Mohammadianpanah: data collection, writing and revising the manuscript, and approval of the final version.

Funding/Support

This study was supported by the Colorectal Research Center, Shiraz University of Medical Sciences, project number 90-5680.

References

1. Hoseini S, Moaddabshoar L, Hemati S, Mohammadianpanah M. An Overview of Clinical and Pathological Characteristics and Survival Rate of Colorectal Cancer in Iran. *Annals of Colorectal Reaerch.* 2014;**2**(1).
2. Iranian Cancer Association.. *[Official record of cancer in Iran, 1388]*.. Available from: http://en.ica.org.ir/.
3. Zheng QQ, Wang P, Hui R, Yao AM. Prognostic analysis of ovarian cancer patients using the Cox regression model. *Ai Zheng.* 2009;**28**(2):170–2.
4. Prinja S, Gupta N, Verma R. Censoring in clinical trials: review of survival analysis techniques. *Indian J Community Med.* 2010;**35**(2):217–21.
5. Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Med Res Methodol.* 2010;**10**:20.
6. ATA N, DEMĐRHAN, H. . Weighted Estimation in Cox Regression: An Application to Breast Cancer Data. *Gazi U Sc J.* 2013;**26**(1).
7. MA P, A P, M V, B MD, A S, S A, et al. Alternative for the Cox Regression model: using Parametric Models to Analyze the Survival of Cancer Patients. *Iranian Canc Prev.* 2011;**4**(1).
8. Stratford JK, Bentrem DJ, Anderson JM, Fan C, Volmar KA, Marron JS, et al. A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma. *PLoS Med.* 2010;**7**(7).
9. MT H, HB D, MH B. Neural Network Design: Pws Pub. . London: Boston; 1996.
10. Amiri Z, Mohammad K, Mahmoudi M, Parsaeian M, Zeraati H. Assessing the effect of quantitative and qualitative predictors on gastric cancer individuals survival using hierarchical artificial neural network models. *Iran Red Crescent Med J.* 2013;**15**(1):42–8.
11. Biglarian A, Hajizadeh E, Kazemnejad A, Zali M. Application of artificial neural network in predicting the survival rate of gastric cancer patients. *Iran J Public Health.* 2011;**40**(2):80–6.
12. Ada K, R. Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier. *Int App Innov E manage J.* 2013;**2**(6).
13. AH H, B B, M R, A B, E Z K. Comparison of Artificial Neural Networks and Cox Regression Models in Prediction of Kidney Transplant Survival. *Int Adv Biol Biomed Res.* 2013;**1**(10).
14. Zhu L, Luo W, Su M, Wei H, Wei J, Zhang X, et al. Comparison between artificial neural network and Cox regression model in predicting the survival rate of gastric cancer patients. *Biomed Rep.* 2013;**1**(5):757–60.
15. Ansari D, Nilsson J, Andersson R, Regner S, Tingstedt B, Andersson B. Artificial neural networks predict survival from pancreatic cancer after radical surgery. *Am J Surg.* 2013;**205**(1):1–7.
16. Ghahramani L, Moaddabshoar L, Razzaghi S, Hamedi SH, Pourahmad S, Mohammadianpanah M. Prognostic Value of Total Lymph Node Identified and Ratio of Lymph Nodes in Resected Colorectal Cancer. . *Ann Colorectal Res.* 2013;**1**(3).
17. Moghimi-Dehkordi B, Safaee A, Zali MR. Prognostic factors in 1,138 Iranian colorectal cancer patients. *Int J Colorectal Dis.* 2008;**23**(7):683–8.
18. Hai-liang C, Xing-wen Z, Qian L, Hua zZ. COX regression analysis of factors affect prognosis of colorectal cancer with time covariate. *Hainan Medic Univ J.* 2012;**8**(15).
19. Biglarian A, Bakhshi E, Gohari MR, Khodabakhshi R. Artificial neural network for prediction of distant metastasis in colorectal cancer. *Asian Pac J Cancer Prev.* 2012;**13**(3):927–30.
20. Ahmed FE. Artificial neural networks for diagnosis and survival prediction in colon cancer. *Mol Cancer.* 2005;**4**:29.
21. Kelly MM. *Hepatocellular Carcinoma:Targeted Therapy and Multidisciplinary Care.*: Springer Science & Business Media; 2010.