Prognostic classification index in Iranian colorectal cancer patients: Survival tree analysis

Amal Saki Malehi[1,2], **Fakher Rahim**[1]

Abstract

Aims: The aim of this study was to determine the prognostic index for separating homogenous subgroups in colorectal cancer (CRC) patients based on clinicopathological characteristics using survival tree analysis. **Methods:** The current study was conducted at the Research Center of Gastroenterology and Liver Disease, Shahid Beheshti Medical University in Tehran, between January 2004 and January 2009. A total of 739 patients who already have been diagnosed with CRC based on pathologic report were enrolled. The data included demographic and clinical-pathological characteristic of patients. Tree-structured survival analysis based on a recursive partitioning algorithm was implemented to evaluate prognostic factors. The probability curves were calculated according to the Kaplan–Meier method, and the hazard ratio was estimated as an interest effect size. **Result:** There were 526 males (71.2%) of these patients. The mean survival time (from diagnosis time) was 42.46± (3.4). Survival tree identified three variables as main prognostic factors and based on their four prognostic subgroups was constructed. The log-rank test showed good separation of survival curves. Patients with Stage I-IIIA and treated with surgery as the first treatment showed low risk (median = 34 months) whereas patients with stage IIIB, IV, and more than 68 years have the worse survival outcome (median = 9.5 months). **Conclusion:** Constructing the prognostic classification index via survival tree can aid the researchers to assess interaction between clinical variables and determining the cumulative effect of these variables on survival outcome.

Key words: Classification, colorectal cancer, Iran, prognostic index, survival tree

Introduction

Colorectal cancer (CRC) is classified as the third common cancer worldwide with nearly 1.4 million new cases in 2012.[1] It is also one of most malignancies cancers in Iran stands after breast cancer in females and the fourth main cancer in males (8/100,000 in male and female).[2] Estimation of diagnosed new cases of CRC in Iran was reported more than 3641 each year.[3] Recent epidemiological studies have reported the increasing incidence trend of CRC in Iran.[4,5] CRC causes to 2262 deaths annually, and it is the sixth leading cause of cancer death in Iran.[2,3] Estimated mean survival time of CRC was 105 months (confidence interval: 95.1–115.1), and the overall 5 years survival was 61.0%.[6,7] However, variation of clinicopathologic characteristics of patients leads to different survival times in several subgroups of patients that defined by different values of prognostic factors.[3] Hence, assessing prognostic factors constitutes one of the principal tasks in clinical cancer research. Evaluating prognostic factors can provide prognostic indices (PI). PI as clinical tool aid clinicians in predicting the survival outcome and prognosis of patients with aggressive diseases. The PI should be defined with good ability in grouping patients with well-separated survival distributions.

There are many prognostic evaluation methods in survival analysis, Cox proportional hazards regression model and its extensions (introduced in a seminal paper by Cox, 1972), broadly applicable and the most commonly used methods. However, Cox proportional hazards model needs to satisfy various assumptions which an underlying assumption is the proportional hazards. As well as it forces a particular link between covariates and the responses.[8] In the last two decades, tree-based models as nonparametric alternatives to parametric and semi-parametric models are developed to relax the restrictive assumptions.[9,10] Tree-based models are implemented regarding several clinicopathologic variables through recursively partitioning the covariates. Survival tree is the most popular use of the tree-based methods in survival analysis in biomedical studies.[11-13] It is an analysis that enables to determine the PI and the natural identification of prognostic groups among patients. Such grouping is important because of the patient's heterogeneity in terms of disease-free survival outcome and allows physicians to make early prudent decisions regarding adjuvant, combination therapies.[14]

The aim of this study was to characterize the PI and identify prognostic subgroups of Iranian colorectal patients to predict survival outcome and time to an event of interest.

Methods

Patients

This study was performed on patients referred to Cancer Registry Center of the Research Center of Gastroenterology and Liver Disease (RCGLD), Shahid Beheshti Medical University in Tehran, from January 2004 to January 2009. The diagnosis of CRC confirmed based on the pathology report of a cancer registry. The survival time of patients was considered from the date of diagnosis up to January 2009. These patients were treated and referred to this cancer registry of 10 public and private collaborative hospitals. After preliminary assessment, a total of 739 patients were engaged. This study was approved by the Ethic Committee of RCGLD, and all participants signed the informed consent prior to enrollment.

Deaths were confirmed through the telephonic contact to relatives of patients. Survival time as the primary outcome was calculated in months. Demographic information such as age at diagnosis, sex, race, education and marital statuses were obtained from the hospital records. The clinicopathological characteristics regarding family history of cancer, tumor grade, tumor size, pathologic stage,[15] and histopathology report were

Access this article online
Quick Response Code:
Website: www.sajc.org
DOI: 10.4103/2278-330X.179703

[1]Research Center of Thalassemia and Hemoglobinopathy, Health Research Institute, Ahvaz Jundishapur University of Medical Sciences, [2]Department of Biostatistics and Epidemiology, School of Public Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
Correspondence to: Dr. Amal Saki Malehi, E-mail: amalsaki@gmail.com

How to cite this article: Malehi AS, Rahim F. Prognostic classification index in Iranian colorectal cancer patients: Survival tree analysis. South Asian J Cancer 2016;5:23-6.

also recorded. Pathologic stage of tumor was defined based on (T) primary tumor, size, and invasiveness, (N) the extent of spread to the lymph nodes, (M) presence or absence of distant metastasis, including lymph nodes that are not regional.

Survival tree

Survival tree analysis is used to model the relationship between survival time and several potential prognostic factors nonparametrically. In this method, the patients were recursively partitioned into homogenous subgroups based on important prognostic factors. Survival tree selected predictors with the highest power to discriminate between good and bad survival as prognostic factors. The result of this analysis represented by terminal nodes which are characterized by a set of predictors and their values and is simultaneously associated with a distinct survival curve. Each terminal node defined as a class of patients with clearly separated survival curve.

Statistical analysis

Survival tree model was performed for the overall survival time, from initial diagnosis to death or censored time (end of Follow-up time). Survival probability is estimated by Kaplan–Meier method for each subgroup and represented as mean (±standard deviation). Log-rank test was used to compare the survival distributions of subgroups of PI and hazard ratio (HR) were estimated as the interested effect size. Data were analyzed using R and SPSS version 19 software (SPSS Inc., Chicago, IL, USA). $P < 0.05$ was considered as significant.

Results

A total of 739 patients were followed over the study period. The mean age at diagnosis was 59.67 ± 12.85 years (range 20–88), 526 (71.2%) were males, and 213 (28.8%) were females.

The estimated mean and median (±standard error) survival time (from diagnosis time) was 42.46± (3.4) and 22.8± (2.27), respectively, and an estimated 5 years overall survival rate was 30%. The baseline and clinical characteristics of patients and result of univariate analysis are reported in Table 1. Survival tree model was fitted based on significant variables in the univariate test.

Table 1: Baseline and clinicopathologic characteristics of the study groups

Characteristic	Categories	*n* (%)	Mean of survival time	SD	*P*
Marital status	Single	30 (4.1)	33.31	3.79	0.053
	Married	705 (95.9)	42.04	3.43	
Sex	Male	526 (71.2)	42.16	2.99	0.846
	Female	213 (28.8)	44.45	3.49	
Race	Fars	241 (33.3)	44.74	4.83	0.083
	Kordish	101 (13.9)	38.08	4.28	
	Turkish	224 (30.9)	45.67	5.18	
	Other	159 (21.9)	28.36	3.96	
BMI	<18.5	106 (22)	28.07	4.08	0.056
	18.5-25	305 (63.3)	46.99	5.31	
	25-30	60 (12.4)	45.75	5.02	
	>30	11 (2.3)	28.65	5.99	
Family history of cancer	Yes	176 (25.3)	42.09	4.40	0.199
	No	519 (74.7)	44.08	4.80	
Tobacco use	Previous or current user	252 (35.6)	32.33	3.65	0.265
	Never used	455 (64.4)	46.88	4.22	
Alcohol use	Previous or current user	45 (6.5)	52.89	8.07	0.913
	Never used	647 (93.5)	38.85	3.35	
Tumor size	<35	91 (27.3)	56.68	9.72	0.012
	≥35	242 (72.7)	49.45	6.99	
Tumor grade	Well differentiated	112 (24.1)	45.24	6.65	0.035
	Moderately differentiated	140 (30.2)	53.29	7.47	
	Poorly differentiated	200 (43.1)	28.64	2.36	
	Undifferentiated	12 (2.6)	41.67	4.89	
TNM stage 7th edition	I	68 (9.2)	25.99	2.15	<0.001
	IIA	84 (11.4)	37.34	3.95	
	IIB	38 (5.1)	28.69	3.71	
	IIC	140 (19)	31.88	2.80	
	IIIA	177 (24)	50.16	7.31	
	IIIB	68 (9.2)	36.05	5.64	
	IIIC	0 (0)	-	-	
	IV	163 (22.1)	29.74	4.23	
Histology	Adenocarcinoma NOS	527 (72.5)	40.88	3.57	0.021
	Signet ring cell carcinoma + Mucin-producing Adenocarcinoma	71 (9.8)	52.52	8	
	Other type of histology	129 (17.7)	28.54	2.91	
First treatment	Chemotherapy and radiation	112 (15.2)	19.05	1.75	<0.001
	Surgery	452 (61.2)	45.14	4.09	
	Biopsy	175 (23.6)	46.15	5.74	

SD=Standard deviation, BMI=Body mass index, TNM=Tumor, node, metastasis, NOS=Not otherwise specified

The diagram of survival tree is shown in Figure 1. It has an initial split on tumor, node, metastasis (TNM) stage as the principal prognostic factor. Survival tree identified two other variables that play important roles in survival time are age at diagnosis and the first treatment protocol. Finally, the patients were divided into homogenous subgroups based on these variables [Table 2]. Subgroup IV has a better survival outcome while subgroup II has worse survival time than other subgroups. Thus, patients with Stage IIIB-IV and more than 68 years with 9.5 months as median survival time have the lowest survival outcome [Table 3]. Estimated HRs of these subgroups showed greater risk for all subgroups than the fourth subgroup [Table 3].

The curves of cumulative hazard functions were drawn in Figure 2. According to these findings, we found that patients with Stage I-IIIA and surgery and biopsy as the first treatment (subgroup IV) has the lowest hazard rate.

The value of the overall log-rank test was 68.64 ($P < 0.001$) and revealed a significant difference between the subgroups. This means that survival tree leads to classify the patients with highly significant difference in survival outcome. In addition, Table 4 shows pairwise comparisons among the subgroups. According to these findings, subgroup II exhibited high-risk, subgroups I and III showed intermediate risk, and subgroup IV determined with low-risk.

Discussion

Beside investigation on etiology and epidemiology, identifying and evaluating the prognostic factors are one of the major tasks in clinical cancer research. In many studies, several prognostic factors and PI for survival have been identified in patients with CRC.[6,7,16-18] However, because of geographic disparities in CRC survival[19] and heterogeneity in biological and clinical pathological characteristics in patients with CRC, the survival times are different in subgroups of patients, and it is difficult to use this information to predict an individual patient's prognosis.

In this study, we evaluated prognostic factors in Iranian patients using tree-based models. The basic idea of the tree-based models to construct the subgroups based on prognostic factors that are internally as homogenous as possible with regard to their response and externally as separate as possible.[14] Recently, the tree-based model has been highlighted in predicting outcomes in cancer patients in several biomedical studies.[20-22] Survival tree analysis is utilized to homogenize the data by separating the data into different subgroups on the basis of similarity of survival outcome and determined the prognostic factors and the subgroup of patient simultaneously.[11] Evaluating the constructed prognostic subgroups via survival tree would aid the researchers to assess interaction between clinical variables, determining the cumulative effect of these variables on survival, and translating this information into appropriate management.[23] In this study, based on survival tree, TNM staging, age of diagnosis with cut of 68-year-old, and the first treatment protocols identified as prognostic factors and characterized prognostic classification index.

Based on our result, HR of patients with chemotherapy and radiation was 1.97 times than patients with surgery. There are some arguable results,[24,25] but some studies reached the same result.[6] Such reversal results may be related to molecular characteristic of the tumor.

Table 2: Subgroups for prognostic index of survival tree

Group	Node	Terminal node
I	3	Stage IIIB-IV + <68 years
II	4	Stage IIIB-IV + more than 68 years
III	6	Stage I-IIIA + chemotherapy and radiation as the first treatment were used
IV	7	Stage I-IIIA + surgery and biopsy as the first treatment were used

Table 3: Descriptive statistics and hazard ratio for each subgroup

Subgroup[a]	*n*	Mean±SE	β	Hazard ratio (95% CI)	*P*
I	167	38.17±4.75	0.549	1.73 (1.31-2.28)	<0.001
II	64	12.7±1.59	1.343	3.83 (2.71-5.41)	<0.001
III	57	21.54±2.51	0.649	1.92 (1.27-2.89)	0.002
IV	451	47.96±5.74	-	-	

[a]Subgroup 4 is reference subgroup. CI=Confidence interval, SE=Standard error

Table 4: Pairwise comparisons by log-rank test

	II	III	IV
I	χ^2=15.27 *P*<0.001	χ^2=0.072 *P*=0.789	χ^2=15.54 *P*<0.001
II		χ^2=8.92 *P*=0.003	χ^2=68.39 *P*<0.001
III			χ^2=10.82 *P*=0.001

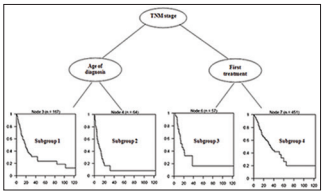

Figure 1: Survival tree. Kaplan–Meier curve inside each terminal node

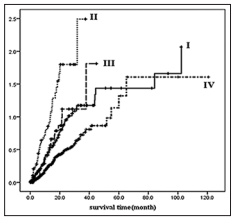

Figure 2: Cumulative hazard rate for four subgroups generated by survival tree

TNM staging was confirmed as the most prognostic factor in several studies.[6,26-28] However, there were few studies that showed inconsistent results.[29]

There were some controversy findings about age of diagnostic;[6,28] however, numerous studies were agreement with our result.[7,29] Various cut points in categorizing the age may be led to different results in the survival studies.

More investigating the result based on cumulative hazard rate curves and log-rank test showed the high-risk, intermediate and low risks subgroup of patients. The patient with Stage I-IIIA + surgery and biopsy as the first treatment was used identified as lower-risk group, and the patient with Stage IIIB-IV + more than 68 years explained as high risk.

Conclusion

Because of patient's heterogeneity in terms of overall survival outcome, using the survival tree to construct the prognostic classification index would aid the researchers to assess interaction between clinical variables and determining the cumulative effect of these variables on survival outcome.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Ferlay JS, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, *et al*. GLOBOCAN Cancer Incidence and Mortality Worldwide: IARC CancerBase. Lyon, France: International Agency for Research on Cancer; 2013.
2. Mousavi SM, Gouya MM, Ramazani R, Davanlou M, Hajsadeghi N, Seddighi Z. Cancer incidence and mortality in Iran. Ann Oncol 2009;20:556-63.
3. Malekzadeh R, Bishehsari F, Mahdavinia M, Ansari R. Epidemiology and molecular genetics of colorectal cancer in Iran: A review. Arch Iran Med 2009; 12: 161-9.
4. Kolahdoozan S, Sadjadi A, Radmard AR, Khademi H. Five common cancers in Iran. Arch Iran Med 2010; 13: 143-6.
5. Hassanzade J, Molavi E Vardanjani H, Farahmand M, Rajaiifard AR. Incidence and mortality rate of common gastrointestinal cancers in South of Iran, a population based study. Iran J Cancer Prev 2011;4: 163-9.
6. Moghimi-Dehkordi B, Safaee A, Zali MR. Prognostic factors in 1,138 Iranian colorectal cancer patients. Int J Colorectal Dis 2008;23:683-8.
7. Safaee A, Moghimi-Dehkordi B, Fatemi S, Ghiasi S, Zali M. Pathology and prognosis of colorectal cancer. Iran J Cancer Prev 2009;2: 137-41.
8. Bou-Hamad I, Larocque D, Ben-Ameur H. A review of survival trees. Stat Surv 2011;5:44-71.
9. Zhang H, Singer BH. Recursive Partitioning and Applications. 2nd ed. New York: Springer; 2010.
10. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and Regression Trees. New York: Chapman and Hall/CRC; 1984.
11. Banerjee M, Noone AM. Tree-based methods for survival data. In: Biswas A, editor. Statistical Advances in the Biomedical Sciences. New Jersey: John Wiley and Sons, Inc.; 2008. p. 265-81.
12. Ibrahim NA, Kudus A, Daud I, Abu Bakar MR. Decision tree for competing risks survival probability in breast cancer study. Int J Biol Med Sci 2008;3:25-9.
13. Al-Nachawati H, Ismail M, Almohisen A. Tree-structured analysis of survival data and its application using SAS software. J King Saud Univ (Science) 2010;22:251-5.
14. Schumacher M, Holländer N, Schwarzer G, Sauerbrei W. Prognostic factor studies. In: Crowley J, editor. Handbook of Statistics in Clinical Oncology. New York: Marcel Dekker, Inc.; 2001. p. 321-78.
15. Edge S, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. AJCC Cancer Staging Manual. 7th ed. New York: Springer; 2010.
16. Elsamany SA, Alzahrani AS, Mohamed MM, Elmorsy SA, Zekri JE, Al-Shehri AS, *et al.* Clinico-pathological patterns and survival outcome of colorectal cancer in young patients: Western Saudi Arabia experience. Asian Pac J Cancer Prev 2014; 15:5239-43.
17. Forones NM, Tanaka M, Falcão JB. CEA as a prognostic index in colorectal cancer. Sao Paulo Med J 1997; 115: 1589-92.
18. Moghimi-Dehkordi B, Safaee A. An overview of colorectal cancer survival rates and prognosis in Asia. World J Gastrointest Oncol 2012;4:71-5.
19. Henry KA, Niu X, Boscoe FP. Geographic disparities in colorectal cancer survival. Int J Health Geogr 2009;8:48.
20. Barnholtz-Sloan JS, Guan X, Zeigler-Johnson C, Meropol NJ, Rebbeck TR. Decision tree-based modeling of androgen pathway genes and prostate cancer risk. Cancer Epidemiol Biomarkers Prev 2011;20: 1146-55.
21. Shen C, Yang H, Chang Y. A decision tree-based approach for cervical smears. Int J Innov Comput Inf Control 2012;8:3251-63.
22. Malehi AS. Diagnostic classification scheme in Iranian breast cancer patients using a decision tree. Asian Pac J Cancer Prev 2014; 15:5593-6.
23. Valera VA, Walter BA, Yokoyama N, Koyama Y, Iiai T, Okamoto H, *et al.* Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. Ann Surg Oncol 2007; 14:34-40.
24. Oñate-Ocaña LF, Montesdeoca R, López-Graniel CM, Aiello-Crocifoglio V, Mondragón-Sánchez R, Cortina-Borja M, *et al.* Identification of patients with high-risk lymph node-negative colorectal cancer and potential benefit from adjuvant chemotherapy. Jpn J Clin Oncol 2004;34:323-8.
25. Coutinho AK, Rocha Lima CM. Metastatic colorectal cancer: Systemic treatment in the new millennium. Cancer Control 2003; 10:224-38.
26. Shayanfar N, Shahzadi SZ. Immunohistochemical assessment of neuroendocrine differentiation in colorectal carcinomas and its relation with age, sex and grade plus stage. Iran J Pathol 2009;4: 167-71.
27. Oh HS, Chung HJ, Kim HK, Choi JS. Differences in overall survival when colorectal cancer patients are stratified into new TNM staging strategy. Cancer Res Treat 2007;39:61-4.
28. Aghili M, Izadi S, Madani H, Mortazavi H. Clinical and pathological evaluation of patients with early and late recurrence of colorectal cancer. Asia Pac J Clin Oncol 2010;6:35-41.
29. Molaei M, Mansoori BK, Ghiasi S, Khatami F, Attarian H, Zali M. Colorectal cancer in Iran: Immunohistochemical profiles of four mismatch repair proteins. Int J Colorectal Dis 2010;25:63-9.