The Estimation of Survival Function for Colon Cancer Data in Tehran Using Non-parametric Bayesian Model

Alireza Abadi[1], Farzaneh Ahmadi[2], Hamid Alavi Majd[2], Mohammad Esmaeil Akbari[3], Zainab Abolfazli Khonbi[4], Esmat Davoudi Monfared[1]

Abstract

Background: Colon cancer is the third cause of cancer deaths. Although colon cancer survival time has increased in recent years, the mortality rate is still high. The Cox model is the most common regression model often used in medical research in survival analysis, but most of the time the effect of at least one of the independent factors changes over time, so the model cannot be used. In the current study, the survival function for colon cancer patients in Tehran is estimated using non-parametric Bayesian model.

Methods: In this survival study, 580 patients with colon cancer who were recorded in the Cancer Research Center of Shahid Beheshti University of Medical Sciences since April 2005 to November 2006 were studied and followed up for a period of 5 years. Survival function was plotted with non-parametric Bayesian model and was compared with the Kaplan–Meier curve.

Results: Of the total of 580 patients, 69.9% of patients were alive. 45.9% of patients were male and the mean age of cancer diagnosis was 65.12 (SD= 12.26) and 87.7 of the patients underwent surgery. There was a significant relationship between age at diagnosis and sex and the survival time while there was a non-significant relationship between the type of treatment and the survival time. The survival functions corresponding to the two treatment groups cross, in comparison with the patients who had no surgery in the first 30 months, showed a higher level of risk in the patients who underwent a surgery. After that, the survival probability for the patients undergoing a surgery has increased.

Conclusion: The study showed that survival rate has been higher in women and in the patients who were below 60 years at the time of diagnosis.

Keywords: Non-parametric Bayesian; Colon cancer; Survival; Iran; Tehran

1. Dept. of Community Medicine and Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. Dept. of Biostatistics, School of Paramedical, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3. Cancer Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
4. Dept. of English Language, Kashan University of Medical Sciences, Kashan, Iran

Corresponding Author:
Farzaneh Ahmadi, MS.c;
Email: ahmadi.farzan@gmail.com

Introduction

In Iran, cancer is the third leading cause of death [1] and among all types of cancer, colon cancer, after lung and stomach, is the third highest cause of death in the world [2] and the third deadliest cancer in men and the fourth one in women [3].

In Iran, the 5-year survival rate of colon cancer was approximately 60 percent [4]. Colorectal cancer survival time has increased in recent years, but the mortality rate remains high. Although studies have determined a number of factors that can predict the survival of patients after diagnosis, life expectancy has not been increased dramatically. It seems that among the prognostic factors explored so far, the most important ones are those that relate to early diagnosis of cancer. Colon cancer is more common in the elderly, although approximately 43 percent of colorectal cancer in Iran occurs before 50 years of age [5]. It is well established that colorectal cancer is one of those cancers that can largely be prevented by the early detection and removal of adenomatous polyps [6].

In survival analysis, the aim is the survival time models as a function of covariates; the most commonly used model in medical research is the Cox Semi-parametric model [7]. The regression model is

based on the assumption that the ratio of hazard functions for two different values of the covariates is proportional (PH). In cases where the PH assumption doses not satisfy, the extended Cox model can be applied. In the extended Cox model, the variable that the PH assumption does not hold for it interacts with the regression model as a function of time[8]. This function is often considered as an indicator function that often causes a jump in the hazard function [9]. If survival curves are crossed, the other semi-parametric models such as the Accelerated Failure Time (AFT) model or Proportional Odds (PO) model will also fail under such circumstances. Moreover, the application of survival analysis models which do not have such restrictions is important.

Using a Bayesian approach has some advantages over the classics, but the Bayesian methods often require complicated computations and are much less used. In this study, non-parametric Bayesian model was used to estimate the 5-year survival function of colon cancer patients in Cancer Research Center of Shahid Beheshti University of Medical Sciences in Tehran.

Materials and Methods

The sample in this survival study, based on the recorded statistics at the Cancer Research Center of Shahid Beheshti University of Medical Sciences, consisted of the patients with colon cancers diagnosed from April 2005 to November 2006 who were followed up for 5 years. The inclusion criteria were the diagnosis of colon cancer at the time and being resident in Tehran. The minimum sample size, with the proportion of 0.50, a 95% confidence interval (Z= 1.96) and the precision of 0.05 (d), was 385 alive patients with 5-year survival time. From April 2005 to November 2006, 1700 patients referred to the Cancer Research Center of Shahid Beheshti University of Medical Sciences. To collect data, telephone interview was done with the patients (if he/she was dead, the interview was done with the family and relatives). Out of 1700 contacts with patients, the 580 were successful. Out of 580 patients, 389 patients were alive and 191 patients have died during these 5 years. Information on the age of the patients at diagnosis, their sex, and their type of treatment (surgery - other treatments) was collected.

In this study, the non-parametric Bayesian model with the dependent Dirichlet process was used. It was supposed that the survival times were mixture distribution of normal densities [10, 11] and the mixture was defined by Dirichlet process [12, 13]. Conjugate prior distributions were considered with the normal and inverse gamma distributions [9, 14-16]. The choice of parameter values of priors was based on a method that was introduced by de Carvalho et al [17].

After the determination of parameters of prior distributions, to evaluate the performance of the non-parametric Bayesian model, the estimated survival function of the model was compared with the Kaplan - Meier curve. Data analysis was carried out by using R software and significant levels were considered 0.05.

Results

Out of the 580 patients, 69.9% were alive. The range of patients' age at diagnosis was 24 to 90 years with the mean of 65.1 (SD= 12.3) that was 31.1% less than 60 years. 45.9% of them were male. 87.7% of the patients had a surgery, while 12.3% did not. 46.7% of women and 53.3% of men underwent a surgery and respectively, 19.9% of the patients aged less than 60 years at diagnosis and 8.7% of patients who were older than 60 years at diagnosis did not have one. Log-rank test indicated a significant relationship between sex, age diagnosis and the type of treatment with survival time. For these variables, satisfying the PH assumption by using Schoenfeld residuals were assessed and this assumption does not hold to the type of treatment (P= 0.011).

Plots in Figure 1 show the survival function for the type of treatment using the non-parametric Bayesian model and the Kaplan-Meier method (left) and also the hazard function with a 95% credible interval (right). As shown Figure1, the survival function under the Bayesian non-parametric model is close to the Kaplan-Meier curve. Up to 30 months, the hazard is higher for the patients who had a surgery than the other group and after 30 months it is reversed. Given that the credible interval does not overlap between the two treatment groups, the hazard for the two groups differs.
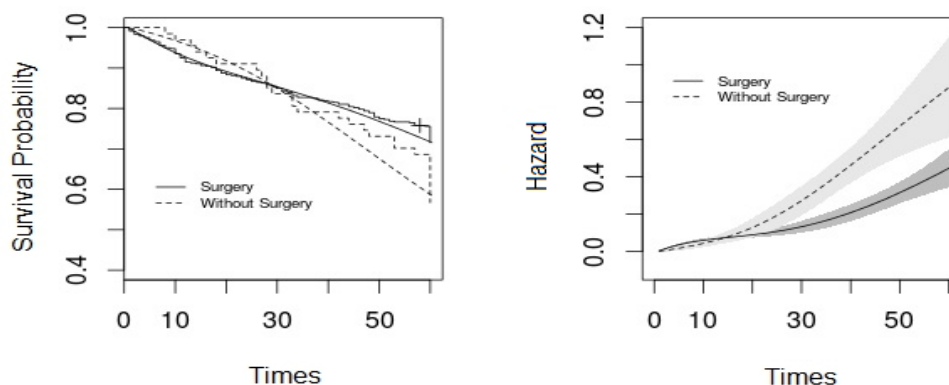
Figure 1. It shows estimated survival function under non-parametric Bayesian model and of Kaplan-Meier curve (left panel) and the estimated hazard function with 95% credible interval (right panel) for the type of treatment of colon cancer patients.

The fitted non-parametric Bayesian model was as:
$$E(Survival\ Time)$$
$$= m + \alpha_1 Type\ of\ Treatment$$
$$+ \alpha_2\ Sex + \alpha_3 Age\ at\ diagnosis$$

The results of non-parametric Bayesian are demonstrated in Table 1. The mean survival time for the patients who had a surgery was one month more than the ones not having a surgery; and for women, it was four months more than men. Regarding the patients with a cancer diagnosis age above 60, the mean survival time was five months more than the patients with cancer diagnosis age of below 60 years of age.

Table 1. Results of non-parametric Bayesian model for treatment, sex and age at diagnosis of colon cancer patients

Variable	Coefficient	S.E	95% Credible interval
Intercept	55.89	1.60	(52.88,58.99)
Type of Treatment [1]	1.01	2.24	(-5.41,3.35)
Sex [2]	-3.97	1.61	(-7.17,-0.89)
Age at diagnosis[3]	-5.04	1.46	(-7.89,-2.24)

Ref; 1: Had no surgery; 2: Male; 3: Less than 60 years old.

Figure 2 shows the estimated survival function under non-parametric Bayesian model for the two treatment groups of women and men in the patients with diagnosis ages below 60 years old. Figure 3 shows the same, but for the patients with cancer diagnosis ages of above 60. As is evident in Figure 2, in the two treatment groups, survival probability was more in women than in men. In women, after 40 months and in men, after 30 months from the study, the survival probability was more for the patients who had undertaken a surgery than those who had not, but later on, it reversed. On the other hand, the difference between survival probabilities, in the two treatment groups, was less in women than men.

In the patients with diagnosis age of above 60 years old, in the two treatment groups the survival probability in women was higher than men. In both women and men, for nearly 30 months after the study, the survival probability was higher for the patients who had a surgery than those who had not, but later on, it reversed. In addition, in the two treatment groups the difference in survival probabilities in women was less than men (Figure 3). A comparison between Figures 2 and 3 shows that survival probability for patients in the age group of below 60 years old, is higher than those in the age group of above 60.
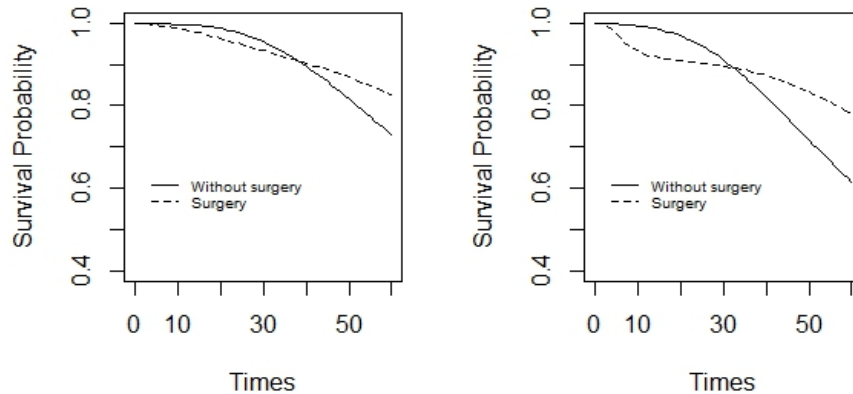
Figure 2. It shows estimated survival function under Bayesian non-parametric model in two treatment groups of men (right panel) and women (left panel) with diagnosis age of below 60 for colon cancer patients.
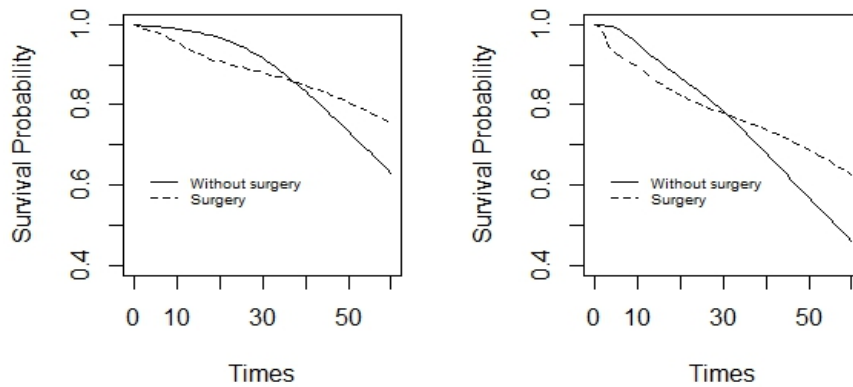

Figure 3. It shows estimated survival function under Bayesian non-parametric model in two treatment groups of men (right panel) and women (left panel) with diagnosis age of above 60 for colon cancer patients.

Discussion

In the current study, the relationship between the 5-year rsurvival of patients with colon cancer was analyzed by sex, age at diagnosis and type of treatment. There was a significant relationship between sex, age at diagnosis and the survival time, but there was non-significant relationship between type of treatment and the survival time. Furthermore, the non-parametric Bayesian model was used for estimating the survival function and it was compared with the Kaplan-Meier curve.

In this study, the age at diagnosis was significantly associated with the survival time; the lower the age at diagnosis, the higher the survival time. Mehrkhani et al., Rosenberg et al., Moradi et al., and Luhavinchi also showed that there is a significant association between age at diagnosis and the survival time [18-21] and Fang, strange, Moghimi-Dehkordi, Zhang and Wang in their studies found a non-significant relationship between the age at diagnosis and the patient survival time [22-26]. Another variable that had a significant effect on the survival of patients was the gender variable where

the survival time was found to be higher in women than men. The result of this study is in accordance with Moradi et al. and Ghazali et al. where the gender variable was significant [20, 27] but in contrast with the results of some other studies [18, 19, 21, 28]. Treatment type variable had no significant relationship with survival time and the result of this study was in contrast with Moghimi-Dehkordi et al. [24].

The estimated survival function under the non-parametric Bayesian approach for the treatment type is similar to the Kaplan-Meier curve. It indicates that the non-parametric Bayesian model suits the data for this study; hence, this model was used to estimate the survival function. Advantages of the non-parametric Bayesian survival regression model include the ease of interpretability and computational tractability. In case the PH assumption is not satisfied, the non-parametric Bayesian model with the dependent Dirichlet processis suggested [9].

Treatment type for each patient is based on various factors such as patient's age, health and disease stage. In this study, 87.8% of the patients

underwent a surgery, and only 12.2 percent had no surgery. Most cases of colon cancer undergo a surgery in addition to other treatments. Cases with no surgery can occur for three reasons: either the disease has been locally advanced, or the patient's overall health is not in a situation to tolerate a surgery, or it is due to underlying diseases such as diabetes or high age, or because the patient does not refer to and it is often due to socioeconomic factors. As it was observed in the beginning, the probability survival in patients undergoing a surgery was less than the patients who have not had surgery and it was reversed at the end of the study. For both treatment groups the probability survival in men is less than women and difference in survival probabilities in women was less than men.

Participants with high socioeconomic levels participated more in colon cancer screening programs [29] while the detection of tumors at advanced stages is more prevalent in patients at low socioeconomic level. The factors that affect patients' treatment [30] and hence the hazard of death from cancer is higher for patients at lower socioeconomic levels [31]. Due to the patients' various socioeconomic levels there are differences in hazard between men and women and also in age at diagnosis.

Acknowledgment

Conflict of Interest

The authors have no conflict of interest in this article.

Authors' Contribution

Alireza Abadi and Farzaneh Ahmadi analyzed the data. Mohammad Esmaeil Akbari and Esmat Davoudi Monfared interpreted the results. Hamid Alavi Majd and Farzaneh Ahmadi revised and edited the manuscript and Zainab Abolfazli Khonbi translated it.

References

1. Mathers CD, Boschi-Pinto C, Lopez AD, Murray CJ. Cancer incidence, mortality and survival by site for 14 regions of the world. Geneva: World Health Organization. 2001;8.

2. World Health Organisation. Cancer. [Cited 2009 April 21]. Available from: http://www.who.int/cancer/en/.

3. Sajadi A, Nouraie M, Mohagheghi M, Mousavi-Jarrahi A, Malekezadeh R, Parkin D. Cancer occurrence in Iran in 2002, an international perspective. Asian Pacific journal of cancer prevention. 2005;6(3):359.

4. Moghimi-Dehkordi B, Safaee A. An overview of colorectal cancer survival rates and prognosis in Asia. World Journal of Gastrointestinal Oncology. 2012;4(4):71.

5. Safaee A, Moghimi-Dehkordi B, Fatemi S, Ghiasi S, Nemati-Malek F, Zali M. Characteristics of colorectal mucinous adenocarcinoma in Iran. Asian Pac J Cancer Prev. 2010;11:1373-5.

6. Fatemi SR, Shivarani S, Malek FN, Vahedi M, Maserat E, Iranpour Y, et al. Colonoscopy screening results in at risk Iranian population. Asian Pac J Cancer Prev. 2010;11:1801-4.

7. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society Series B (Methodological). 1972:187-220.

8. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York: Springer-Verlag; 2000.

9. De Iorio M, Johnson WO, Müller P, Rosner GL. Bayesian nonparametric nonproportional hazards survival modeling. Biometrics. 2009;65(3):762-71.

10. Lo AY. On a class of Bayesian nonparametric estimates: I. Density estimates. The annals of statistics. 1984;12(1):351-7.

11. Tokdar ST. Posterior consistency of Dirichlet location-scale mixture of normals in density estimation and regression. Sankhyā: The Indian Journal of Statistics. 2006:90-110.

12. Ferguson TS. A Bayesian analysis of some nonparametric problems. The annals of statistics. 1973:209-30.

13. Ferguson TS. Prior distributions on spaces of probability measures. The annals of statistics. 1974:615-29.

14. De Iorio M, Müller P, Rosner GL, MacEachern SN. An ANOVA model for dependent random measures. Journal of the American Statistical Association. 2004;99(465):205-15.

15. Escobar MD. Estimating normal means with a Dirichlet process prior. Journal of the American Statistical Association. 1994;89(425):268-77.

16. West M. Hyperparameter estimation in Dirichlet process mixture models. Discussion paper: Duke University ISDS, USA; 1992.

17. de Carvalho VI, Jara A, Hanson TE, de Carvalho M. Bayesian Nonparametric ROC Regression Modeling. Bayesian Analysis. 2013;1(1):1-21.

18. Mehrkhani F, Nasiri S, Donboli K, Meysamie A, Hedayat A. Prognostic factors in survival of colorectal cancer patients after surgery. Colorectal Disease. 2009;11(2):157-61.

19. Rosenberg R, Friederichs J, Schuster T, Gertler R, Maak M, Becker K, et al. Prognosis of patients with colorectal cancer is associated with lymph node ratio: a single-center analysis of 3026 patients over a 25-year time period. Annals of surgery. 2008;248(6):968-78.

20. Moradi A, Khayamzadeh M, Guya MM, Mirzaei HR, Salmanian R, Rakhsha A, et al. Survival of colorectal cancer in Iran. Asian Pac J Cancer Prev. 2009;10:583-6.

21. Sudsawat Laohavinij M, Maneechavakajorn J. Prognostic factors for survival in colorectal cancer patients. J Med Assoc Thai. 2010;93(10):1156-66.

22. Fang H, Wang X, Feng F, Wang J. [Prognostic analysis of patients with liver metastases from colorectal cancer treated with different modes of therapy]. Zhonghua zhong liu za zhi [Chinese journal of oncology]. 2010;32(1):67-70.

23. Gharbi O, Chabchoub I, Limam S, Hochlef M, Ben Fatma L, Landolsi A, et al. [Prognostic factors and survival of metastatic colorectal cancer in the Sousse University Hospital (Tunisia): comparative study of two treatment period of 200 patients]. Bulletin du cancer. 2010;97(4):445.

24. Moghimi-Dehkordi B, Safaee A, Zali MR. Prognostic factors in 1,138 Iranian colorectal cancer patients. International journal of colorectal disease. 2008;23(7):683-8.

25. Zhang S, Gao F, Luo J, Yang J. Prognostic factors in survival of colorectal cancer patients with synchronous liver metastasis. Colorectal Disease. 2010;12(8):754-61.

26. Wang Z, Zhou Z, Liang J, Bai X, Bi J. [Prognostic factors of colorectal cancer patients with synchronous liver metastasis treated with simultaneous liver and colorectal resection]. Zhonghua zhong liu za zhi [Chinese journal of oncology]. 2008;30(5):372-5.

27. Ghazali AK, Musa KI, Naing NN, Mahmood Z. Prognostic factors in patients with colorectal cancer at Hospital Universiti Sains Malaysia. Asian Journal of Surgery. 2010;33(3):127-33.

28. Nan K-J, Qin H-X, Yang G. Prognostic factors in 165 elderly colorectal cancer patients. WORLD JOURNAL OF GASTROENTEROLOGY. 2003;9(10):2207-10.

29. Wardle J, McCaffery K, Nadel M, Atkin W. Socioeconomic differences in cancer screening participation: comparing cognitive and psychosocial explanations. Social science & medicine. 2004;59(2):249-61.

30. Groome P, Schulze K, Keller S, Mackillop W. Demographic differences between cancer survivors and those who die quickly of their disease. Clinical Oncology. 2008;20(8):647-56.

31. Vinnakota S, Lam N. Socioeconomic inequality of cancer mortality in the United States: a spatial data mining approach. International journal of health geographics. 2006;5(1):9.