Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2023;9(1):11-19

Original Article

Survival Rate Estimation in Patients with Colorectal Cancer by applying Fuzzy Product Limit Estimator

Galawezh Khedrizadeh[1], Tohid Jafari Koshki[1], Roya Dolatkhah[2], Saeid Mousavi[1*]

[1]Department of Statistics and Epidemiology, Tabriz University of Medical Sciences, Tabriz, Iran.
[2]Hematology, and Oncology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.

ARTICLE INFO

Received 09.10.2022
Revised 09.12.2022
Accepted 19.01.2023
Published 15.03.2023

Key words:
Colorectal cancer;
Survival;
Fuzzy logic;
FPLE

ABSTRACT

Introduction: The survival probability curve is used to show the progress of the disease and the effect of treatments. Estimating survival probabilities, especially in the presence of highly censored data is challenging. In this study, the Fuzzy Product Limit Estimator (FPLE) is applied to mitigate the challenge in Colorectal Cancer (CRC) survival data.

Methods: In a longitudinal study, data from 173 CRC patients were analyzed. To estimate survival probabilities, mean and median survival time, the FPLE and Kaplan – Meier (KM) methods, were applied to the data. The FPLE uses the information contained in the data and the knowledge of user to provides a smooth survival probability curve.

Results: One-year survival rate for CRC patients was estimated to be 83% using the FPLE and KM methods. The five-year survival rate was estimated to be 37% and 52% by the FPLE and KM methods, respectively. The largest observed time in data (71.96 months) was censored, so the survival rate after 71.96 months was not estimable by the KM method. But 10-year and 20-year survival rates estimates by FPLE was 0.21 and 0.09, respectively. The mean (median) survival time was estimated 45.97 (65) and 82.69 (41.70) months by KM and FPLE methods, respectively.

Conclusion: The FPLE method gives a smooth survival curve for CRC patients. The curve provides an estimate at each time point event after the largest observed time. These probabilities could be informative about the future status of the CRC patients. The smaller estimates by the FPLE at 5-year could be considered as warning that the actual survival rate is lower than that reported by the KM method.

*.Corresponding Author: musavis@tbzmed.ac.ir

Introduction

Colorectal cancer (CRC) is the third most common cancer in the world (10% of all cancers) and the second leading cause of cancer death in both sexes, according to Global Cancer Statistics in 2020. The age-specific rate of colorectal cancer was 19.5 per 100000.[1] Despite the progress in understanding the epidemiology, etiology, molecular biology, and clinical aspects of colorectal cancer over the past few decades; annually, 1.8 million new cases are diagnosed worldwide. CRC is often diagnosed in advanced clinical stages; therefore, the mortality rate is high. Keum in 2019 reported that approximately 900,000 people annually die from colorectal cancer worldwide.[2] The global burden of CRC is projected to reach 2.2 million new cases and 1.1 million deaths per year by 2030.[3]

Most CRC cases occur in developed countries. However, its incidence is constantly increasing in developing countries.[4] According to the annual reports of the Cancer Registry of Iran, CRC is the fourth most common cancer in men after gastric, bladder, and prostate cancers and the second most common cancer in women after breast cancer. It is also the third leading cause of cancer deaths in both sexes.[5]

The incidence of CRC in Iran is lower than in developed countries, but this rate has increased significantly in the last decade.[6] Annual percentage change (APC) in age-specific incidence rate (ASIR) was reported to be 13.74 for women (CI: 10.5-17.1) and 16.4 (CI: 12.4-20.5) for men.[7] The incidence of CRC varies in different geographical areas due to environmental, social, and behavioral factors. The highest incidence of CRC in Iran is observed in the central, northern, and western provinces.[8] Thus, the study of CRC in Iran is important as a public health issue.

In general, the survival rate of CRC varies worldwide. Survival rates in Eastern Mediterranean countries, especially in 5-year survival, are lower than in Europe and the United States.[9] In a meta-analysis study by Maajani et.al survival rates of one-year, three-year, and five-year CRC in Iran have been reported as 84%, 64%, and 54%, respectively.[10] In this study survival data of CRC patients in East Azerbaijan province (Located in the northwest of Iran) is studied. According to the results of population-based cancer registration in Iran, CRC is the second most common cancer in East Azerbaijan province in both sexes.[11]

Determining the survival rate of cancer patients is one of the methods to help expand health care, implement cancer programs and evaluate the effectiveness of new treatments. The most common method of estimating the survival rate is the Kaplan-Meier (KM) estimator.

The KM estimator is a non-parametric method used to estimate the survival probability at concise event time intervals.[12] The shorter the time intervals, the more accurate the KM survival probability estimation.[13] In presence of heavy censoring, the distance between two consecutive events increases, thus the probability of survival between two consecutive events can be unreliable, and the variance is estimated to be less than the actual value.[14] In addition, heavy censoring cause bias in estimates and overestimate the survival probabilities.[15,16] Besides when the largest observed time is censored the survival rate is undefined beyond that time and the continuation of survival curve is undetermined. Another method for estimating survival rate is the Fuzzy Product Limit Estimator (FPLE) of

Pokorney.[17] The FPLE is a non-parametric, data-driven fuzzy logic based estimator introduced for samples with heavy censoring. According to the simulation studies of Pokorney[17] and Musavi[18] the estimator provides more reliable estimates compared to KM in presence of heavy censoring. Also, the continuation of the survival curve is not a concern in the FPLE estimator.

In this study, the FPLE was used to handle the heavy censoring and continuation of survival curve problems in CRC survival data.

Methods

Study Design and Data Collection

All CRC cases referred to two main hospitals (Imam Reza and Shahid Ghazi) and oncology clinics in Tabriz, from January 2016 to November 2018 were included in the study. Out of 280 patients, 173 were eligible, based on CT-scan and MRI diagnosis, to enter the study. The survival time is measured as the time from diagnosis to death due to colorectal cancer or loss to follow-up from any reason. Besides, demographic, anthropometric, and pathological variables were registered for patients.

Statistical analysis

FPLE is a data-driven fuzzy logic based estimator of survival rate. In the FPLE methodology, assuming that the event does not occur immediately after the time of censoring, a number between 0 and 1 will be assigned to show the possibility of survival after the moment of censoring. To estimate survival rate, two fuzzy sets will be defined. A rectangular set will be defined for event times and a modified sigmoid function for censored times. Membership value (MV) of the fuzzy sets will show the possibility of survival. For event time, MV is one in the time interval from diagnosis to event times and 0 for the time points after the event occurred. In censored cases, the MV is one from diagnosis to the moment of censoring, after that point, a value between 0 and 1 will be defined by the sigmoid function. For times points immediately after censoring MV is close to one; as time increases, the MV approaches zero. Finally, the survival rate is estimated by aggregating the membership values at each time point. FPLE survival curve is the plotting of aggregated MVs versus the time. The area under the curve is an estimation of the mean survival time. The median survival time is the point with survival rate equal to 0.5. Membership functions are calculated as follows:

Membership value for censored time:

$$\mu_i(x) = \begin{cases} 1 & if\ 0 \le x < t_i \\ \dfrac{1}{1+\left(\dfrac{x-t_i}{TOT}\right)^{c_i} exp\left(\dfrac{x-t_i}{a_i}\right)} & if\ t_i \le x < \infty \end{cases} \quad (1)$$

TOT in Eq.1 is the total time on the study (sum of observed survival times), t_i is the observed survival time, a_i and c_i are defined using the following formulas:

$$c_i = \frac{(z_i - t_i)\ln\left(\dfrac{1-U}{U}\right) - (w_i - t_i)\ln\left(\dfrac{1-L}{L}\right)}{(z_i - t_i)\ln\left(\dfrac{w_i - t_i}{TOT}\right) + (w_i - t_i)\ln\left(\dfrac{TOT}{z_i - t_i}\right)} \quad (2)$$

$$a_i = \frac{z_i - t_i}{\ln\left[\left(\dfrac{1-L}{L}\right)\left(\dfrac{z_i - t_i}{TOT}\right)^{-c_i}\right]} \quad (3).$$

In Eq.2, w_i is a time point with the possibility of

survival equal to *U* and is obtained as follows:

$$w_i = t_i + \left(\frac{cen+n}{2n} \right) \frac{1}{n} \sum_{j=i}^{n} t_i \qquad (4)$$

In Eq.4 "*cen*" is the number of censored times in the data set. Zi in Eq2-3 is a point on the time axis with the possibility of survival, equal to L:

$$z_i = \mathrm{w_i} + \frac{\mathrm{TOT}}{\mathrm{f_i} + 1} \qquad (5)$$

f_i in Eq.5 is the number of events that occurred after t_i. If no events occurred after t_i there is more optimism in the possibility of survival compared to those t_i's that there are many events after that time. *L* and *U* in Eq.2 and Eq.3 are specified by the user to determine the shape of the membership function. *U* is a value between 0.5 and 1 and *L* is a positive value indicating the degree of membership at the point z in Eq.2. With a fixed value of *U*, as *L* decreases the degree of membership approaches zero faster. While *L* is fixed and *U* is closer to one the degree of membership approaches zero very slowly and there is more optimism in survival after censoring.

In brief, the survival membership function for patients with censored and event time is:

$$g_i(\mathrm{x}) = \begin{cases} \grave{\imath}_{\,i}(\mathrm{x}) & ,\textit{if } \mathrm{t_i}\ \textit{is censored} \\ \begin{cases} 1 & , \ x < \mathrm{t_i} \\ 0 & , \ x \geq \mathrm{t_i} \end{cases} & ,\textit{if } \mathrm{t_i}\ \textit{is event} \end{cases} \qquad (6).$$

After constructing the membership functions for all t_i's, the survival probability estimate is defined by Eq.7. The resulting estimator is called the Fuzzy product limit estimator (FPLE).[17]

$$\tilde{s}(x) = \frac{\sum_{i=1}^{n} g_i(x)}{n} \qquad (7)$$

The area under the curve (AUC) by Eq.7 is defined as the mean survival time like KM curve. Notice that if all t_i's are event times, the FPLE and KM estimates are equivalent.

In advance of applying FPLE on CRC survival data, U and L would be selected by literature review on CRC survival rate and the KM curve for CRC patients. Mean, Median, one-year, three-year, five-year, 10-year and 20-year survival rate will be calculated by the FPLE and KM methods.

Result

From a total of 173 cases, 97 (56.1%) of patients were male, and 76 (43.9) were female. The mean age of CRC patients at diagnosis time was 59.05±12.93 years, and the median age was 59 years. At the end of follow-up, 50 cases have died due to CRC, and 123 (71.1%) were censored.

The KM estimator was applied to the CRC patient's survival time data. Mean and median survival times was estimated to be 45.97 and 65 months, respectively. KM curve is plotted in Figure 1. Considering the KM curve and previous studies on survival rate of CRC patients, U and L in the FPLE method were chosen to be 0.85 and 0.0005, respectively. Then, the survival rate estimates calculated by the FPLE method. The FPLE survival curve is shown in Figure 1 along with KM curve. The curve is included in 95% CI of KM estimates. Mean and median survival time estimated 82.69 and 41.7 months by the FPLE, respectively. The result is reported in Table 1 with 95% CI.

Mean and median survival times for male/female

and age groups (less than 50 and greater than 50 years-old) with 95% CI interval is reported in Table 1. In FPLE method CI is calculated by bootstrap method with 500 replications. 2.5 and 97.5 percentile of bootstrap distribution was reported as 95% CI. In both methods mean survival time is larger for males and age group under 50 years old.

One-year, three-year, five-year, 10-year, and 20-year survival rates with the FPLE and KM estimators are reported in Table 2. The KM estimator was unable to estimate 10 and 20-year survival rates. The largest observed time, 71.96 months, was censored and the continuation of the KM curve was undetermined beyond that time. But, the FPLE method estimated 10 and 20- year survival rates, as 0.21 and 0.09. Survival curve by the FPLE estimator is plotted in Figure 2 with 95% confidence interval estimated by Bootstrapping.

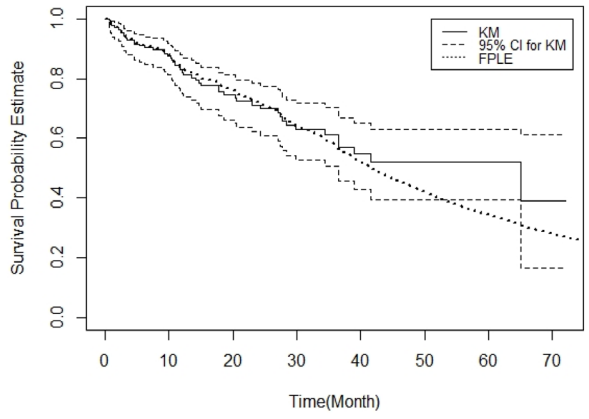

Figure 1. Survival probability estimates for CRC patients by KM (solid line) and FPLE (dotted line) estimators

Table 1. Mean and median survival time estimates with 95% CI for CRC patients by FPLE and KM methods

			Mean		Median	
Method	Variable	Category	Estimate	95% CI	Estimate	95% CI
KM	Gender	Male	46.31	39.71-52.92	65	22.10-107.89
		Female	43.48	34.97-52	39.03	28.77-49.28
	Age	<50	55.70	45.58-65.81	-	-
		50<=	41.59	35.65-47.54	41.63	27.33-55.92
	Overall		45.97	40.38-51.57	65	27.67-102.37
FPLE	Gender	Male	77.03	55.67 -103.98*	45.50	35.89-65.00*
		Female	60.62	43.43-77.73*	36.50	31.09-48.82*
	Age	<50	84.99	52.50 – 103.89*	55	39.10-79.52*
		50<=	66.14	50.52-91.20*	37.10	31.39-50.80*
	Overall		82.69	61.29-103.58*	41.70	36.59-53.91*

*Estimated by 500 bootstrap replicates

Table 2. CRC patient's survival rate estimates by the FPLE and KM methods

Time (Month)	FPLE	KM
12	0.83	0.83
36	0.58	0.57
60	0.37	0.52
120	0.21	-
240	0.09	-

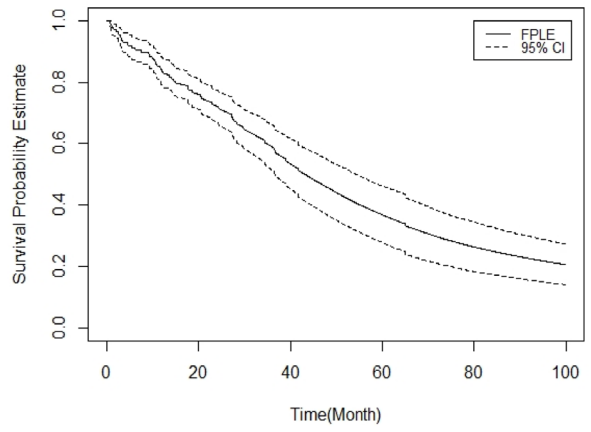

Figure 2. Survival probability estimates for CRC patients by the FPLE estimators with 95% CI

Discussion

This study aimed to provide a smooth survival probability for the CRC patients in presence of highly censored data. The FPLE estimator was used for this purpose. The difference between the KM and FPLE is in the way of estimation. The KM estimates survival probability in event times. In intervals between event times, the survival probability is equal to the previous event times survival probability. In the FPLE method survival probability is estimable at each time point. Therefore, the estimator provides a smooth survival probability curve.

In studies that the largest observation is event time, the KM curve drops to zero, meaning that survival probability beyond that time is zero. In case when it is censored time, the continuation of the KM curve is undetermined.[19] In the FPLE method membership functions provide MV for each time point, so the survival curve continues until it approaches zero. Thus, the continuation of the survival curve after largest observation is not a concern in the FPLE method.

Mean survival time in the FPLE method is expected to be greater than the KM method, because of the continuation of the curve. The result of this study showed that mean survival time estimated by the FPLE was larger than the estimates by the KM. Also, median survival time is always estimable in the FPLE method unlike the KM method. In the FPLE method, median survival time is always smaller than mean survival time as expected in right skewed data. In the KM method sometimes the median survival time is greater than the mean survival time. In this study median survival time was greater than mean survival time. Also, the median survival time was not estimable for the age group under 50 years old.

Comparing the methods in aspect of computation time, the FPLE needs more computation time than the KM method. Estimating the variances of estimates and confidence interval is not direct in the FPLE. The CI is estimated by bootstrap method.

In presence of high censoring, the KM estimator can be very unstable in times close to the end of the follow-up.[20] According to the study of Ming Zhong et al. mean survival time is overestimated as the censorship rate increases. The bias amount varies for different distributions, so that as the skewness increases the bias increases. Also, with a censoring rate higher than 30% the bias amount gets higher.[21] The study of Musavi et al. and Pokorny et al. also showed that the estimates of mean survival time by the FPLE in presence of highly censored data are more reliable than KM.[17,18] This is seen also in the confidence interval of the KM curve. The CI at the larger times is wider than FPLE's CI.

In this study, one and three-year survival rates by the KM and the FPLE are very close and similar to the result of a study by Majani.[10] In the study of Saadati et al. five-year survival rate

was reported to be in a range of 15-93%[22] which is consistent with the result of the current study. The five-year survival rate based on FPLE and KM were 37 and 52 percent, respectively. Looking at the survival curves by the two method, as the time intervals between event times are small the curves are very close. But at the end of the study time when the event time intervals are larger, the FPLE estimates are smaller than the KM's. In the FPLE method, at each time point the possibility of survival is estimable unlike the KM. Therefore, while in the KM curve the survival probability stands constant in the interval between two consecutive event times, the estimates by the FPLE smoothly goes toward zero. This causes difference in estimates by the two methods at five years. If we assume that a proportion of cases will never experience the event, then the cure model occurs and the FPLE method is not appropriate. Otherwise the FPLE method could help to have survival estimates beyond the largest observed survival time.
The limitation of this study was that follow-up time in this study was less than six years, therefore, comparison of survival rate by the two methods at 10 and 20 years was not possible.

Conclusion

In presence of highly censored survival data, the FPLE method provides acceptable estimates of CRC patients survival rate. Also, the continuation of survival curve was estimated after largest observed time. These probabilities could be informative about the future status of the CRC patients. The smaller estimates by the FPLE at 5-year could be considered as warning that the actual survival rate is lower than that of reported by the KM method. Thus, in treatment of patients with colorectal cancer this finding should be considered.

Conflict of interest

The authors declare no conflict of interest.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–249.

2. Keum NN, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol. 2019;16(12):713–32.

3. Kalyan A, Kircher S, Shah H, Mulcahy M, Benson A. Updates on immunotherapy for colorectal cancer. J Gastrointest Oncol. 2018;9(1):160–9.

4. Cancer IA for R on. Globocan 2018: Cancer Fact Sheets—Colorectal Cancer. IARC http//gco iarc fr/today/data/factsheets/cancers/10_8_9-Colorectum-fact-sheet pdf. 2018;

5. Zendehdel K. Cancer Statis. Basic Clin Cancer Res. 2019;11(1):1–4.

6. Dolatkhah R, Dastgiri S, Somi M hossein, Jabbarpour Bonyadi M, Kermani IA, Farassati F. Colorectal cancer in Iran: Molecular epidemiology and screening strategies. J Clin

Oncol. 2019;33(15_suppl):e12513–e12513.

7. Rafiemanesh H, Pakzad R, Abedi M, Kor Y, Moludi J, Towhidi F, et al. Original article : Colorectal cancer in Iran : Epidemiology and morphology trends. EXCLI J. 2016;15:738–44.

8. Khosravi Shadmani F, Ayubi E, Khazaei S, Sani M, Mansouri Hanis S, Khazaei S, et al. Geographic distribution of the incidence of colorectal cancer in Iran: a population-based study. Epidemiol Health. 2017;39:e2017020.

9. Nikbakht HA, Hassanipour S, Shojaie L, Vali M, Ghaffari-fam S, Ghelichi-ghojogh M, et al. Survival Rate of Colorectal Cancer in Eastern Mediterranean Region Countries: A Systematic Review and Meta-Analysis. Cancer Control. 2020;27(1):1–15.

10. Maajani K, Khodadost M, Fattahi A, Shahrestanaki E, Pirouzi A, Khalili F, et al. Survival rate of colorectal cancer in Iran: A systematic review and meta-analysis. Asian Pacific J Cancer Prev. 2019;20(1):13–21.

11. Somi MH, Dolatkhah R, Sepahi S, Belalzadeh M, Sharbafi J, Abdollahi L, et al. Cancer incidence in the East Azerbaijan province of Iran in 2015-2016: Results of a population-based cancer registry. BMC Public Health. 2018;18(1):1–13.

12. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. J Am Stat Assoc. 1958;53(282):457.

13. Kishore J, Goel M, Khanna P. Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res. 2010;1(4):274.

14. Does your data violate Kaplan-Meier assumptions?

15. Breslow NE. Introduction to Kaplan and Meier (1958) Nonparametric estimation from incomplete observations. In: Breakthroughs in Statistics. Springer; 1992. p. 311–8.

16. Shafiq M, Shah S, Alamgir M. Modified Weighted Kaplan-Meier Estimator. Pakistan J Stat Oper Res. 2007;3(1):39.

17. Pokorny K, Sule D. Empirical fuzzy estimate of the survival curve. Int J Uncertainty, Fuzziness Knowlege-Based Syst. 2004;12(3):347–56.

18. Musavi S, Pokorny KL, Poorolajal J, Mahjub H. Fuzzy survival analysis of AIDS patients under ten years old in Hamadan-Iran. J Intell Fuzzy Syst. 2015;28(3):1385–92.

19. Laux SE. Techniques for. IEEE Transactions on Electron Devices. 1985. 2028–2037 p.

20. Shen Y, Fleming TR. Weighted mean survival test statistics: A class of distance tests for censored survival data. J R Stat Soc Ser B Stat Methodol. 1997;59(1):269–80.

21. Zhong M, Hess KR. Collection of Biostatistics Research Archive COBRA Preprint Series Mean Survival Time from Right Censored Data Mean Survival Time from Right Censored Data. 2009;99.

22. Mozafar Saadati H, Khodamoradi F,

Salehiniya H. Associated Factors of Survival Rate and Screening for Colorectal Cancer in Iran: a Systematic Review. J Gastrointest Cancer. 2020;51(2):401–11.