ISSN 2250-3153

Non Parametric Method of Estimation of Survival and Hazard Functions for Patients with Uterine Fibroid

Dr. J. Jothikumar [*], V. Sumathy [**]

[*] Associate Dean, Faculty of Science and Humanities, SRM University, Kattangulathur, jjothikumar@yahoo.co.in
[**] Assistant Professor in Statistics, Kumararani Meena Muthiah College of Arts and Science, Adyar, Chennai, v.sumathy@yahoo.com

Abstract-The present study aims at the analysis of data obtained from patients with Uterine Fibroid. The studies consist of data obtained from 105 patients in respect of clinical and follow up actions for the treatment of Fibroid. Non-parametric methods viz., Kaplan-Meier, Product-limit method, Log Rank Test and Cox Regression models have been used for analyzing the Fibroid data. The analysis revealed the significant role played by certain variables in the survival of patients with Fibroids.

Index Terms- Distribution-Free methods, Cox regression, Survival function Log Rank Test, Uterine Fibroid.

I. Introduction

In survival analysis we fit statistical distributions for empirical data and assess the fitness of the theoretical distribution form. There are some situations in which the experimenter has no knowledge of the form of the theoretical distribution for which the data has to be fitted and the validity has to be assessed. In such situations the experimenter resort to apply non-parametric or distribution free methods for analyzing the survival data. In the present study we have collected data from a sample of 105 patients with Uterine Fibroid disease. An attempt is made to arrange the data in a retrospective way. Using non-parametric methods we have computed estimates based on survival function. A comparison among the groups of patients is also carried out using Log Rank Test. Based on Cox regression model we have carried out multivariate analysis and detected the independent risk factors among the patients.

A. Non parametric approach in life time data analysis

Kaplan and Meier [10] have made significant contributions for the study of survival analysis and in its related areas. They have identified Product-limit (PL) method for the estimating survival function with graphical representation. In the subsequent decades attempts were made for estimating survival function through different methods and the actuarial life method has become popular Berkson and Gage [2], Cutler and Ederer [8] have developed life tables and estimated the survival function. Gehan [9] in his celebrated paper has given procedures for the estimation of the basic functions viz., survival function, hazard function and density function. Breslow and Crowley [3] and Meier [12] have shown that under certain conditions the Kaplan –Meier estimate is consistent and asymptomatically normal. Recently Zou et al.,[4] have applied non-parametric maximum likelihood approach to study the multiple change point problems.

B. Comparison of two groups of survival data

Cox [6] made initial attempts for studying the comparison of survival data between two groups of patients. Mantel [11] has generalized the Savage Test which is termed as Log Rank Test. Subsequently contributions for the study of log rank test were made by Cox [7], Peto and Peto [13] and Peto et al., [14]. The details of the log rank test are discussed in the subsequent sections.

C. Objective of the study

The objectives of the study are;

(i) Estimation of the proportion of population which will survive past a certain time.
(ii) Computation of the survival rate and the recurrence rate
(iii) To take into account the multiple causes of death and its implications or not.
(iv) To identify the circumstances this causes either increase or decrease in the odds of survival.

Basics of survival analysis

Let the random variable T denotes the survival time. The distribution function of the survival time is given by

$$F(t) = P(T < t) = \int_0^t f(u).du, \tag{1.1}$$

The density function is viz.,

$$f(t) = \frac{dF(t)}{du}, \tag{1.2}$$

We can define the survival function as,

$S(t) = P(T \geq t) = 1 - F(t).$ (1.3)

The hazard function h(t) of the survival time T gives the conditional failure rate and the same is given by

$h(t) = \lim_{\delta t \to 0}\{\frac{P(t \leq T < t + \delta t | T \geq t}{\delta t}\} = \frac{f(t)}{S(t)}$ (1.4)

Then it follows that

$h(t) = -\left[\frac{dS(t)/dt}{S(t)}\right]$ (1.5)

and

$S(t) = exp\left[-\int_0^t h(u)du\right]$ (1.6)

II. NON PARAMETRIC APPROACH FOR CENSORED SURVIVAL DATA

A. Kaplan-Meier estimation

The Kaplan- Meier (KM) Survival Analysis procedure is a method of estimating time-to-event models in the presence of censored cases. The Kaplan-Meier model is based on estimating conditional probabilities at each time point when an event occurs and taking the product limit (any KM formula for a survival probability is limited to product terms up to the survival week being specified. Then the KM formula is often referred to as a "product-limit" formula) of those probabilities to estimate the survival rate at each point in time.

To compute KM curves, we must form a data layout that orders the failure times from smallest to largest. For each ordered failure time, the estimated survival probability is computed using the **product limit formula**.

$t(j)$: j th ordered failure time

$\hat{S}\,(t(j)) = \prod_{i=1}^{j} \hat{P}\mathrm{r}(T > t(i) | T \geq t(i))$

$= \hat{S}\,(t(j-1)) \times \hat{P}\mathrm{r}(T > t(i) | T \geq t(i))$ (2.1)

Alternatively, this estimate can be computed as the product of the survival estimate for the previous failure time multiplied by the conditional probability of surviving past the current failure time.

Let $t_{(1)} < t_{(2)} < \cdots < t_{(n)}$ denote the distinct ordered times of death. Let d_i be the number of deaths at $t_{(i)}$, and let n_i be the number alive just before$t_{(i)}$. This is the number exposed to risk at time$t_{(i)}$. Then the KaplanMeier or product limit estimate of the survivor function is

$\hat{S}(t) = \prod_{i:t_{(i)<t}}(1 - \frac{d_i}{n_i})$ (2.2)

B. Cox regression model

The hazard function for the i^{th} individual can be written as:

$h_i(t) = (e^{\beta' x_i})h_0(t),$ (2.3)

Suppose the data consists of n observed survival times, denoted by $t_1, t_2, \ldots, t_n$ and δ_i is an indicator variable

$\delta_i = \begin{cases} 1, & if uncensored \\ 0, & if censored \end{cases}$

Collett[5] derived the likelihood function for the Cox Proportional Hazard model is given by

$L(\beta) = \prod_{i=1}^{n}\left[\frac{e^{\beta' X_{(i)}}}{\sum_{l \epsilon R(t_{(i)})} e^{\beta' X_{(l)}}}\right],$ (2.4)

where $R(t_i)$ is the risk set at time t_i. The log likelihood function is given by:

$logL(\beta) = \prod_{i=1}^{n} \delta_i \left\{\beta' X_i - log \sum_{l \epsilon R(t_{(i)})} e^{\beta' X_{(l)}}\right\}$ (2.5)

C. Comparison of two survival distributions

The problem of comparing survival distributions arises often in biomedical research. Invariably, the disease free or survival times of the different groups vary. These differences can be illustrated by drawing graphs of the estimated survivorship functions, but that gives only a rough idea of the difference between the distributions. It does not reveal whether the differences are significant or merely chance variations. A statistical test is necessary. To evaluate whether or not KM curves [1] for two or more groups are statistically equivalent, we use the most popular testing method viz., Log Rank Test.

Logrank test

The Log Rank Test statistic *S* is equal to the sum of the failures observed minus the conditional failures expected computed at each failure time, or simply the difference between the observed and expected failures in one of the groups. The Log Rank Test is a Chi-square test. It Provides an overall comparison of KM curves. It uses observed versus expected cell outcomes over categories of outcomes. The categories for the log rank statistic are defined by each of the ordered failure times for the entire set of data being analyzed.

The Log Rank Test statistics compares estimates of the hazard functions of two groups at each observed event time based on the sum S of the 'w' scores of the two groups. The permutational variance of S is given by,

$$Var(S) = \frac{n_1 n_2 \sum_{i=1}^{n_1+n_2} w_i^2}{(n_1+n_2(n_1+n_2-1)} \quad (2.6)$$

The test statistics $L = \frac{S}{\sqrt{Var(S)}}$ has an asymptotically standard normal distribution.

III. Data Description and Data Base

3.0 Meaning of Fibroid

Fibroids are muscular tumors that grow in the wall of the uterus (womb). In medical terminology Fibroids is also called "leiomyoma" (leye-oh-meye-OH-muh) or just "myoma". Fibroids are almost always benign (not cancerous). Fibroids can grow as a single tumor, or there can be many of them in the uterus. They can be as small as an apple seed or as big as a grapefruit. In unusual cases they can become very large.There are factors that can increase a woman's risk of developing Fibroids.

The general features of Fibroid include the following:

(i) **Age**: Fibroids become more common as women age, especially during the 30s and 40s through menopause. After menopause, fibroids usually shrink.

(ii) **Family history**: Prior information on family history will reveal the severity of Fibroids.

(iii) **Ethnic origin:** It is stated that African-American women are more likely to develop fibroids than white women.

(iv) **Obesity:** Women who are overweight are at higher risk for fibroids. It may be noted that for very heavy women, the risk is two to three times greater than average.

(v) **Eating habits**: This plays a greater role with higher risk of Fibroids. Most fibroids grow in the wall of the uterus.

Medical practitioners classify the Fibroids as given below:

(i) **Submucosal** (sub-myoo-KOH-zuhl)**-** Fibroids grow into the uterine cavity.

(ii) **Intramural** (ihn-truh-MYOOR-uhl) **-**Fibroids grow within the wall of the uterus.

(iii) **Subserosal** (sub-suh-ROH-zuhl) **-**Fibroids grow on the outside of the uterus.

Some fibroids grow on stalks that grow out from the surface of the uterus or into the cavity of the uterus. They might look like mushrooms. These are called **pedunculated** fibroids.

Most Fibroids do not cause any symptoms. The consequences of Fibroids with some women are noted below:

- Heavy bleeding (which can be heavy enough to cause anemia) or painful periods
- Feeling of fullness in the pelvic area (lower stomach area)
- Enlargement of the lower abdomen
- Frequent urination
- Pain during sex
- Lower back pain
- Complications during pregnancy and labor, including a six-time greater risk of cesarean section
- Reproductive problems, such as infertility, which is very rare

Medical researchers are trying to find the probable causes for the Fibroids. They have identified two main factors viz., Hormonal (affected by estrogen and progesterone levels) and Genetic (runs in families) . Fibroids grow rapidly during pregnancy, when hormone levels are high. They shrink when anti-hormone medication is used. They also stop growing or shrink once a woman reaches menopause

A.Data Base for the study and Data description.

The reference period of the study is 2007-2012. The data relating to fibroids were collected from a sample of 105 women who have undergone treatment for uterine fibroid in a private hospital, Chennai. About 50% of the patients had fibroid in intramural position,30% in submucal position and 20% in subserosal position. As far as the severity of the symptoms about 26% of the patients suffered from severe pain in abdomen during periods,51% complaints with heavy bleeding and they seem to be aneamic,1% with inconception,1% bleeding after menopause,1% misconception etc.,

The variables collected under study are Age, Blood group, Age of menarche(age at puberty), Age at marriage, Family income, Place of residence, Position of Fibroid in uterus, Fibroid size, Symptoms persisted due to Fibroid, Number of children of the patient,

The time reference considered in the study are :
Time of diagnosis: It is the month and the year in which the abnormality is diagnosed.
Time of treatment: After diagnosis it is the time when treatment is given.
Waiting time: It is duration in months between the time when fibroid is diagnosed to the time when treatment is taken.
The basic statistics with reference to the variables are given in the following Table 1.

Table 1. Basic Statistics of covariates

Variables		Mean $\bar{x}$	Standard deviation S
Age(years)	$[x_1]$	38	11.63
Age at menarche(years)	$[x_2]$	13	1.7
Age at marriage(years)	$[x_3]$	26	0.5
Number of children	$[x_4]$	1	1.5
Waiting time(months)	$[x_5]$	21	19.22

It is observed from above Table 1, there is greater variation in respect of the variables viz., Age and Waiting Time.

B. Empirical Analysis

Computation of estimate of survival function and hazard function

The survival time forms the basis for the computation of the functions viz., Survival function and Hazard function. The survival times observed among the women with Fibroids are not uniform. The estimates of the survival function, hazard function and survival standard error are given in Table 2.

We observe the following based on the results given Table 2:

(i) Estimates of the survival function tends to zero as the survival time increases in a non-uniform manner

(ii) Estimates of the hazard function tend to one as the survival time increases in non-uniform manner.

The following figures [Fig.1- Fig.6] illustrate the survival plots under different factors.
The steps clearly reveals the Waiting time of patients (in months) in Figure 1. In Figure 2, we observe that the hazard function exhibits an approximate S curve. Figure 3, the shape of survival function reveals the same as that of waiting time curve. Figure 4, gives the estimated survivor curves for the two age of menarche of patients in two groups less than or equal to 13 and more than 13.

Table 2. Estimated Survival and hazard function of patients with Uterine fibroid

SurvivalTime	**Survival function** $\hat{S}(t)$	**Hazard function** $\hat{h}(t)$	**Survival Standard Error**
0	1.0000	0	0
1	0.9810	0.0190	0.0133
2	0.8190	0.1810	0.0376
3	0.7333	0.2667	0.0432
5	0.7048	0.2952	0.0445
6	0.6571	0.3429	0.0463
10	0.6476	0.3524	0.0466
11	0.6286	0.3714	0.0472
12	0.5714	0.4286	0.0483
13	0.5429	0.4571	0.0486
14	0.5328	0.4672	0.0487
16	0.5115	0.4885	0.0491
17	0.5008	0.4992	0.0492
18	0.4582	0.5418	0.0494
21	0.4262	0.5738	0.0493

24	0.3917	0.6083	0.0492
27	0.3341	0.6659	0.0482
28	0.3110	0.6890	0.0476
30	0.2995	0.7005	0.0472
33	0.2875	0.7125	0.0468
35	0.2750	0.7250	0.0464
41	0.2475	0.7525	0.0456
42	0.2338	0.7662	0.0451
51	0.2192	0.7808	0.0446
58	0.1948	0.8052	0.0458
59	0.1623	0.8377	0.0483
72	0	1.0000	.

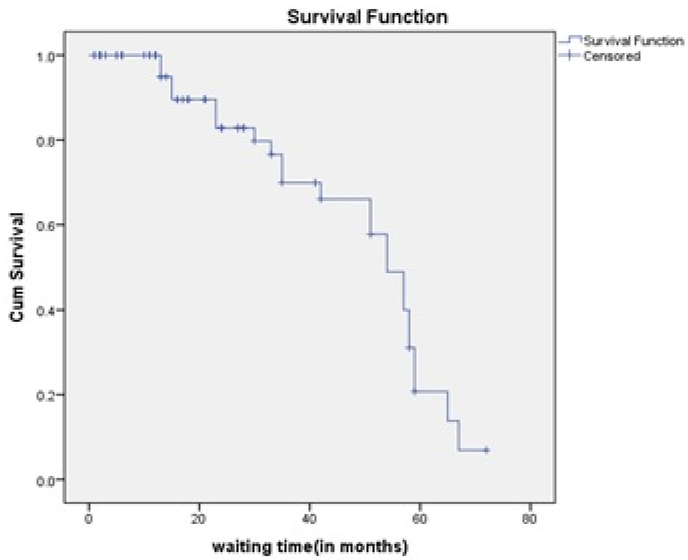

Figure 1. Survival curve of 105 patients with uterine fibroid

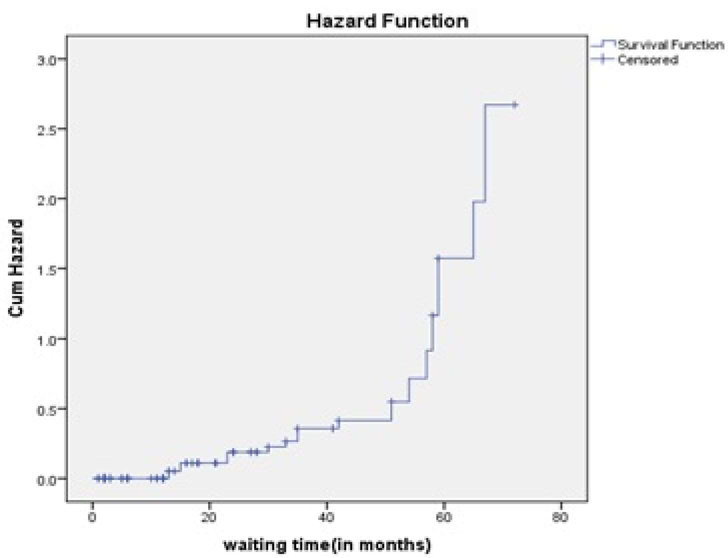

Figure 2. The hazard curve of 105 patients with Uterine fibroid

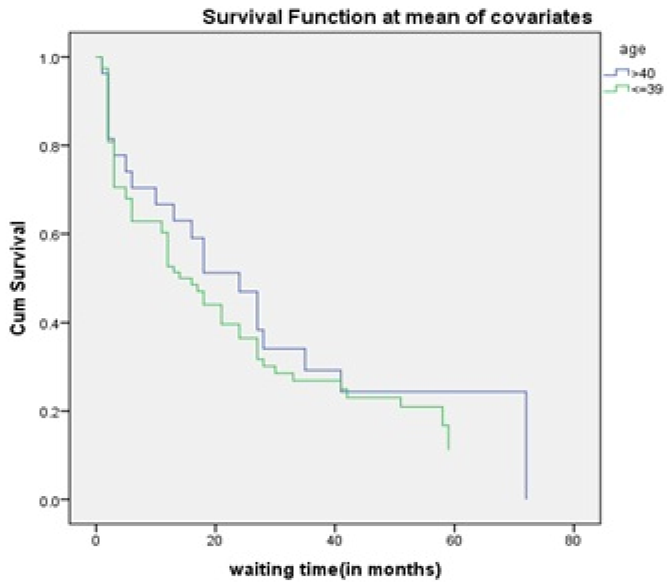

Figure 3. Overall survival curves of patients with uterine fibroid in different age group

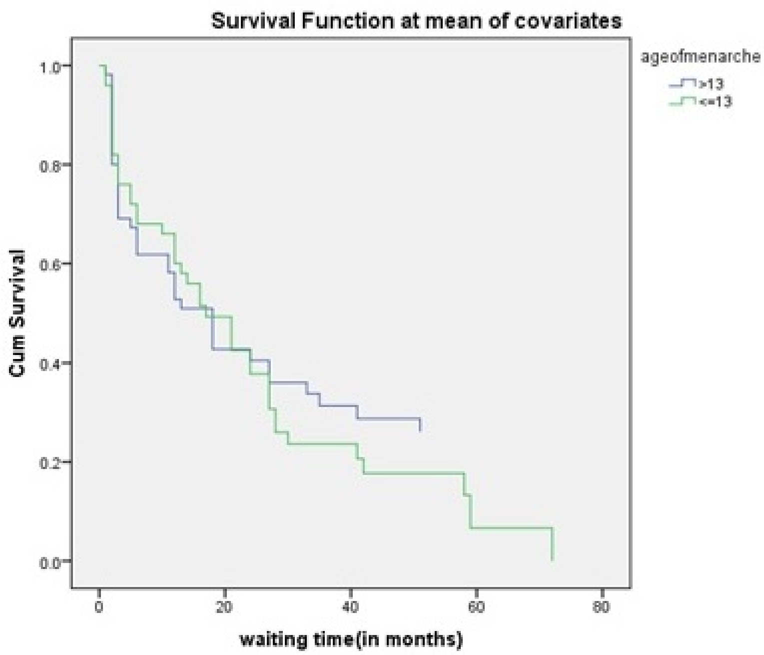

Figure 4. Overall survival curves of patients with Uterine fibroid in different age group of menarche

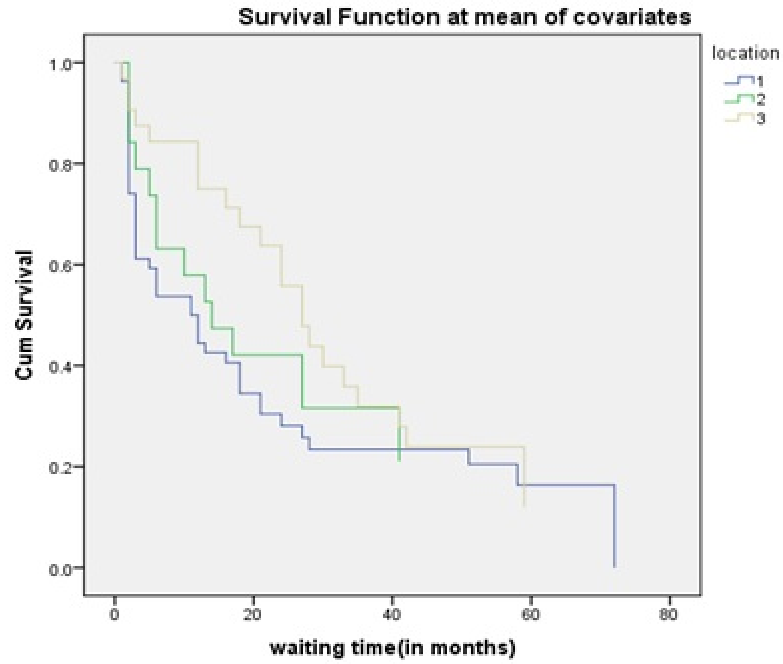

Figure 5. Overall Survival curves of patients with Uterine fibroid with fibroid in different locations.

1-intramural,2-subserosal,3-submucal.

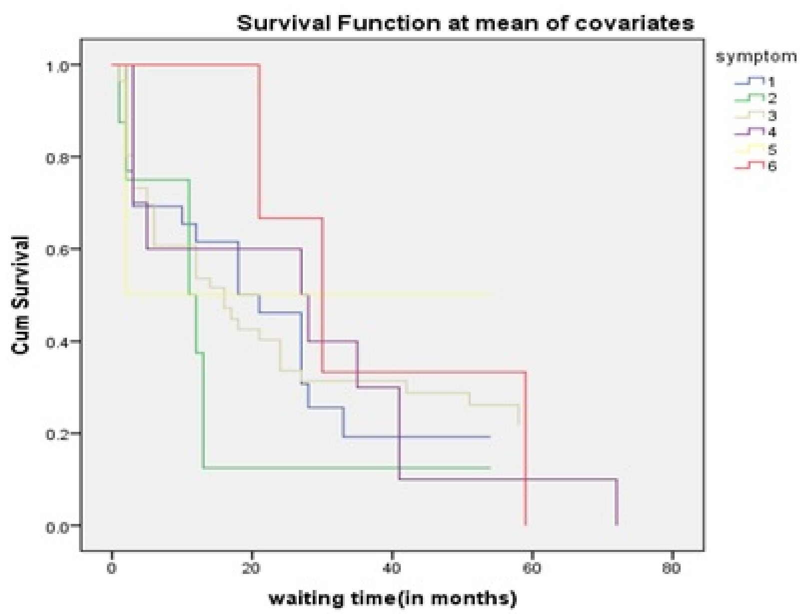

Figure 6 Overall survival curves of patients with Uterine fibroid with different symptoms.

1-pain in abdomen,2-inconception,3-heavy bleeding,anemic,4-Bleeding after menopause,5-irregular periods,6-misconception

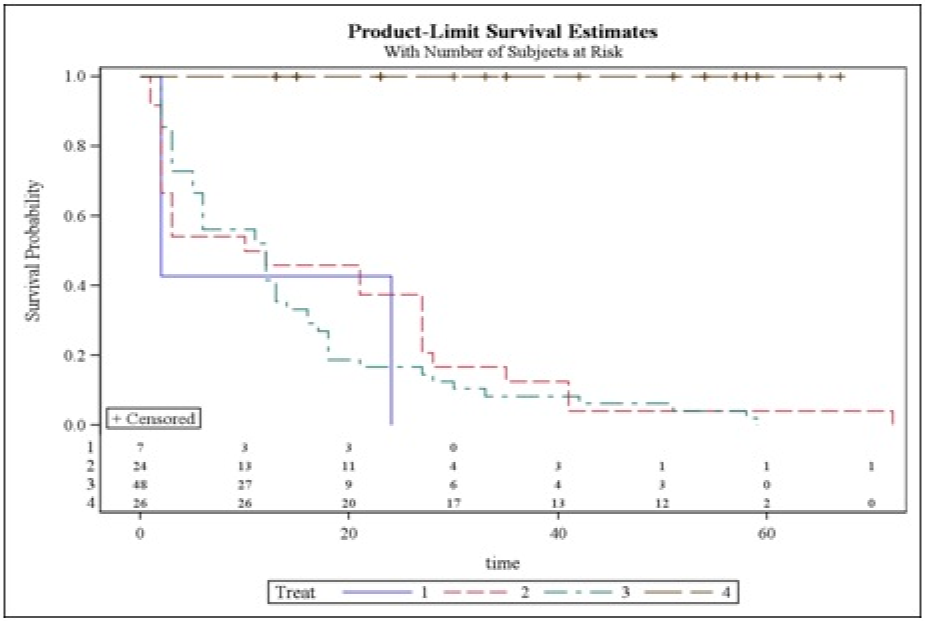

Figure 7. Overall Survival curves of patients with Uterine fibroid with different modes of treatment taken.

The graph of the survivor function of the two age group < = 13 and > 13 more or less coincide till the waiting time of 25 months. However, after 25 months the survivor function of the group > 13 lies above the < = 13. Also it is noted the event of interest of the age group is > 13.

In Figure 5 the graph of survival function of patients with Fibroid in submucal position lies above the survival curves of subserosal and intramural positions. However, the survival curve of subserosal position lies above intramural position. The functions due to three different positions of Fibroid coincide at waiting time 1 month, 40 months and 60 months.

The symptom of Fibroid in respect of the survival time and hazard shows a mixed pattern in Figure 6. Figure 7 gives the survival curves of different types of treatments. Survival curves of treatment myomectomy and hysterectomy experience more or less same survival time except those in the waiting period 15 months to 30 months.

Univariate analysis of the basic data

In this section we have attempted Univariate analysis of the data with respect to the variables viz., Age, Age of menarche, Age of marriage, Location of Fibroids and Symptoms for a deeper understanding of the survival rate. It is important to note that we have used the grouping the data in respect of the 3 variables viz., age, age of menarche and age of marriage in two distinct categories. In respect of position of Fibroid, we have adopted 3 distinct categories. For the variable viz "Symptoms" 6 categories have been considered. The usage of different categories adopted here in conforming with the standard analysis carried out in survival studies. The results are presented in the following Table 3.

Table 3. Univariate analysis of survival of 105 patients with Uterine fibroid.

Variable	Cases	Survival rate	-2log likelihood	P value (log rank)
Age				
≤39	59	23.980	544.147	0.6121
>39	20	29.510		
Age of menarche				
≤13	41	23.711	523.234	0.0267*
>13	38	27.133		
Age at marriage				
≤23	28	25.831	477.303	0.9306
>23	43	26.760		
Position of fibroid				
1-Intramural	43	22.394	434.131	0.2965
2-Subserosal	14	25.053		
3-Submucal	22	30.090		
Symptoms				
1-pain in abdomen	19	23.282		

2-inconception	17	15.000		
3-Heavy bleeding,aneamic	39	26.188	488.130	0.7265
4- bleeding after menopause	10	25.800		
5-irregular periods	1	28.000		
6-misconception	3	36.667		

*p<0.05

Interpretation of results.

We observe that age of menarche plays an important role and the results relating to this category are significant. The other variables viz., Age, Age at marriage, Position of Fibroid and symptoms do not reveal significant effect.

Multiple comparisons of treatments using log rank test:

We have attempted to compare the types of treatments using Log Rank Test. We have classified the treatments as given below:

Treatment 1: Control Treatment
Treatment 2: Myomectomy
Treatment 3: Hysterectomy
Treatment 4: Gnrh

The data relating to the 4 categories of treatments were treated us 4 different strata. The data analysis has been carried out using Chi-square Test and the results are presented in the following Table 4.

Table 4. Multiple comparison of treatments using Log rank test

Multiple Comparisons for the Logrank Test			
Strata Comparison		**Chi-Square**	**p-Values**
Treat	**Treat**		
1	**2**	0.9972	0.3180
1	**3**	12.4764	0.0004
1	**4**	56.1662	<.0001
2	**3**	4.0507	0.0442
2	**4**	38.3361	<.0001
3	**4**	50.7488	<.0001

1-Control Treatment,2-myomectomy,3-hysterectomy,4-GnrH

A closer analysis of the results presented in the above table reveals the following:

(i) The Chi-square values for Control Treatment with GrnH, Myomectomy and GnrH and Hysterectomy and GrnH are significant. This clearly reveals the major effect of GrnH.

(ii) Chi-square values for Control Treatment with Myomectomy and Myomectomy and Hysterectomy are not significant.

Cox-Regression Model for patients with Uterine Fibroid

We have also made attempts for analyzing the effects of multi factors involved in this study. Cox-Regression Model is fitted using the variables viz., Age (X_1), Age of Menarche (X_2), Age at marriage (X_3), Location (X_4), and Symptoms (X_5) and the results are presented in the following Table 5.

As noted earlier here also we observe that the variables viz., Age of Menarche plays the key role and it reveals that the effect is significant.

Table 5. Cox regression analysis of patients with Uterine fibroid

Variable	**B**	**Standard Error**	**P**
Age X_1	-0.0490	0.2415	0.8391
Age of menarche X_2	-0.5727	0.2775	0.0391*
Age at marriage X_3	-0.0183	0.2393	0.9390
Location X_4	-0.1279	0.1589	0.4207
Symptoms X_5	-0.0374	0.0566	0.5087

*p<0.05

Results and Discussion

The analysis and results of the study are given below:-

(i) The percentage of patients with position of fibroids is: intramural (50%), submucal (30%) and subserosal (20%)

(ii) Estimates of the survival function tend to zero as the survival time increase in a non-uniform manner.

(iii) Estimates of the hazard function tend to one as the survival time increases in non-uniform manner.

(iv) The symptoms of Fibroid in respect of survival time and hazard show a mixed pattern.

(v) The variable viz., Age of menarche plays a significant role among the patients with Fibroids.

(vi) Comparison of treatments based on strata analysis clearly reveals the significant effect of the treatment Gonadotrophin releasing Hormone (GnrH) which is a hormone medicine that causes low level of oestrogen in the body. Fibroids shrink if the level of oestrogen falls. This can ease heavy periods and pressure symptoms due to fibroids. However, a oestrogen level can cause symptoms similar to going through the menopause (hot flushes etc.,).It may also increase the risk of ‘thinning’ of the bones (osteoporosis).Therefore, this treatment is given for a maximum of six months.

References

[1] An-Kuiyang, Tian-Run Liu, Fu-Jin chen, Xiu-Fang Ma, Zhu-Ming Guo, MingSong, Qiu-Li Li and Zang Yuan Zeng, “Survival analysis of 229 patients with advanced squamous cell carcinoma of the oral tongue”.

[2] J. Berkson and R.R. Gage, “Calculation of Survival rates for Cancer”, Proceedings of staff Meetings, Mayo Clinic, 25 (1950), 250.

[3] N.E. Breslow and J. Crowley, “A Large sample study of the Life Table and Product limit Estimates under Random Censoring”, Annals of Statistics, 2 (1974) 437-453.

[4] C. Zou, G. Yin, L. Feng and Z. Wang, “Non parametric maximum Likelihood Approach to Multiple change point problems”, The Annals of Statistics,42 (2014), 970-1002.

[5] D. Collett, “Modeling Survival data in Medical Research”, 40-42 (1994), 204-207.

[6] D.R. Cox, “A note on the graphical analysis of survival data”, Biometrika, 66 (1959), 188-90.

[7] D.R. Cox, “Regression Models and life tables”, Journal of Royal Statistical Society, Series B, 34 (1972), 187-220.

[8] S.J. Cutler and F. Ederer, “Maximum Utilization of the Life Table Method in Analyzing Survival”, Journal of Chronic Diseases, 8 (1958), 699-712.

[9] E.A. Gehan, “Estimating survival function from the life table”, Journal of chronic Diseases, 21 (1969), 629-44.

[10] E.L. Kaplan and P. Meier, “Nonparametric estimation from incomplete observations”, Journal of the American Statistical Association, 53 (1958), 457-81.

[11] N. Mantel, “Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration”, Cancer Chemotherapy Reports, 50 (1966), 163-170.

[12] P. Meier, “Statistics and Medical Experimentation”, Biometrics, 31 (1975), 511-529.

[13] R. Peto and J. Peto, “Asymptotically Efficient Rank Invariant Proceedures”, Journal of the royal Statistical Society, Series A, 135 (1972), 185-207.

[14] Peto, M.C. Pike, P. Armitage, N.E. Breslow, D.R. Cox, S.V. Howard, N. Mantel, K. McPherson, J. Peto and Smith, “Design and Analysis of Randomised Clinical trails requiring prolonged Observation of Each patient”, British Journal of Cancer, 34, (1976) 585-612, 35 (1977), 1-39.

Authors

First Author - Dr. J. Jothikumar, Associate Dean, Faculty of Science and Humanities, SRM University, Kattangulathur.
Email id : jjothikumar@yahoo.co.in.

Secind Author -V. Sumathy, Assistant Professor in Statistics, Kumararani Meena Muthiah College of Arts and Science, Adyar, Chennai.
Email id : v.sumathy@yahoo.com

Correspondance Author : V. Sumathy, Assistant Professor in Statistics, Kumararani Meena Muthiah College of Arts and Science, Adyar, Chennai.
Email id : v.sumathy@yahoo.com, Phone num: 9884682386/9941732000.