

Georgia State University

ScholarWorks @ Georgia State University

Mathematics Theses

Department of Mathematics and Statistics

8-6-2007

Study of Factors of Affecting Recurrence of Myoma after Myomectomy

Lu Wang

Follow this and additional works at: https://scholarworks.gsu.edu/math_theses



Part of the [Mathematics Commons](#)

Recommended Citation

Wang, Lu, "Study of Factors of Affecting Recurrence of Myoma after Myomectomy." Thesis, Georgia State University, 2007.

https://scholarworks.gsu.edu/math_theses/33

This Thesis is brought to you for free and open access by the Department of Mathematics and Statistics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Mathematics Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

STUDIES OF FACTORS AFFECTING RECURRENCE OF MYOMA AFTER MYOMECTOMY

By

Lu Wang

Under the director of Yu-Sheng Hsu

This study is performed to evaluate the factors associated with the recurrence of myoma after Myomectomy. Identifying the factors of myoma recurrence will assist the patient and her gynecologist in deciding the most appropriate method of treatment according to her specific social, medical and emotional needs. Multiple logistic regression is used to determine the factors affecting the recurrence. 'Age of Surgery', 'Tumor Size', 'Pelvic Pain' and the interaction between the 'Age of Surgery' and 'Tumor Size' are significant in the final model. Kaplan-Meier method is used to calculate the cumulative recurrence rate. The 5 year cumulative recurrence rate is 24.32% and the 10 year cumulative recurrence rate is 32.57%.

INDEX WORDS: Logistic regression, Kaplan-Meier, Myomectomy

**STUDIES OF FACTORS AFFECTING RECURRENCE OF MYOMA AFTER
MYOMECTOMY**

By

LU WANG

A Thesis Submitted In Partial Fulfillment of the Requirement for the Degree of

Masters of Science

In the College of Arts and Sciences

Georgia State University

2007

**STUDIES OF FACTORS AFFECTING RECURRENCE OF MYOMA AFTER
MYOMECTOMY**

By

LU WANG

Major Professor : Dr. Yu-Sheng Hsu

Committee: Dr. Xu Zhang

Dr. Jia-Wei Liu

Electronic Version Approved By :

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
Aug, 2007

TABLE OF CONTENTS

LIST OF TABLES.....	v
LIST OF FIGURES.....	vi
CHAPTERS	
1 INTRODUCTION.....	1
Background.....	1
Source of Data.....	6
2 METHODOLOGY.....	11
Method Of Analysis.....	11
Statistical Analysis.....	11
Goodness of Fit.....	18
Cross Validation.....	22
Survival Analysis.....	24
3 RESULTS.....	27
Odds Ratio.....	27
4 DISCUSSION.....	28
REFERENCES.....	29
APPENDICES.....	31
SAS Code.....	31

LIST OF TABLES

Table 1. BMI versus Recurrence Rate.....	7
Table 2. Age Groups versus BMI.....	8
Table 3. Tumor Size versus Recurrence Rate.....	9
Table 4. Pregnancy Before versus Recurrence Rate.....	9
Table 5. Results from Univariate Test.....	12
Table 6. Results from stepwise selection with SLE=0.4, SLS=0.3. . .	12
Table 7. Variable ‘Age of Surgery’ Transformation.....	15
Table 8. Correlation of the variables.....	16
Table 9. Final Model of Logistic Regression.....	16
Table 10. HL Goodness fit test.....	19
Table 11. Sensitivity and Specificity with different probability Level.....	20
Table 12. ROC Value.....	22
Table 13. Example of Cross Validation.....	23

LIST OF FIGURES

Figure 1. UAE.....	3
Figure 2. Types of fibroids are classified by their location.....	4
Figure 3. Sensitivity and Specificity.....	21
Figure 4. Receiver Operating Characteristic Curve (ROC).....	21
Figure 5. Kaplan-Meier Cumulative Recurrence Rate.....	24
Figure 6. Kaplan-Meier Survival Rate.....	26

FACTORS AFFECTING RECURRENCE OF LEIOMYOMA AFTER MYOMECTOMY

Chapter 1

INTRODUCTION

Uterine myomas, more commonly called fibroids, are benign neoplasms of uterine smooth muscle. They are the most common pelvic tumors in women. In the United States, uterine myomas are a leading cause of hospitalizations for gynecologic disorders unrelated to pregnancy and the most frequent reason for Hysterectomy, accounting for 33% of all such surgeries annually. In principle, the fibroid tumors can cause infertility, distorting the uterine cavity or occluding tubes and altering vascularization. Women with these tumors experience heavy period, menorrhagia, anaemia and pelvic pain. Ultrasound is generally been used to detect fibroids. It is a painless, safe and reliable way to assess the uterus, ovaries and to look for uterine fibroids. In this process, sound waves, not radiation, are used to create pictures of uterus. When the finding fibroids are small and cause no symptoms, no treatment is required. However, in presence of symptomatic or large fibroids, hysterectomy is the conventional treatment, which completely removes the uterus. It is recommended for women older than 40 and women younger than 40 who have completed childbearing. Hysterectomy is the only treatment with the guarantees that there will be no recurrence of fibroids. The drawback is that after hysterectomy, there is an increased prevalence of problems related to sagging pelvic organs such as “dropped bladder” (in medical terms these are known as genital prolapse, such as vaginal vault prolapse, enterocele, cystocele and rectocele). These conditions may cause symptoms, such as urination and defecation disorders, and may require surgical repair. Also it has been shown that after hysterectomy, women tend to enter menopause earlier, by as much as four years according to one study. On top of that, as more

women have become better informed on health issues recent years, they have sought treatments that preserve the pelvic organs, especially in the presence of a benign disease. In principle, many women are strongly opposed to the removal of any organs, genital or otherwise, unless absolutely necessary.

Uterine fibroid embolization (also known as uterine artery embolization) represents a fundamentally new approach to the treatment of fibroids. Embolization is a minimally invasive means of blocking the arteries that supply blood to the fibroids. The procedure was first used in fibroid patients in France as a means of decreasing the blood loss that occurs during myomectomy. It was discovered that after the embolization, while awaiting surgery, many patient's symptoms went away and surgery was no longer needed. The blockage of the blood supply caused degeneration of the fibroids and this resulted in resolution of their symptoms. This has led to the use of this technique as a stand-alone treatment for symptomatic fibroids.

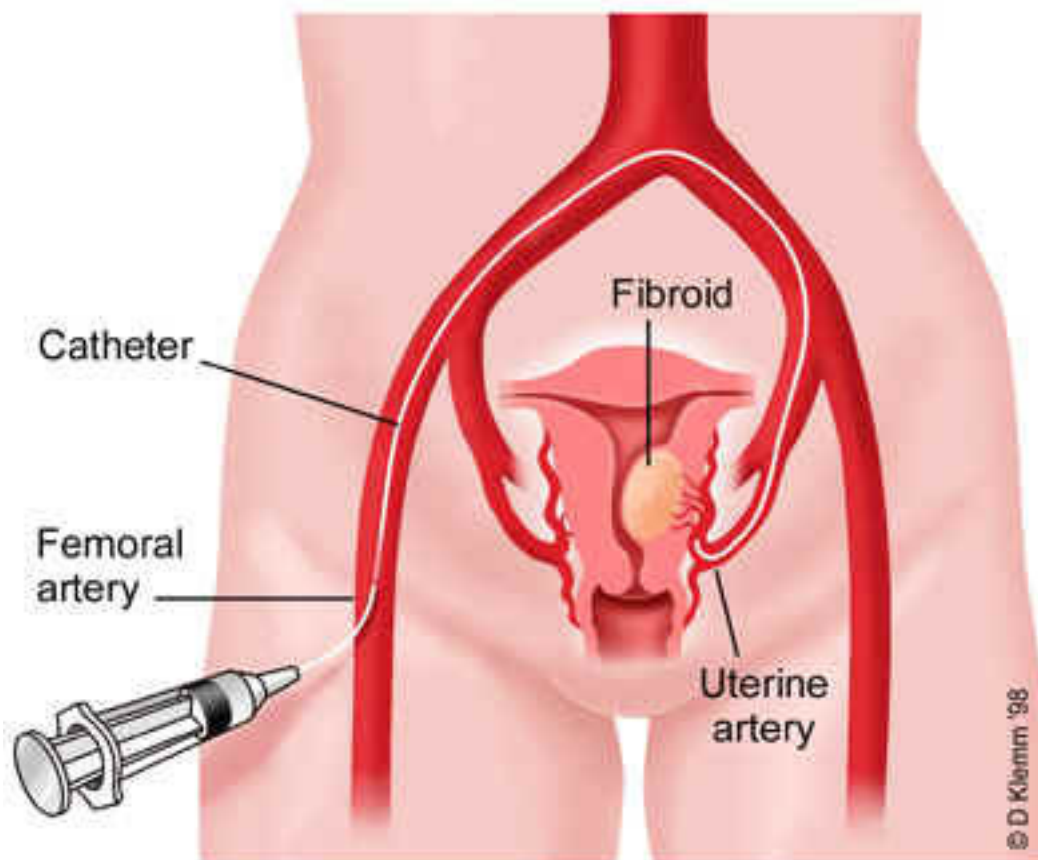


Figure 1: UAE: A tiny particle is injected through blood vessels to block the arteries supplying fibroids

As of this time, approximately 2000 to 3000 patients have had this procedure worldwide. The expected average reduction in the volume of the fibroids is 50% in three months, with reduction in the overall uterine volume of about 35%. The long-term outcome is not known, in that the arteries could reopen or collateral vessels could be recruited which might allow regrowth of the fibroids. Therefore, it is not yet known if the fibroids can regrow. The large majority of the patients that had this procedure are finished with childbearing and so few women have tried to become pregnant after this procedure. Thus far, at least a dozen patients have become pregnant after this procedure worldwide. As the outcome of pregnancy following UAE is not known, the procedure is not recommended for women who plan to have children.

When the conditions occur as seeking for fertility, Myomectomy is considered. It is a surgical procedure that allows the gynecologist surgeon to remove the fibroids and preserve the fertility. There are three methods for Myomectomy: Hysteroscopy, which involves inserting a lighted viewing instrument through the vagina and into the uterus; Laparoscopy, which uses a lighted viewing instrument and one or more small cuts (incisions) in the abdomen; Laparotomy, which uses a larger incision in the abdomen. The method used depends on the size, location and number of fibroids. Hysteroscopy can be used to remove fibroids on the inner wall of the uterus that have not grown deep into the uterine wall, such as intracavitary myomas that are inside the uterus and submucous myomas that are partially in the cavity and partially in the wall of uterus. Laparoscopy is usually reserved for removing one or two fibroids, up to about 2 inches across, that are growing on the outside of the uterus, such as pedunculated myomas. Laparotomy is used to remove large fibroids, many fibroids at once or fibroids that have grown deep into the uterine wall.

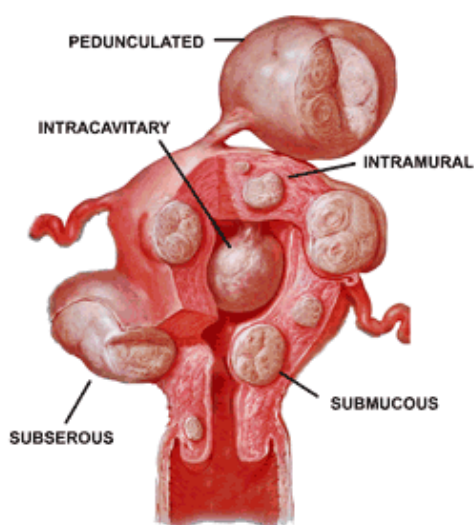


Figure 2: Types of fibroids are classified by their location.

Myomectomy requires more expertise and surgical skills than Hysterectomy; it is more likely to result in prolonged surgery, significant blood loss and other postoperative complications. Therefore, it is not surprising that the average gynecologist tends to offer hysterectomy as a treatment for fibroids, rather than Myomectomy.

Myomectomy preserves the fertility for women, however, it has a big drawback: the fibroids may recur. Clinical studies have reported that the recurrence rates ranging from 5 to 30 % (Buttram & Reiter 1981) and the most updated paper showed 30% of recurrence rate. This wide range may be explained by different criteria used in diagnosis and deficiencies in long-term follow-up. The effect of potential prognostic factors has still not been properly investigated.

Much research has been devoted to finding the factors affecting the recurrence after Myomectomy. A paper “Recurrence of fibroids after Myomectomy” published in 1991 showed that there were no associations between risk of recurrence and age of the surgery or number of pregnancy before Myomectomy. It was also said that there was a steady increase in recurrence-ten years after Myomectomy and fewer recurrences in women who had had a single fibroid removed. It also gave an explanation for the results: First, pregnancy may have a long lasting protective effect on the uterus that inhibits the growth of fresh fibroids. Second, women with small fibroids that have gone undetected at the time of surgery are both less likely to identifiable fibroids in the future. Third, women who have had one or more successful pregnancies after Myomectomy are more likely to use combined oral contraceptives. Another paper from Clinique Universitaire Baudeocque (Paris, France) published in year 2000 showed that Age, number of tumors are significant factors of recurrence; the pregnancy after Myomectomy is associated with a lower risk of recurrence

and it stated that the real risk of recurrence after Myomectomy is difficult to assess. A paper published in 1991 from several Docrots in Milan, Italy also found out that women gave birth to a child has a lower rate than those who did not; women with a single fibroid tended to experience a lower rate and the cumulative 10-year recurrence rate after Myomectomy was 27% and the number of myomas removed at Myomectomy has been associated positively with the risk of recurrence. In 1998, "Recurrence rate after laparoscopic Myomectomy" was published and stated that out of 114 patients, the cumulative risk of recurrence rate was 10.6% after 1 year, 31.7% after 3 years and 51.4% after 5 years; risk factors were number of myomas and depth of penetration.

A retrospective study of factors affecting recurrence after Myomectomy is conducted of 257 patients underwent Myomectomy from January 1st, 1992 to May 1st, 2007. All surgeries were done by one surgeon, Dr Hanafi.

Sources of Data

Between 1992 and 2007, 257 women underwent Myomectomy with Dr. Hanafi at Saint Joseph Hospital. Patients' surgery information has been collected from their charts and survey has been conducted through mails, phone calls and personal interview to get the patients' pregnancy information after first Myomectomy. Due to the lack of participation of the survey and consequently censored at the time of their last follow-up visit, plus that the criteria at least one myoma is larger than 2cm, out of the 257 patients, only 109 patients' information were fully obtained at the cut off time May 1st 2007.

The patients' data include BMI, their age at first Myomectomy surgery, number of pregnancy before the surgery, whether or not have had deliveries before the surgery, number

of myoma, the largest size of the myoma, willingness of conceive after the surgery, pregnancy after the surgery, five symptoms before the surgery (heavy period, irregular period, dysmen, dyspar, pelvic pain) and the response variable Recurrence. The age of those 109 patients is ranged from 23 to 50 and their average age is 36. The BMI is ranged from 16.91 to 56.68 and has an average of 29.5.

Based on Dr. Hanafi 30 years of experience, BMI, willingness of conceive, size of tumor and the age at first surgery are clinically important. And the paper published before showed that the number of Myomas is the significant factor. BMI has been discreted into four groups: Less than 18.5-Light weight; Between 18.5 and 24.9-Normal Weight; Between 25 and 29.9-Overweight; greater than 30-Obesity. Age and tumor size are discreted into three groups: 20-30; 30-40 and 40-50 for age and 2-4cm; 4-6cm and greater than 6cm. Graph these three discreted factors associate with their recurrence rate, interesting patterns can be seen.

BMI Group	BMI	N	Rec	RR
Light Weight	<18.5	2	1	50%
Normal	18.5-24.9	30	11	37%
Over weight	25-29.9	31	8	26%
Obesity	>30	46	16	35%

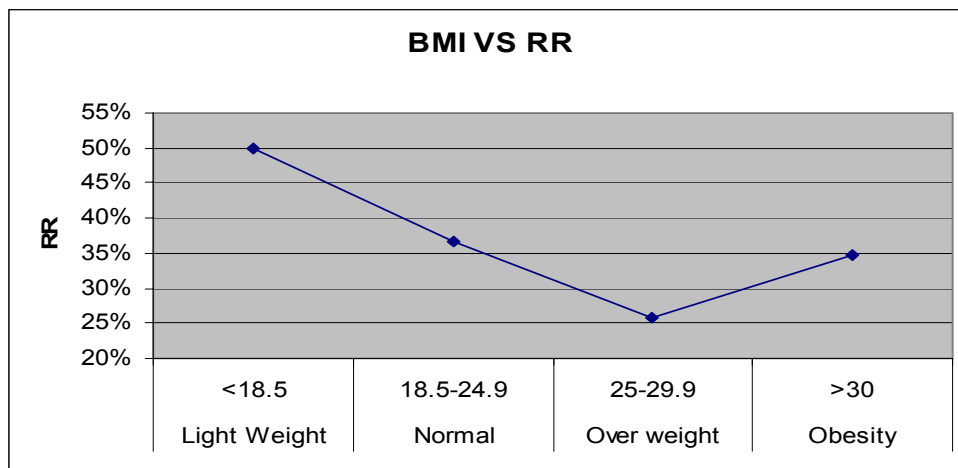


Table1: BMI versus Recurrence Rate

From the table1, it is surprising that the heavier the weight the lower the rate of recurrence. And the lowest recurrence rate occurred at the Over weight group, which is 26%.

Age Group	N	Rec	RR
20-30	9	3	33%
30-40	77	30	39%
40-50	23	3	13%

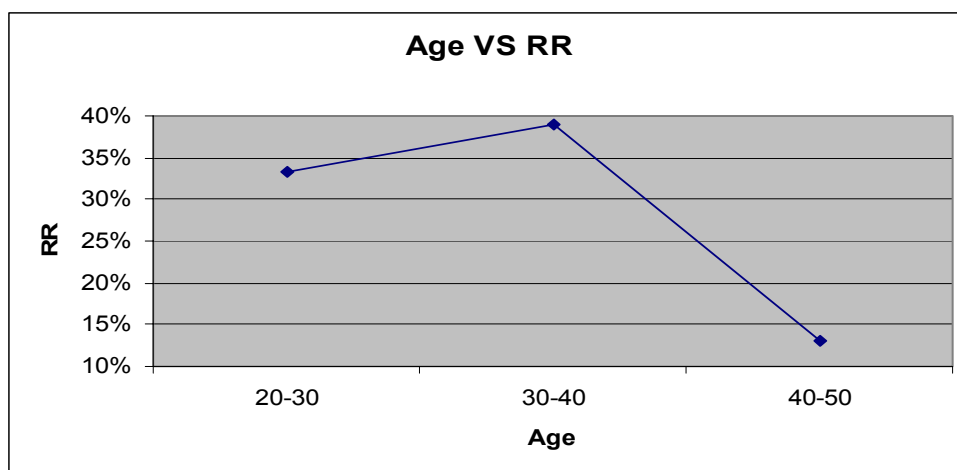


Table2: Age Groups versus BMI

The table 2 has shown increasing rate from 20-30 to 30-40 and decreasing rate from 30-40 to 40-50. However, for the age group of 20-30, there are only 9 observations.

Size Group	N	Rec	RR
2-4 cm	59	24	41%
4-6 cm	27	8	30%
6+ cm	23	4	17%

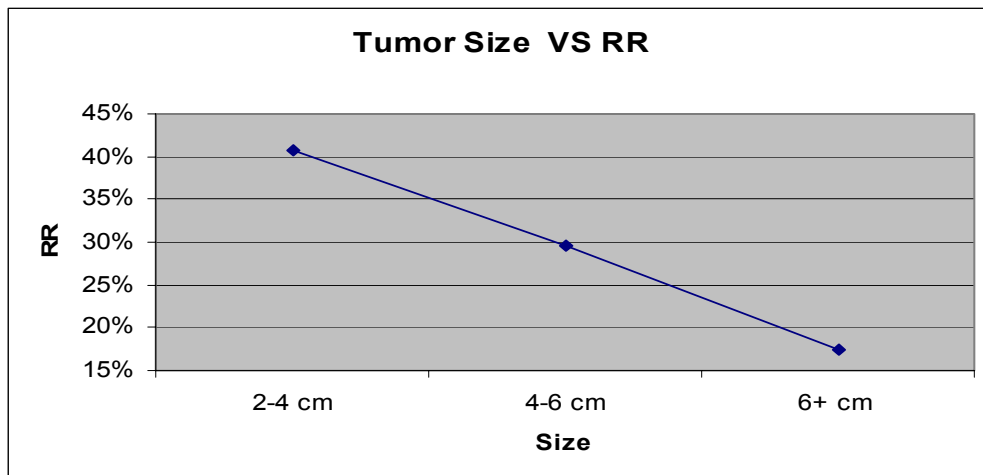


Table 3: Tumor Size versus Recurrence Rate

Table 3 has shown a linear relationship between the tumor size and the rate of recurrence. This indicates that as the tumor size gets larger, the rate of recurrence gets lower.

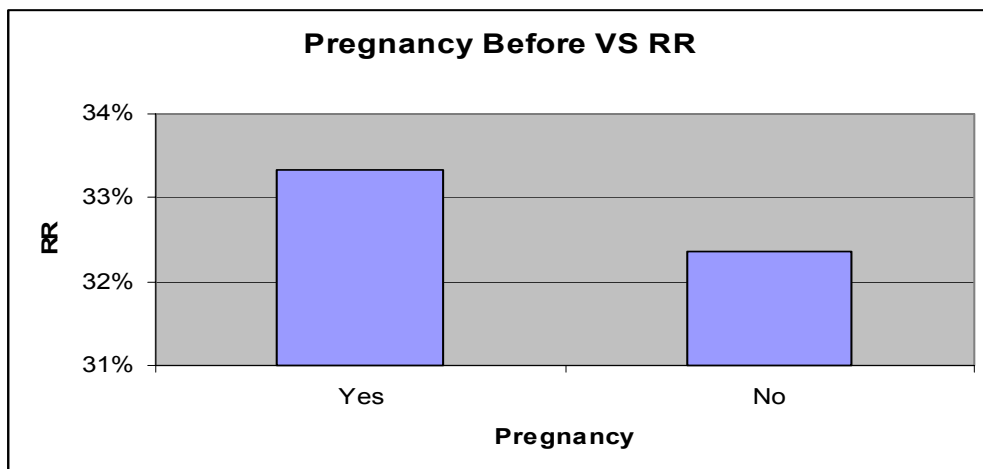
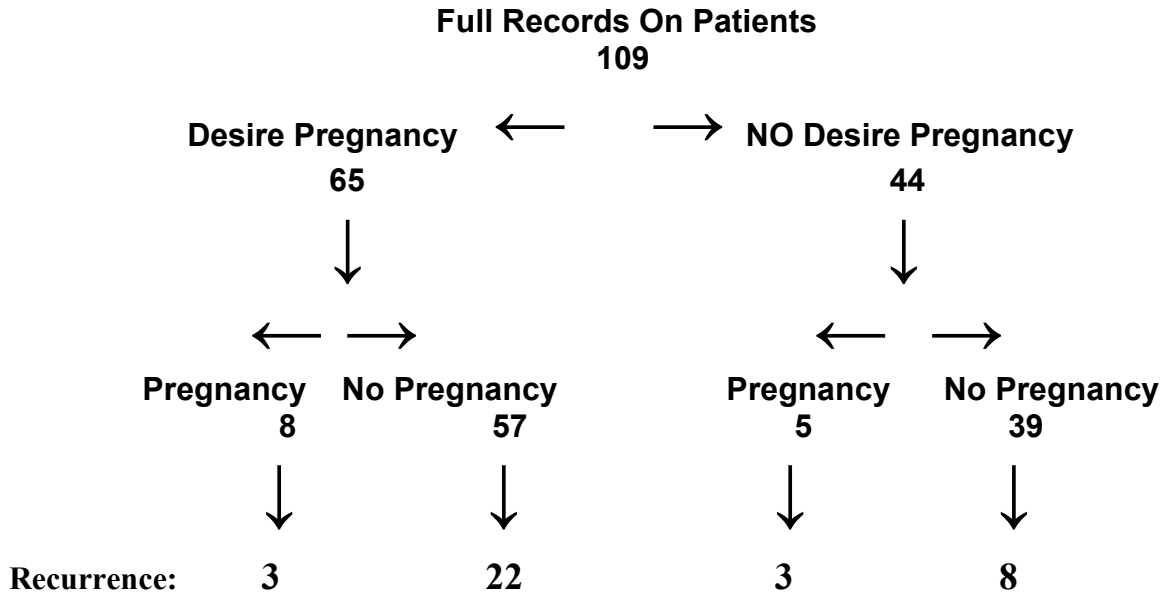


Table 4: Pregnancy Before versus Recurrence Rate

Table 4 has shown that whether the patients have pregnancy before the first surgery does not make much difference. The recurrence rate for pregnancy before is 33% compare with non-pregnancy 32%.

A classification table is made for the factor willingness of conceive.



With patients who desire pregnancy, the recurrence rate is 37.5% for those patients actually got pregnancy and 38.6% who did not get pregnancy. And among those patients who do not desire pregnancy, the recurrence rate is 60% for those who became pregnancy and 20.5% who did not get pregnancy. And there were 38.5% of patients who desire to conceive had a recurrence and only 25% of those who did not desire to conceive experienced recurrence.

Chapter 2

METHODOLOGY

Method of Analysis

The data of patients' consists primarily of dichotomous variables. Variables such as Pregnancy-before, willingness-to-conceive, whether or not having any symptoms, the response variable-recurrence is also a binary variable. Many different methods are available to analyze this data; however, for this study multiple logistic regression has been chosen as an optimal method to use for the dichotomous variables. The goal of the method is to select those variables that result in a best model to predict the factors that affecting the recurrence of Myomectomy.

Statistical Analysis

A univariate analysis was conducted on each of the independent variables to determine which variables were significant in predicting the recurrence after Myomectomy. Variables whose univariate test has a P-value < 0.25 is a candidate for the model along with all variables of known clinical importance. The 0.25 level is suggested to be used as a screening criterion for variable selection is based on work by Bendel and Afifi(1977) on linear regression and on the work by Mickey and Greenland (1989) on logistic regression.

Label	Test	Chi-Square	DF	P> Chi-Square
BMI	Likelihood Ratio	0.0012	1	0.9726
Age of Surgery	Likelihood Ratio	5.1041	1	0.0239
Pregnancy Before	Likelihood Ratio	0.9364	1	0.3332
Delivery	Likelihood Ratio	0.0591	1	0.8079
Number of Myoma	Likelihood Ratio	0.0027	1	0.9585
Tumor Size	Likelihood Ratio	7.3906	1	0.0066
Single Tumor	Likelihood Ratio	0.0007	1	0.9784
Willing To Conceive	Likelihood Ratio	2.1917	1	0.1388
Pregnancy After Surgery	Likelihood Ratio	1.1	1	0.2943
Heavy Period	Likelihood Ratio	0.3278	1	0.5669
Irregular Period	Likelihood Ratio	0.2353	1	0.6276
Dyspar	Likelihood Ratio	0.002	1	0.9643
Dysmen	Likelihood Ratio	1.3333	1	0.2482
Pelvic Pain	Likelihood Ratio	1.0118	1	0.3145

Table 5: Results from Univariate Test.

Another method to variable selection is to use a stepwise method in which variables are selected from the model in a sequential fashion based solely on statistical criteria. Stepwise selection of variables is widely used in linear regression. All major software packages have either a separate program or an option to perform this type of analysis for the logistic regression.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr> Chi-Square
Intercept	1	3.5732	2.0102	3.1596	0.0755
Age of surgery	1	-0.1046	0.0494	4.4799	0.0343
Tumor Size	1	-0.3734	0.1323	7.9728	0.0047
Willing to conceive	1	0.7306	0.4924	2.202	0.1378
Dysme	1	0.6641	0.4877	1.8542	0.1733
Pelvic Pain	1	1.1277	0.5696	3.9206	0.0477

Table 6: Results from stepwise selection with SLE=0.4, SLS=0.3.

Table 5 and Table 6 showed that Age of surgery, tumor size and pelvic pain are potentially significant.

Variable Transformation

Once a model that contains the essential variables has been established, the assumption of linearity in the logit should be checked. Two methods to assess this problem: Design variables and fractional polynomials. A design variable uses a scatter plot smooth, plotting the results on the logit scale. Unfortunately, scatter plot smoothing methods are not easily extended to multivariable models and thus have limited applicability in the model refinement step. Another more analytic approach is to use the method of fractional polynomials, developed by Royston and Altman (1994) to do transformation. We wish to determine what value of x^p yields the best model for the covariate. Royston and Altman propose replacing full maximum likelihood estimation of power by a search through a small but reasonable set of possible values. For a single continuous covariate, the logit, which is linear in the covariate is

$$g(x, \beta) = \beta_0 + x\beta_1$$

Where β denotes the vector of model coefficients. One way to generalize this function is to specify it as

$$g(x, \beta) = \beta_0 + \sum_{j=1}^J F_j(x)\beta_j$$

The functions $F_j(x) = x^{P_j}$ are a particular type of power function. The value of the first function is $F_1(x) = x^{P_1}$. In theory, the power P_1 , could be any number, but in most applied settings it makes sense to try to use something simple. Royston and Altman (1994)

propose restricting the power to be among those in the set $\mathcal{G} = (-2, -1, -0.5, 0, 0.5, 1, 2, 3)$.

Where $P_1=0$ denotes the log of the variable. The remaining functions are defined as

$$F_j(x) = \begin{cases} x^{p_j}, p_j \neq p_{j-1} \\ F_{j-1}(x) \ln(x), p_j = p_{j-1} \end{cases}$$

for $j=2 \dots J$ and restricting powers to those in \mathcal{G} .

Implementation of the method requires, for $J=1$, fitting 8 models that is $P_1 < \mathcal{G}$. The best model is the one with the largest log likelihood. The process is repeated with $J=2$ by fitting the 36 models obtained from the distinct pairs of powers, that is, $(p_1, p_2) \in \mathcal{G} \times \mathcal{G}$ and the best model is again the one with the largest log likelihood.

The relevant question is whether either of the two best models is significantly better than the linear model. Let $L(1)$ denote the log likelihood for the linear model, $J=1$, and $P_1=1$, and $L(P_1)$ denote the log likelihood for the best $J=1$ model and $L(P_1, P_2)$ denote the log likelihood for the best $J=2$ model. Then compare the partially likelihood ratio test with the linear model to the best $J=1$ model, is approximately distributed as chi-square with 1 degree of freedom under the null hypothesis of linearity in x . The partial likelihood ratio test comparing the best $J=1$ model to the best $J=2$ model is approximately distributed as chi-square with 2 degrees of freedom. Similarly, the partial likelihood ratio test comparing the linear model to the best $J=2$ model is distributed approximately as chi-square with 3 degrees of freedom.

‘Age of surgery’ and ‘Tumor size’ are the continuous variables in the model that need to be checked for linearity.

Transformation Models		-2 Log L	
Best J=2 model:	$Age = x^3 + x^3 * \ln(x)$	130.0698	
Best J=1 model:	$Age = x^3$	131.9099	
Linear model:		133.189	
Partially Likelihood Test			
DF		Chi-Square	P>Chi-Square
1	$G(1, p_1)$	1.2791	0.2580
2	$G[p_1, (p_1, p_2)]$	1.8401	0.3984
3	$G[1, (p_1, p_2)]$	3.1192	0.3736

Table 7: Variable ‘Age of Surgery’ Transformation

Age transformation table has shown the best J=1 model, best J=2 model and the comparisons with the linear model. However, the transformation models are not significantly different than the linear model; therefore, linear model is selected. Same process for the variable ‘Tumor size’, and the linear model is selected too.

Interaction

Interaction effects are an important consideration for any model and must be thoroughly explored to determine if there is a significant interaction that should be included in the final model. In any model an interaction between two variables implies that the effect of one of the variables is not the same over levels of the each other. The final decision as to whether an interaction term should be included in a model should be based on statistical as well as practical considerations. The correlation for each of the variable is calculated to assess the significant interaction.

	Recurrence	BMI	Age of Surgery	Pregnancy Before	Delivery	Number of Myoma	Tumor Size	Single Tumor	Willing To Conceive	Pregnancy After	Heavy Period	Irregular Period	Dyspar	Dysmen
Recurrence	1.00													
BMI	0.00	1.00												
Age of Surgery	-0.21	0.11	1.00											
Pregnancy Before	0.09	-0.03	0.19	1.00										
Delivery	-0.02	0.07	0.19	0.64	1.00									
Number of Myoma	0.00	-0.01	0.00	-0.23	-0.31	1.00								
Tumor Size	-0.24	0.00	-0.08	-0.13	-0.06	0.17	1.00							
Single Tumor	0.00	-0.13	0.02	0.20	0.14	-0.62	0.02	1.00						
Willing To Conceive	0.14	-0.05	-0.16	0.14	-0.02	0.05	-0.04	0.06	1.00					
Pregnancy After	0.10	-0.23	-0.11	-0.01	-0.17	0.03	-0.01	0.02	0.01	1.00				
Heavy Period	-0.06	0.13	0.09	-0.19	-0.02	0.03	0.04	-0.04	-0.21	-0.22	1.00			
Irregular Period	0.05	0.05	0.00	-0.08	0.00	0.05	-0.02	-0.03	0.07	-0.12	0.06	1.00		
Dyspar	0.00	-0.03	-0.17	0.01	0.08	0.00	0.04	-0.09	-0.11	-0.12	0.24	-0.21	1.00	
Dysmen	0.11	-0.02	-0.19	-0.19	-0.11	-0.01	-0.01	-0.02	-0.17	-0.09	0.23	-0.08	0.37	1.00
Pelvic Pain	0.10	-0.09	0.06	0.06	-0.06	-0.07	0.16	0.17	-0.01	0.01	0.12	-0.02	-0.11	-0.23

Table 8: Correlation of the variables

The correlations among variables indicate the possible multicollinearity problem. The final model should be carefully examined and the interpretation also should be done with caution. A stepwise logistic regression with all the main variables and the interactions is conducted to determine whether any of those interactions are significant with SLE=0.15 and SLS=0.15. Only the interaction between Age of surgery and Tumor size turn out to be significant in the model.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr> Chi-Square
Intercept	1	11.6422	4.9733	5.4802	0.0192
Age of Surgery	1	-0.3102	0.1375	5.0905	0.0241
Tumor Size	1	-1.8179	0.9624	3.5683	0.0589
Pelvic Pain	1	0.9076	0.5326	2.9036	0.0884
Age * Tumor size	1	0.0412	0.0264	2.4401	0.1183

Table 9: Final Model of Logistic Regression

Based on the results of the variable selection process, a final model has been developed to determine the risk factors associated with recurrence of fibroids. The final logistic regression model is:

$$\hat{\pi}(x) = \frac{e^{11.6422 - 0.3102 * Age - 1.8179 * Tumorsize + 0.9076 * pelvicpain + 0.04128 * Age * Tumorsize}}{1 + e^{11.6422 - 0.3102 * Age - 1.8179 * Tumorsize + 0.9076 * pelvicpain + 0.04128 * Age * Tumorsize}}$$

Goodness of Fit

Once the model is established, it is important to determine if the model is a good fit of the data. Also, we need to know if the model is a good predictor for the response variable. Finally, we like to check if the overall patterns of the model are relevant in representing the predicted outcomes. Hosmer Lemeshow(HL) goodness-of-fit can be used for some of the above purposes. The null hypothesis being tested with HL test is that the model is a good fit. HL divided the data into $g=10$ deciles of risk groups to do the comparisons between the observed to the expected number of outcomes for value of x . The first group containing the $n_1=n/10$ subjects having the smallest estimated probabilities, and the last group containing the $n_{10}=n/10$ subjects having the largest estimated probabilities. The HL test uses a Pearson Chi-Square statistic

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Where n'_k is the total number of subjects in the K^{th} group, C_k denotes the number of covariate patterns in the K^{th} decile,

$$O_k = \sum_{j=1}^{C_k} y_i$$

is the number of response among the C_k covariate patterns, and

$$\bar{\pi}_k = \sum_{j=1}^{C_k} \frac{m_j \hat{\pi}_j}{n'_k}$$

is the average estimated probability.

The distribution of \hat{C} was determined by simulation studies to be approximately χ^2 with 8 degrees of freedom (Hosmer and Lemeshow, 1980). The HL test rejects the fit of logistic model at α level when $C > \chi_{8, 1-\alpha}^2$.

Partition for the Hosmer and Lemeshow Test					
Group	Total	Recurrence = 1		Recurrence = 0	
		Observed	Expected	Observed	Expected
1	11	0	0.69	11	10.31
2	11	2	1.44	9	9.56
3	11	1	1.93	10	9.07
4	11	3	2.54	8	8.46
5	11	4	3.06	7	7.94
6	11	4	3.69	7	7.31
7	11	4	4.32	7	6.68
8	11	6	5.43	5	5.57
9	11	6	5.85	5	5.15
10	10	6	7.06	4	2.94

Hosmer and Lemeshow Goodness-Of-Fit Test		
Chi-Square	DF	Pr>Chisq
2.7828	8	0.9472

Table 10: HL Goodness fit test

The large P-value from Table proved that the model is a good fit of the data.

ROC

Sensitivity and specificity are two diagnostic measurements used to assess the model in terms of reliability the model to classify observations correctly. Sensitivity is the probability of correct classification among all true positives. On the other hand, the specificity is the probability of correct classification among all true negatives. Sensitivity and Specificity rely on a single cut point to classify a test results as positive. The most commonly

used measurement to gauge the classification method is the area under the ROC curve (Receiver operating Characteristic), which is a plot of sensitivity versus 1-specificity overall possible cut points. This curve, ranges from zero to one, provided a measure of the model's ability to discriminate between those subjects who experience the outcome of interest versus those who don't. If the objective was to choose an optimal cut point for purpose of classification, one might select a cut point that maximizes one of sensitivity and specificity, while fixed the other one. The other way to gauge the discrimination power is to find the point, where sensitivity equals to the specificity. For example, if the cutoff point yield sensitivity =specificity=0.9, then the method is very good.

Prob Level	Correct Identifies		Incorrect Identifies		Sensitivity	Specificity	False Positive	False Negative
	Recur=1	Recur=0	Recur=1	Recur=0				
5%	36	7	66	0	100	9.6	64.7	0
10%	34	12	61	2	94.4	16.4	64.2	14.3
15%	32	21	52	4	88.9	28.8	61.9	16
20%	30	31	42	6	83.3	42.5	58.3	16.2
25%	28	36	37	8	77.8	49.3	56.9	18.2
30%	25	41	32	11	69.4	56.2	56.1	21.2
35%	23	52	21	13	63.9	71.2	47.7	20
40%	20	56	17	16	55.6	76.7	45.9	22.2
45%	16	57	16	20	44.4	78.1	50	26
50%	11	61	12	25	30.6	83.6	52.2	29.1
55%	8	64	9	28	22.2	87.7	52.9	30.4
60%	7	66	7	29	19.4	90.4	50	30.5

Table 11: Sensitivity and Specificity with different probability Level

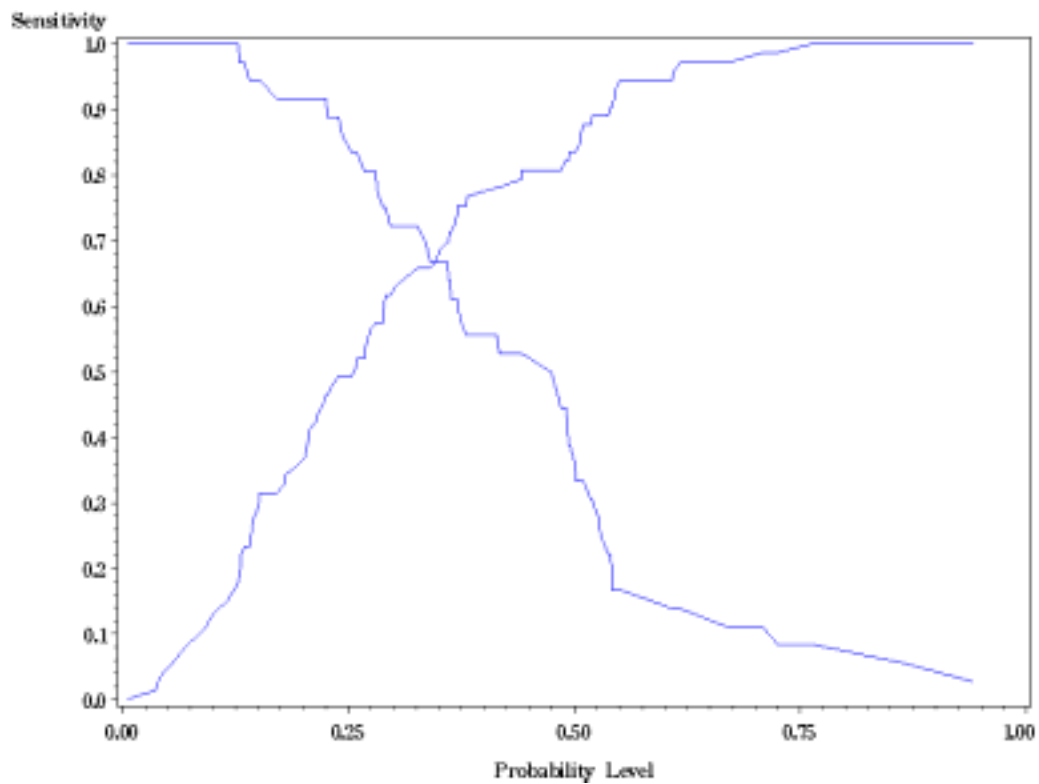


Figure 3: Sensitivity and Specificity

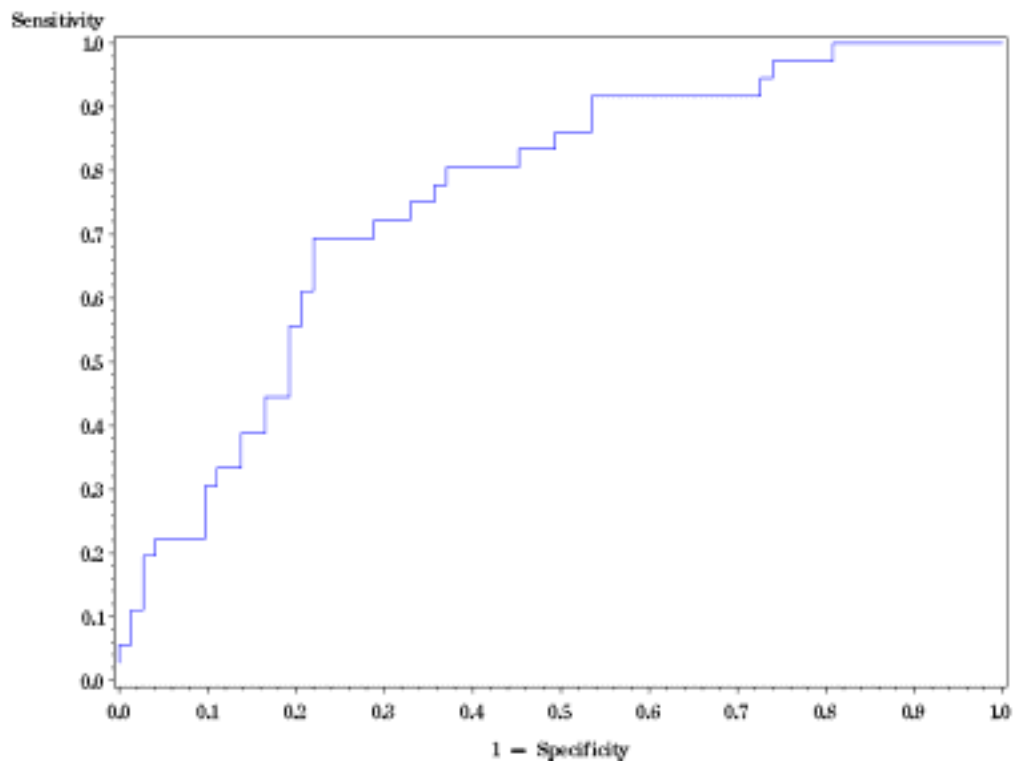


Figure 4: ROC Curve

As a general rule,

If $ROC=0.5$: this suggests no discrimination

If $0.7 < ROC < 0.8$, this is considered acceptable discrimination

If $0.8 < ROC < 0.9$, this is considered excellence discrimination

If $ROC > 0.9$, this is considered outstanding discrimination.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	73.4	Somers' D	0.471
Percent Discordant	26.3	Gamma	0.473
Percent Tied	0.3	Tau-a	0.21
Pairs	2628	C	0.736

Table 12: ROC Value

The Area under the curve is 0.736, which is considered acceptable discrimination.

Cross Validation

Five fold Cross-Validation is used to examine the accuracy of the ROC found in the final model. The original 109 patients' data has been partitioning into five sub-data. Of the five sub-data, a single sub-data is retained as the validation data for testing the model, and the remaining four are used as training data. The same main effects and interactions terms used to find the final model are adopted here for the training data. A new model is created and substituted into the retained testing data. A final ROC value is calculated for this testing data. The cross-validation process repeated five times with each of those sub-data used exactly once as testing data. The five ROC results are averaged and compare with the final model in order to calculate the shrinkage. Randomization plays an important role in this cross-

validation procedure. The original data is randomly assigned to those five sub-data. Proc Plan is used in SAS to accommodate this process.

	Training Data	Subdata
	0.814	0.718
	0.735	0.667
	0.747	0.688
	0.766	0.556
	0.75	0.726
Average	0.7624	0.671
Shrinkage	0.7624-0.671=0.0914	

Table 13: Example of Cross Validation

Repeat the process five times; the average shrinkage of AUC is 0.12.

Survival Analysis

Survival analysis has been conducted in order to find out the cumulative proportion of the patients who did not have the recurrence of fibroids. The Kaplan-Meier estimator (also known as product limit estimator) is used. It estimates the survival function for life-time data. In medical research, it might be used to measure the fraction of patients living for a certain amount of time after surgery. An economist might measure the length of time people remain unemployed after a job loss. An engineer might measure the time until failure of machine parts.

An important advantage of the Kaplan-Meier curve is that the method can take into account "censored" data — losses from the sample before the final outcome is observed (for instance, if a patient withdraws from a study). On the plot, small vertical tick-marks indicate losses, where patient data has been censored. When no truncation or censoring occurs, the Kaplan-Meier curve is equivalent to the empirical distribution.

A plot of the Kaplan-Meier estimate of the survival function is a series of horizontal steps of declining magnitude which, when a large enough sample is taken, approaches the true survival function for that population. The value of the survival function between successive distinct sampled observations is assumed to be constant.

Let $S(t)$ be the probability that an item from a given population will have a lifetime exceeding t . For a sample from this population of size N let the observed times until death of N sample members be

$$t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_N.$$

Corresponding to each t_i is n_i , the number "at risk" just prior to time t_i , and d_i , the number of deaths at time t_i .

The nonparametric maximum likelihood estimate of $S(t)$ is

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

When there is no censoring, n_i is just the number of survivors just prior to time t_i . With censoring, n_i is the number of survivors less the number of losses (censored cases). It is only those surviving cases that are still being observed (have not yet been censored) that are "at risk" of an (observed) death

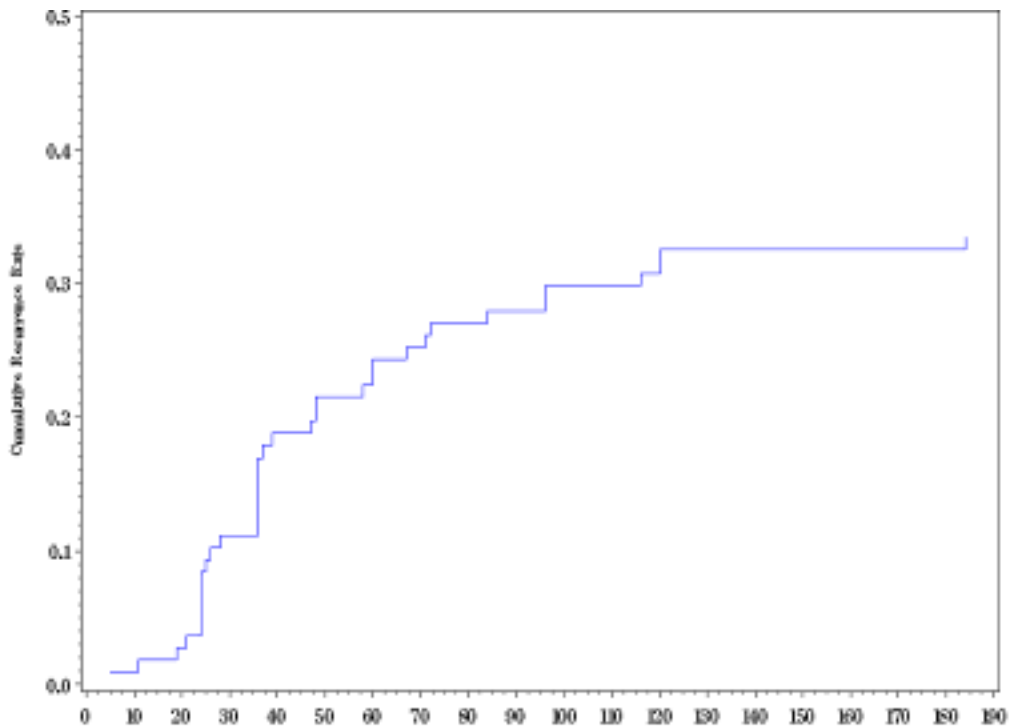


Figure 5: Kaplan-Meier Cumulative Recurrence Rate

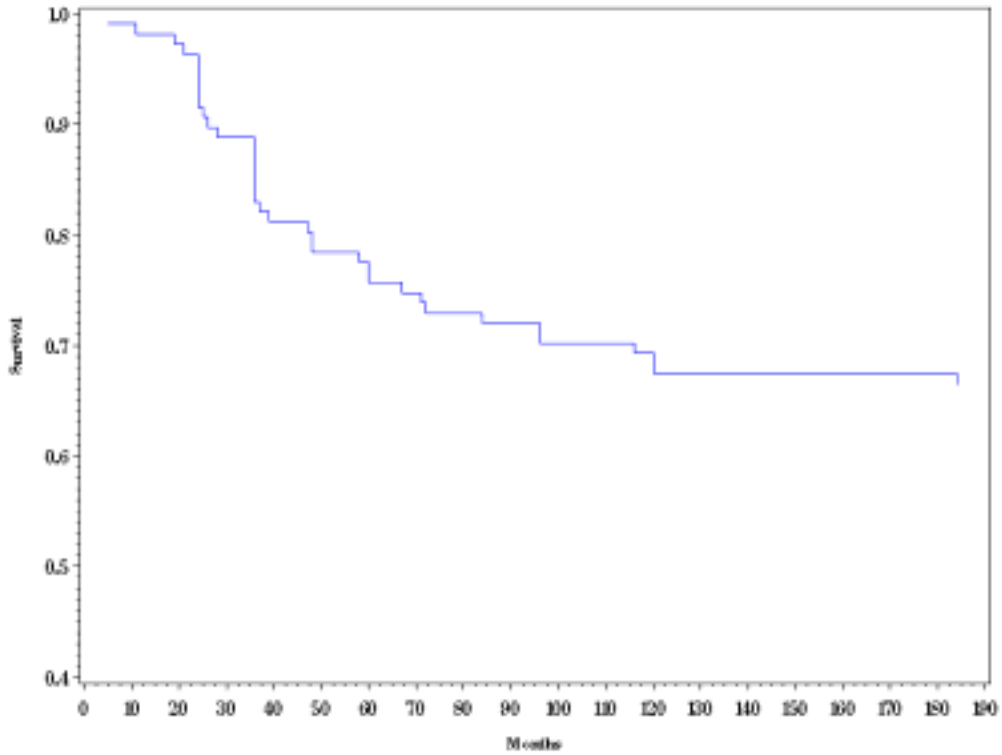


Figure 6: Kaplan-Meier Survival Rate

Figure 6 shows the survival curve. It seems that the recurrence rate is much higher for the first 4 years comparing with the next 10 years. This could be an important information for medical doctors who give advises to patients.

Chapter 3

RESULTS

Odds Ratio

After check the goodness fit of the model, the consideration of the interpretation of logistic regression coefficients can be made, especially with the independent variable is dichotomous. Odds ratio is a measure of impacts of independent variables which has been widely used, especially in medical field. The odds are a way representing probabilities, it approximates how much more likely (or unlikely) it is for the outcome to be present among those levels of dichotomous variables. In the study, 'Pelvic pain' is a dichotomous variable and it is significant in the final model. The odds ratio for 'Pelvic pain' is 2.478, and the 95% confidence interval is (0.873, 7.40). The interpretation of this could be that the patients who experience the symptom 'Pelvic pain' are approximately 2.5 times of having recurrence to those patients who don't experience the symptoms. However, the confidence interval contains one, the odds ration 2.5 may not be that significant.

From the survival study, the 5 year cumulative recurrence rate is 24.32% and the 10 year cumulative recurrence rate is 32.57%. Factors are significant in the final model are 'Age of surgery', 'Tumor size', 'Pelvic pain' and the interaction between the 'Age of surgery' and 'Tumor size'.

Chapter 4

DISCUSSION

The result from the cross validation has a very big shrinkage in average and this may be caused by the small sized dataset. Partial Least Square method might be an option of this Logistic regression model and will be conducted in a future study.

A patient with symptomatic uterine myoma wants to know the chances of recurrence of myoma after Myomectomy and the factors that may affect recurrence if she chooses this surgery as her treatment option. Recent literature reviews have suggested that the variation in length of follow-up and criteria for recurrence make it difficult to compare the results of previous studies.

Recurrence of myoma after Myomectomy has been reported based on diagnosis by palpation, by systematic transvaginal ultrasonography, or by the need for a second surgery. The current study used transvaginal ultrasonography to diagnose recurrence, counting only tumors that were at least 2cm in diameter. The crude recurrence rate used in other studies represents the minimum myoma recurrence rate and will underestimate the true recurrence rate, which was calculated by method of Kaplan-Meier.

This risk should not rule out the use of Myomectomy. The recurrence occurs after quite a long time after the surgery; patients could use this time to have children. The risk of recurrence is linked with the persistence of small nuclei within the myometrium, which are undetectable during the operation. It would be an advantage to know what the growth factors are, and how to use those factors to identify groups at high risk of recurrence who could then benefit from more suitable treatment strategies or drug treatment to prevent recurrence.

REFERENCES

1. Lynn M. Marshall, Donna Spiegelman, JoAnn E.manson, Marlene B.Goldman, Robert L.Barbieri, Meir J.Stampfer, Walter C.Willett, David J. Hunter (1998) *Epieimiology*. Risk of Uterine Leiomyomata among Premenopausal Women in Relation to Body Size and Cigarette Smoking, 511:517
2. Donna Day Baird and David B. Dunson (2003) *Hypotheses*. Why is Parity Protective for uterine Fibroids? 247:249
3. Magdi Hanafi, M.D. (2005) *American College of Obstetricians and Gynecologists*. Factors Affecting Recurrence of Leiomyoma after Myomectomy, 879:881
4. Arnaud Fauconnier, Charles Chapron, Katayoun Babaki-Fard, Jean B.Dubuisson (2000) *Human Reproduction*. Recurrence of Leiomyomata after Myomectomy 595:602
5. B.Candiani, L.Fedele, F.Parazzini, L.Villa (1991) *The Journal of Obstetrics and Gynecology*. Risk of Recurrence after Myomectomy 385:389
6. Michael Roemisch M.D., Ceana H.Nezhat, M.D.,DanielSeidman, M.D. (1998) *The Journal of the American Association of Gynecologic laparoscopists*. Recurrence Rate after Laparoscopic Myomectomy 237:239
7. Henri Marret, Jean Cottier, Ana Alonso, Bruno giraudeau, Gilles Body (2005) *An international Journal of Obsterics and Gynaecology*. Predictive factors for fibroids recurrence after Uterine Artery Embolisation 461:465

8. Lynn M. Marshall, Donna Spiegelman, Marlene Goldman, Joann Manson, David Hunter, Meir Stamper (1998) American Society for Reproductive Medicine. A prospective study of reproductive factors and oral contraceptive use in relation to the risk of uterine leiomyomata 432: 439
9. Fabio Parazzini, Carlo Vecchia, Eva Negri, Gabriela Cecchetti, Luigi Fedele (1988) Epidemiologic Characteristics of Women with Uterine Fibroids: A Case-Control Study. American College of Obstetricians and Gynecologists 853:857
10. Hosmer, DW., Lemeshow, S., (2000) Applied Logistic Regression 2nd Edition.
11. Marchelle Sanchez, (2006) Analyzing the effects of adolescent risky behaviors on suicidal ideation.
12. Uterine artery Embolization (UAE). Cited 2007, June 7.
URL: www.fibroidworld.com/UAE.htm
13. WebMD Better information better health. Cited 2007, June 26.
URL: www.webmd.com/a-to-z-guides/Myomectomy-11717
14. Micheal E. Toaff, M.D., MSC, Alternatives to Hysterectomy Assessed 2007 June 25.
URL: www.Netreach.net/~hysterectomyedy/myomecto.htm

SAS CODE

```

*****Rename*****;
data lu.finaldata2;
set lu.finaldatafull;
rename ID=Var1
       recurrence=Y
       height=var2
       weight=var3
       BMI=var4
       age_1st_myo=var5
       preg_before=var6
       delivery=var7
       number_of_leiomyoma=var8
       tumor_size=var9
       single_tumor=var10
       willing_to_conceive=var11
       pregnancy_after_myomectomy=var12
       heavy=Var13
       irregular=Var14
       dyspar=Var15
       dysmen=var16
       pelvic_pain=var17
       ;
run;

*****Discretize age, tumor-size , number of tumor, BMI*****;
data lu.discretedata;
set lu.finaldata;
if Age_1st_myo < 30 then Agegroup1=1; else Agegroup1=0;
if Age_1st_myo > 30 and Age_1st_myo < 40 then Agegroup2=1; else
Agegroup2=0;
if BMI < 24.9 then BMIgroup1=1; else bmgrou1=0;
if BMI > 24.9 and BMI < 29.9 then BMIgroup2=1; else bmgrou2=0;
if BMI > 30 then BMIgroup3=1; else bmgrou3=0;
if tumor_size > 1.9 and tumor_size < 4 then sizegroup1=1; else
sizegroup1=0;
if tumor_size > 4 and tumor_size < 6 then sizegroup2=1; else sizegroup2=0;
if number_of_leiomyoma < 4.1 then numbergroup=1; else numbergroup=0;
run;

*****logistic regression model*****;

Proc logistic data=lu.finaldata2 descending ;
model y=var4-var17/selection=stepwise sle=0.4 sls=0.3 lackfit;
run;

Univariate Analysis;

%macro logitreg(label,tvarlist);
ODS OUTPUT Logistic.GlobalTests=testout;
proc logistic descending data=lu.myoma2;
model y= &tvarlist;
run;

```

```

data one;
label="&label";
run;

data testout;
merge one testout(obs=1);
run;

data report;
set report testout;
run;
%mend;

%logitreg(var4,var4);

*****Run var2 first!!!!*****;
Data total;
Set testout;
Run;

%macro lu ;
%do i=3 %to 17;
%logitreg (var&I, var&i);
Data total;
Set total testout;
Run;
%end;
%mend;

%lu;
*****correlations*****;

proc corr data=lu.finaldata outp=lu.Corr noprint;
var recurrence BMI age_1st_myo preg_before delivery number_of_leiomyoma
tumor_size single_tumor
willing_to_conceive pregnancy_after_myomectomy heavy irregular dyspar
dysmen pelvic_pain;
run;
***** Included interaction with BMI, PB*****;
Proc logistic data=lu.finaldata2 descending ;
model y=var4-var17 var5*var9 var5*var17 var9*var17 var4*var5 var4*var9
var4*var17 var6*var5 var6*var9 var6*var17
var8*var5 var8*var9 var8*var17/selection=stepwise sle=0.3 sls=0.2;
run;

*****Variable Transformation*****;

data one;
set lu.finaldata1;
p1=-2;p2=-1;p3=-0.5;p4=0;p5=0.5;p6=1;p7=2;p8=3;
run;
%macro fractional(var);
data out1factor out2factor;
if 1=1 then delete;
run;

```

```

%do n=1 %to 8;
data sample&n;
    set one;
    if &var=0 then u1=0;
    else do;
        if p&n=0 then u1=log(&var);else u1=&var**p&n;
    end;
run;
proc logistic descending data=sample&n outest=out&n;
    model y=u1;
run;
data out&n;
    merge out&n one(obs=1);
    f1=p&n;D=-2*_lnlike_;
    keep f1 D;
run;
data out1factor;
    set out1factor out&n;
run;
%do m=1 %to 8;
data sample&n&m;
    set one;
    /* if p&n=0 then u1=log(var65);else u1=var65**p&n;
    if p&m=p&n then u2=log(var65)*u1;else if p&m=0 then
u2=log(var65);else u2=var65**p&m;*/
    if &var=0 then u1=0;
    else do;
        if p&n=0 then u1=log(&var);else u1=&var**p&n;
    end;

    if &var^=0 then do;
        if p&m=p&n then u2=log(&var)*u1;
        else if p&m=0 then u2=log(&var);else u2=&var**p&m;
    end;
    else u2=0;
run;
proc logistic descending data=sample&n&m outest=out&n&m;
    model y=u1 u2;
run;
data out&n&m;
    merge out&n&m one(obs=1);
    f1=p&n;f2=p&m;D=-2*_lnlike_;
    keep f1 f2 D;
run;
data out2factor;
    set out2factor out&n&m;
run;
%end;
%end;
%mend fractional;
%fractional(var11);

proc iml;
    use out1factor;
    read all var{f1 D} into modell;
    close out1factor;
    L1=modell[1,2]; P1=modell[1,1];

```



```

nrow1=nrow(model1);
do i=2 to nrow1;
    if model1[i,1]=1 then L=model1[i,2];
    if model1[i,2]<L1 then do;
        L1=model1[i,2];P1=model1[i,1];
    end;
end;
use out2factor;
read all var{f1 f2 D} into model2;
close out2factor;
L2=model2[1,3]; P21=model2[1,1];P22=model2[1,2];
nrow2=nrow(model2);
do i=2 to nrow2;
    if model2[i,3]<L2 then do;
        L2=model2[i,3];P21=model2[i,1];P22=model2[i,2];
    end;
end;

if L-L1>3.84 then do; /*chi-sq(.95,1)=3.84; chi-sq(.95,2)=5.99;
chi-sq(.95,3)=7.81;*/
    if L1-L2>5.99 then print 'the best is model2 with power'
P21 'and' P22;
    else print 'the best is model1 with power' P1;
end;
else do;
if L-L2>7.81 then print 'the best is model2 with power' P21
'and' P22;
    else print 'the best is linear model' ;
end;
quit;

*****CROSS VALIDATION*****;

data a;
    set lu.myoma2(keep=var1);run;

proc plan seed=30000;
    factors var1=109;
    output data=a out=b;
run;
data b;
    set b;
    seq=_n_;
    if seq<=22 then group=1;
    else if seq>22 and seq<=44 then group=2;
    else if seq>44 and seq<=66 then group=3;
    else if seq>66 and seq<=88 then group=4;
    else group=5;
run;
proc sort data=b;by var1;run;

data lu.trial6;
    merge lu.myoma2 b;
    by var1;
run;

```

```

proc sort data=lu.trial6;
by group;run;

data group1 group2 group3 group4 group5;
set lu.trial6;
if group=1 then output group1;
else if group=2 then output group2;
else if group=3 then output group3;
else if group=4 then output group4;
else output group5;
run;

data group1234;
set group1 group2 group3 group4;
run;

data group1235;
set group1 group2 group3 group5;
run;

data group1245;
set group1 group2 group4 group5;
run;

data group1345;
set group1 group3 group4 group5;
run;

data group2345;
set group2 group3 group4 group5;
run;
*** For Group 5***;
Proc logistic data=group1234 descending ;
model y=var4-var17 var5*var9 var5*var17 var9*var17 var4*var5 var4*var9
var4*var17 var6*var5 var6*var9 var6*var17
var8*var5 var8*var9 var8*var17/selection=stepwise sle=0.15 sls=0.15;
run;

data group5m;
set group5;
A=-0.1665*var5-0.4706*var9+0.8982*var17;
run;

proc logistic data=group5m descending;
model y=A;
run;

*****for group4*****;
proc logistic data=group1235 desc;
model y=var4-var17 var5*var9 var5*var17 var9*var17 var4*var5 var4*var9
var4*var17 var6*var5 var6*var9 var6*var17
var8*var5 var8*var9 var8*var17/selection=stepwise sle=0.15 sls=0.15;
run;
data group4m;
set group4;
A=-0.0921*var5-0.2866*var9;
run;

```

```

proc logistic data=group4m descending;
model y=A;
run;

*****For group 3*****;
proc logistic data=group1245 desc;
model y=var4-var17 var5*var9 var5*var17 var9*var17 var4*var5 var4*var9
var4*var17 var6*var5 var6*var9 var6*var17
var8*var5 var8*var9 var8*var17/selection=stepwise sle=0.15 sls=0.15;
run;
data group3m;
set group3;
A=-0.1526*var5-0.4979*var9+0.9995*var16+1.8035*var17;
run;

proc logistic data=group3m descending;
model y=A;
run;

*****For group 2*****;
proc logistic data=group1345 desc;
model y=var4-var17 var5*var9 var5*var17 var9*var17 var4*var5 var4*var9
var4*var17 var6*var5 var6*var9 var6*var17
var8*var5 var8*var9 var8*var17/selection=stepwise sle=0.15 sls=0.15;
run;

data group2m;
set group2;
A=-0.3335*var5-1.899*var9+0.2402*var6+0.0464*var5*var9;
run;

proc logistic data=group2m descending;
model y=A;
run;

*****For group 1*****;
proc logistic data=group2345 desc;
model y=var4-var17 var5*var9 var5*var17 var9*var17 var4*var5 var4*var9
var4*var17 var6*var5 var6*var9 var6*var17
var8*var5 var8*var9 var8*var17/selection=stepwise sle=0.15 sls=0.15;
run;

data group1m;
set group1;
A=-0.3798*var5-2.4964*var9+1.0510*var11+5.6351*var17+0.0644*var5*var9-
1.0496*var9*var17;
run;

proc logistic data=group1m descending;
model y=A;
run;

```

```

*****ROC*****;
Ods listing close;
ODS output Logistic.Classification=classification;
Proc logistic data=lu.myoma2 descending ;
model y=var4-var17 var5*var9 var5*var17 var9*var17 var4*var5 var4*var9
var4*var17 var6*var5 var6*var9 var6*var17
var8*var5 var8*var9 var8*var17/selection=stepwise sle=0.15 sls=0.15 ctable
outroc=fit details pprob= (0.05 to 0.6 by 0.05);
run;
ods trace off;

data fit1;
set fit;
if _step_ ^=6 then delete;
run;

proc gplot data=fit1;
plot _sensit_*_lmspec_;
run;

ODS listing close;
ODS output Logistic.parameterestimates=logisticmodelinfo;
Proc logistic data=lu.myoma2 descending ;
model y=var4-var17 var5*var9 var5*var17 var9*var17 var4*var5 var4*var9
var4*var17 var6*var5 var6*var9 var6*var17
var8*var5 var8*var9 var8*var17/selection=stepwise sle=0.15 sls=0.15;
run;
data roc3;
set roc1;
if _step_ ^=6 then delete;
run;

data roc2;
set roc3;
spec=1-_lmspec_;
run;
symbol1 i=join v=none;
proc gplot data=roc2;
plot _sensit_*_prob_=1 spec*_prob_=1 / overlay haxis=0 to 1 by 0.25
vaxis=0 to 1 by 0.1;
run;

*****SURVIVAL*****;

data lu.survival1;
set lu.survival;
Duraition=INTCK('month',first_surgery_date,date);
run;

data lu.kaplan1;
set lu.kaplan;
retain stemp 1;
stemp=stemp*(1-recurrence/atrisk);
survival=stemp;
CRR=1-stemp;

```

```
run;

goptions reset=global
        rotate=landscape
        gunit=pct
        htext=2.5
        ftext=centb
        gsfmode=replace
        device=pscolor
        gaccess=gsasfile;

symbol1 color=blue i=step1j line=1 w=3;
axis1 label=(a=90 h=2 c=black"Survival") order=(0.4 to 1 by 0.1);
axis2 label=(h=2 c=black"Months") order=(0 to 190 by 10);
proc gplot data=lu.kaplan1;
plot survival*duration/vaxis=axis1 haxis=axis2;
run;

*****;
```