Dialysis:

Long-Term Pattern Analysis

NATIONAL HEALTH INSURANCE RESEARCH DATABASE (LHID2000)

Po-Yi Du, Yung-Chun Lee, Po-Yang Yeh, Guan-Lun Jhang
Instructor: Dr. Pei-Fang Su
NATIONAL CHENG KUNG UNIVERSITY | DEPARTMENT OF STATISTICS

Abstract: This research is about 5-year pattern of individuals before undergoing dialysis as treatment for renal failure. By using Taiwan's National Health Insurance Research Date – LHID2000 to study the 5-year medical records of patients undergoing dialysis as treatment within 2002 to 2011, and by applying LASSO Logistic Regression to build a predicting model for finding out the certain kinds of factor combination that would enlarge the probability of suffering from renal failure.

Key words: LHID2000, renal failure, dialysis, LASSO Logistic Regression

1 Introduction

1-1 Motivation

Taiwan is one of the countries in the world with the highest rate of renal failure. Dialysis, a widely used treatment for renal failure, costs the Bureau of National Health Insurance of Taiwan NTD 30 billions a year. [Taiwan Healthcare Reform Foundation, 2015] The medical expenditure has brought a great amount of financial deficit into Taiwan's Health Insurance for years. Although Taiwan is famous for its well-planned health insurance system, those debts, if the issue left unsolved, would finally devastate the entire system. [Chang, 2014]

And also, among the island, the Southern Taiwan seems to have the worst condition. It is very often to hear such sentences as "People in southern Taiwan do the dialysis most often" or "People in some certain groups of age have the higher likelihood of having renal failure".

As a result, to prevent the National Health Insurance System from collapsing, and to find out whether some factors (e.g. age, place of living) or long-term patterns (e.g. drug usage) would somehow cause the renal failure, the research began.

Lastly, we hope that the result of this research can fix this difficult position and make Taiwan a better place.

1-2 Dialysis and Renal Failure

When a patient develops a renal failure, especially at the end stage (ESRD, by the time that someone lose about 85 to 90 percent of the kidney function), he or she are at high risk for chronic hepatitis, liver cirrhosis, and mortality than the general population. Thus, he or she would need to undergo dialysis as a treatment to maintain his or her health. Patients with end-stage renal dialysis (ESRD) are eligible for any type of renal replacement therapy for free of charge; the expenses of chronic dialysis patients are covered by National Health Insurance (NHI). [Chien et al., 2012]

Dialysis functions exactly the same way as the kidney, removing waste, salt and extra water to prevent them from building up in the body. Also, dialysis keeps a safe level of certain chemicals in patient's blood, and help controlling blood pressure. [National Kidney Foundation]

There are two kinds of dialysis, hemodialysis and peritoneal dialysis. In this research, both treatments would be seen as the same treatment, namely "Dialysis".

1-3 Definition of Areas

For not being too trivial or pointless, some parts of the research will be using the 6 areas defined by Bureau of Health Insurance rather than the original regions (except for the specific regions of interested.) The medical areas of Taiwan are divided into following 6 parts:

I. Taipei Medical Area (Shown as purple in Figure 1.2): Keelung City, Taipei City, Taipei County, Yilan County, and Kinmen County

II. Northern Medical Area (Shown as green in Figure 1.2): Taoyuan County, Hsinchu City, Hsinchu County, and Miaoli County

III. Central Medical Area (Shown as blue in Figure 1.2): Taichung City, Taichung County, Nantou County, and Changhua County

IV. Southern Medical Area (Shown as yellow in Figure 1.2): Yunlin County, Chiayi City Tainan City, and Tainan County

V. Kaohsiung and Pingtung Medical Area (Shown as red in Figure 1.2): Kaohsiung City, Kaohsiung County, Pingtung County, and Penghu County

VI. Eastern Medical Area (Shown as bright blue in Figure 1.2): Hualien County, and Taitung County

[Bureau of Health Insurance, Executive Yuan of ROC, 2010]

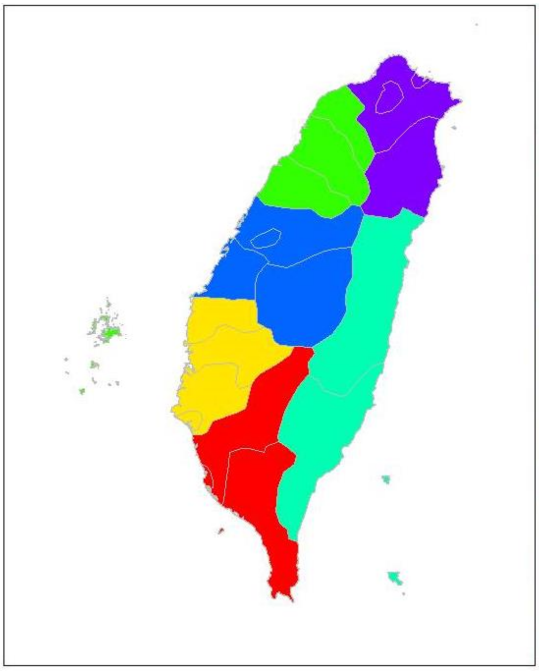

Figure 1.2

1-4 Definition of Age Group

In this research, the groups of age are defined as:

I. "Before middle age" for under 45 years old,

II. "Middle age" for 45 to 65 years old,

III. "Old age" for beyond 65 years old.

[NTNU Holistic Education Program]

The definition is based on both mental and physical considerations. The distribution of three groups of age is shown as Figure 1.3. The group “Under Middle Age” contains most of the population, followed by “Middle Age”, and then “Old Age”.

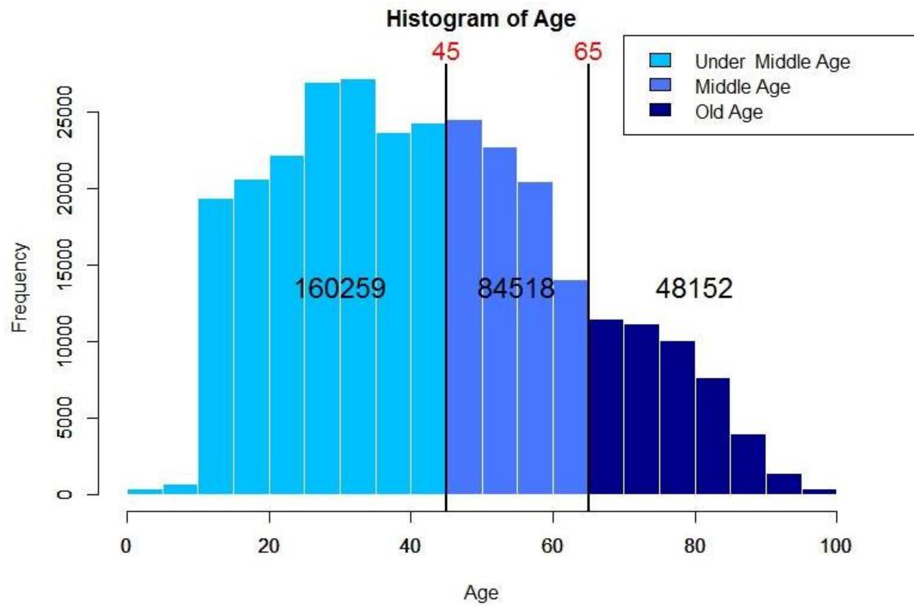

Figure 1.3

2 Methodology

2-1 Source of Data

The National Health Insurance (NHI) program has provided compulsory universal health insurance in Taiwan since 1995. LHID2000, the data for the study, is a dataset obtained from the National Health Insurance Research Database (NHIRD), containing 300 thousands individuals participating in the NHI program. [Chien et al., 2012] The dataset Contains registration files and original claim data for reimbursement from, displayed in many sub-datasets (e.g. ID, CD, DD, OO, HOSB, DO, GO, etc.). [National Health Insurance Research Database, 2010]

This study will be using 4 sub-dataset (ID, CD, DD, and HOSB, details are shown in table 1.1) of LHID2000. Those sub-datasets contain only the records from 1996 to 2011, occupying about 18 Gigabytes of storage.

Sub-dataset	Description	# of variables
ID	*ID* contains the list of personal health insurance information of individuals (300 thousands) that are sampled from the people who joined in the National Health Insurance.	13
CD	*CD* is about the clinical case types, treatments, payments, prescriptions.	37
DD	*DD* is the sub-dataset of payments details while hospitalizing.	70
HOSB	*HOSB* contains the basic information of the health care facilities (including hospitals and clinics).	28

Table 1.1

2-1.1 Extracting Data

Working with The Data without manipulating them beforehand is obviously unworkable. Thus, manipulating The Data into a usable form is definitely needed.

Firstly, the interaction of two specific variables in the CD sub-dataset – case_type and cure_item – was defined to be the indicator of whether an individual suffered from renal failure. While case_type equaling to '05' (case of renal failure) and cure_item equaling to 'D8' (undergoing hemodialysis as treatment) or 'D9' (undergoing peritoneal dialysis treatment), the individual then would be judged as a patient suffering from renal failure and under dialysis treatment.

Secondly, to have a better view of the research topic – long-term pattern analysis –, not only the records of the undergoing dialysis but also the records before undergoing dialysis are needed. Moreover, not every records of patients undergoing dialysis were needed but the first time of undergoing dialysis. Since the research topic was about the long-term pattern, the "causes" of renal failure (or undergoing dialysis as treatment) were the particular things in which were interested. Thus, all of the records would be eliminated except for the first records of individuals (renal failure patients), and then the "first records of undergoing dialysis of individuals" would be created.

Thirdly, to compare every individual on the same basis, the "first records of undergoing dialysis of individual" would be kept as the "trimmed records" only if the date of the record was within Jan 1, 2002 and Dec 31, 2011. Furthermore, the ID of the "trimmed records" would be used to tracing back the medical records within 5 years before an individual start undergoing dialysis as a treatment for renal failure.

Therefore, the records of individuals who were suffered from renal failure and were undergoing dialysis as the treatment, and the records BEFORE individuals undergoing dialysis would be found.

Besides the patient who suffered from renal failure, the latest – 5 – year records of individuals who were healthy enough to avoid suffering from renal failure were collected.

The concept is demonstrated as Figure 1.1:

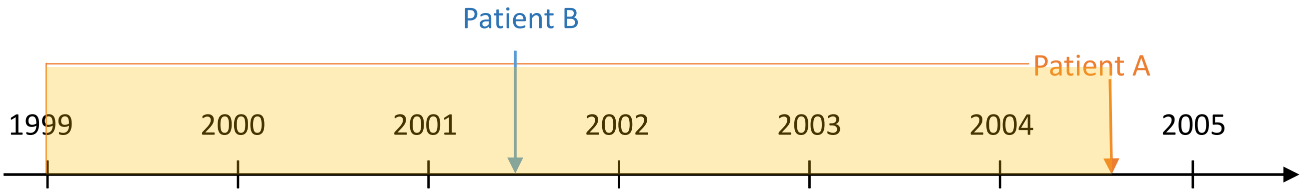

Figure 1.1

Patient A started undergoing dialysis on July 24, 2004, the records of patient A **BETWEEN Jan 1, 1999 AND July 24, 2004** are going to be taken into consideration in this research.

Patient B started undergoing dialysis on June 5, 2001, which was before Jan 1, 2002, so **EVERY RECORDS of patient B** will **NOT** be taken into consideration in this research.

2-1.2 Data Structure

After manipulating the sub-datasets, two data frames would be acquired. One is the clinical records of individual within 5 years, and the other would be the former one compacted into 1 record per ID.

Those data frames would contains the following variables:

id_sex

id_sex represents the individual's gender, denoted 'F' as female and 'M' as male.

H

H is a binary variable indicating that whether an individual suffered from renal failure and undergoing dialysis as treatment. '1' as yes, and '0' as no.

age

age is a continuous variable, represents the individual's age.

AREA_NO_H

AREA_NO_H is a categorical variable, represents the region (in a coded form) of the health care facilities that individuals were at in the last record.

avg_dg_amt

avg_dg_amt is a continuous variable, represents the individual's average annual payment of prescription for the last 5 years.

avg_dg_d

avg_dg_d is a continuous variable, represents the individual's average annual days if prescription (in other words, average days of taking drug annually).

trt1~trt97

Each *trt# (# from 1 to 97)* is a binary variable. It represents that whether an individual had undergone specific treatment within the latest 5 year. The number (#) of the treatment (trt) works as a coded name of each treatment. For example, *trt1* refer to the treatment of *diabetes.* Once an individual has an '1' in *trt1* variable, the individual could be conclude accordingly as a patient of *diabetes*

.

2-2 LASSO Logistic Regression

LASSO Logistic Regression is used to fit the model and select the operator at the same time when the saturated model has too many parameters to use the ordinary least squares regression method (which will fail to use backward elimination for lacking of degrees of freedom).

While the ordinary Least Square (OLS) regression method finds the unbiased linear combination of the x_ij's that minimizes the residual sum of squares, LASSO does it by minimizing the residual sum of squares subject to the sum of absolute values of the coefficient being less than a constant. By doing so, LASSO tends to produce some coefficients to be exactly zero with a little bias sacrificed to reduce the variance of the predicted values and improve the overall prediction accuracy. [Zhao, 2008]

3 Result

3-1 Summary Statistics

The sex ratio shows that of all the individuals suffered from renal failure, male patients occupies larger proportion than female patients **(male/female=1.04)**. However, when it comes to the sex ratio of number of case records, the result is reversed **(# of male's record/ # of female's record =0.85)**.

When comparing the condition of suffering from renal failure between regions, the frequency table (Table 3.1) shows the distribution of the individuals undergoing dialysis as treatment during 2002 to 2011. As presented, big cities such as Taipei City Taipei County, Kaohsiung City, and Taichung County have higher frequencies than other regions; on the other hand, suburban regions like Penghu County, Hualien County, Taitung County, and Kinmen County have lower frequency (only single digits) than other regions.

Region	**Taipei City**	**Kaohsiung City**	**Keelung City**	**Hsinchu County**	**Taichung City**	**Tainan City**	**Chiayi City**	**Taipei County**	**Taoyuan County**	**Hsinchu City**	**Yilan County**	**Miaoli County**
Times	**133**	**70**	19	12	41	49	27	**85**	66	12	17	19
Region	**Taichung County**	**Changhua County**	**Nantou County**	**Yunlin County**	**Chiayi County**	**Tainan County**	**Kaohsiung County**	**Pingtung County**	**Penghu County**	**Hualien County**	**Taitung County**	**Kinmen County**
Times	**60**	37	28	22	12	58	51	46	**1**	**9**	**8**	**1**

Table 3.1

This phenomenon may be due to the population scales of regions; therefore, the comparison of proportion are presented as Figure 3.1. It can be found that the color changes from white to blue as the region goes Southerner. This fact means that there were more proportion of individuals undergoing dialysis as treatment in Southern Taiwan. Therefore, it could be included that there were more people suffered from renal failure or something which could injure the kidneys.

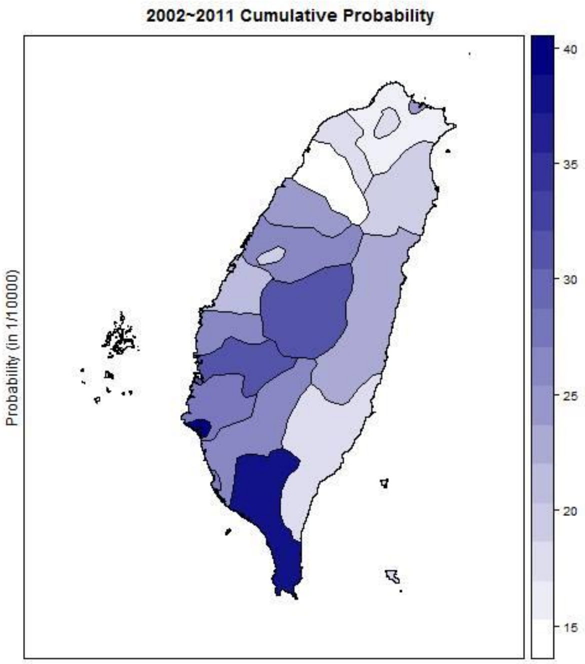

Figure 3.1

To look closer, as presented in Figure 3.2, the southern Taiwan generally displays darker color (or the color that is closer to red, which infer to a higher proportion) when separated by years. This fact may imply that the southern Taiwan suffered worse than other region from undergoing dialysis as treatment of renal failure.

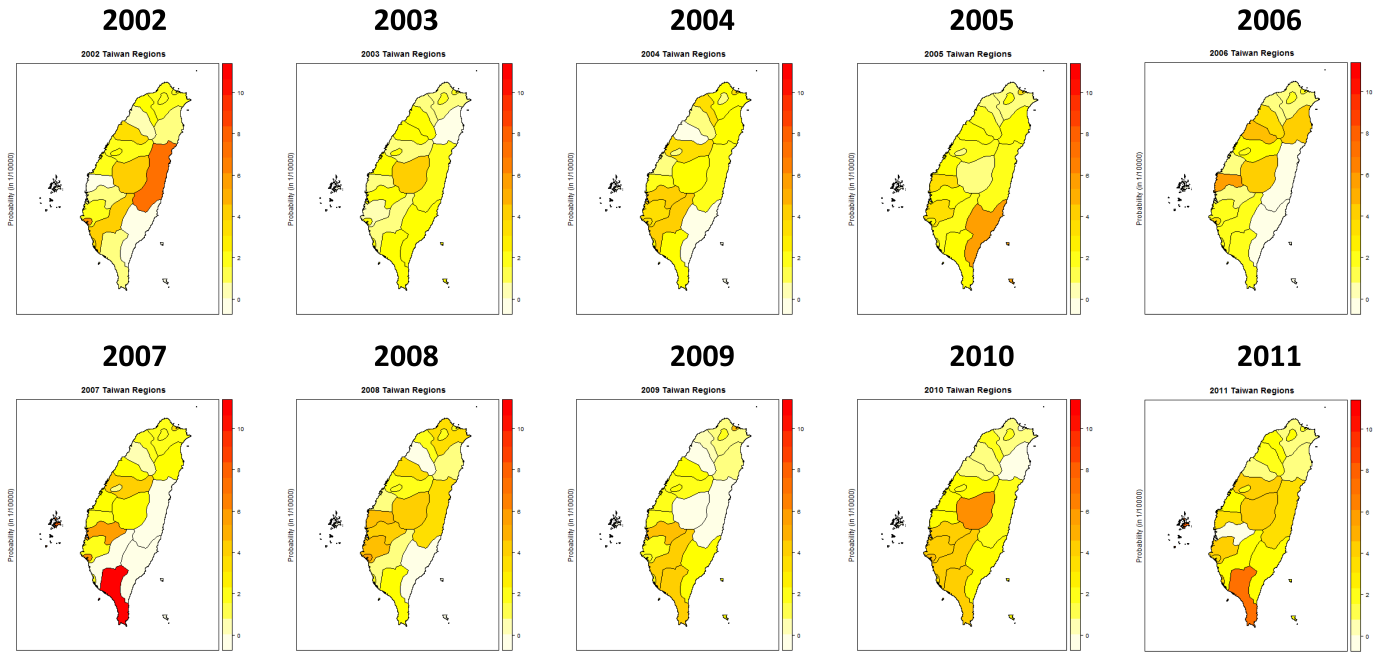

Figure 3.2

If summing the result of Figure 3.2 by year, shown in Figure 3.3 as below, it can be found that there is no obvious difference between years (between 0.2 per mil and 0.4 per mil).

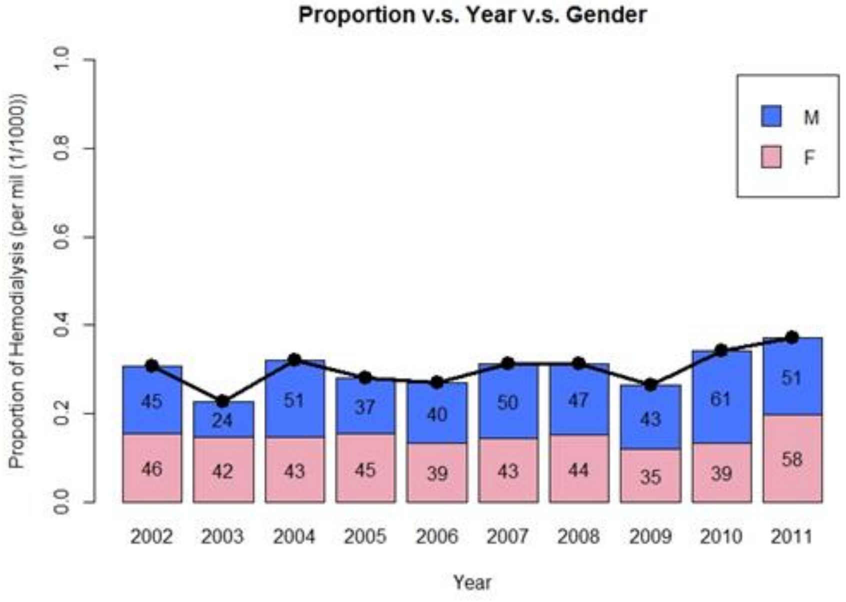

Figure 3.3

*The denominator of the proportion is the total population (different ID) that in the data.

*Note that the unit of y-axis is “per mil”.

As for ages, it seems that the distribution of ages of newly patients undergoing dialysis is quite normal, shown in Table 3.1 and Figure 3.4. The mean is 61.7, which is very close to the median age 63, and minimum age is 8. Note that the kurtosis in the chart has already subtract by 3.

n	Mean	sd	median	min	max	Skew	Kurtosis	Variance
883	61.69	14.6	63	8	98	-0.46	0	213.16

Table 3.2

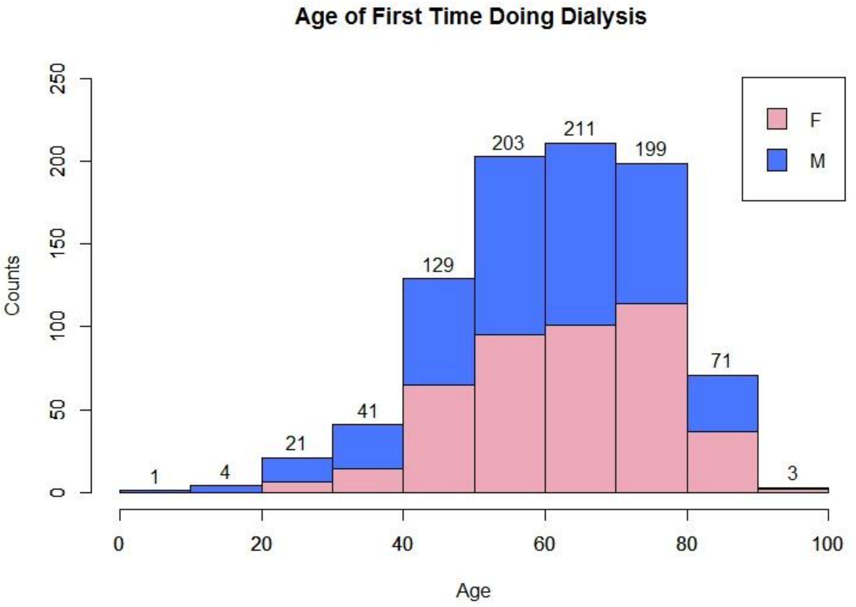

Figure 3.4

When it comes to the prescription days and the previous treatments that individuals have experienced before undergoing dialysis, as shown in Figure 3.5, the drug day “3” has the highest frequency, follow by “0” and “28, and then are “7”, “30”, and “14”. One interesting pattern is that, expect for “3” (the most common prescription day in Taiwan) and “0” (treatments that do not need prescriptions of drug such as physical treatment), “7”, “14”, “28” are the multiples of length of a week (7 days), and that “30” is exactly the length of a month (30 day).

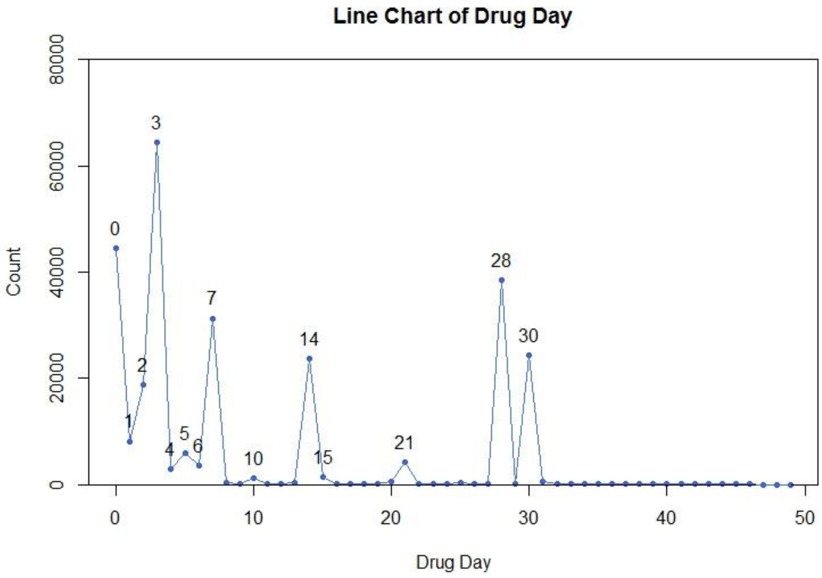

Figure 3.5

The other interesting pattern is the extremely unbalanced distribution of previous treatments, as shown in Figure 3.6, the six highest – proportion treatments are diabetes (16.5%), hypertension (15.1%), chronic pyelonephritis (13.3%), heart disease (5.7%), hyperlipidemia (4.2%), and arthritis (3.1%), occupying about 60% of all of the proportion of treatments, and The other more – than 100 treatments share the rest 40%.

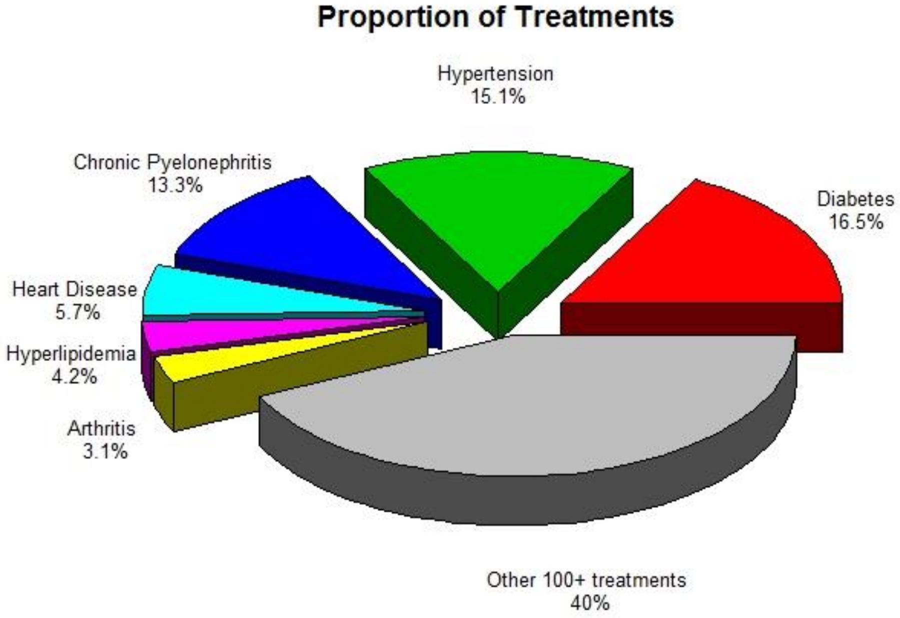

Figure 3.6

To sum this section up, the ages of the patients undergoing dialysis distributed quite normally, the southern Taiwan generally suffer more than any other regions, and there seems to be no difference in proportion between genders and year. Furthermore, the prescription days usually follow the multiple of a week or are exactly a month, and the previous treatments that may be related to dialysis are diabetes, hypertension, chronic pyelonephritis, heart disease, hyperlipidemia, and arthritis.

3-2 Tests

Fisher's Exact Tests were conducted to test for the difference in proportion of individuals undergoing dialysis as a treatment between genders and regions (or areas), respectively. To test the differences between different groups of ages, and difference in proportions of individuals in different year undergoing dialysis as treatment, some Chi-Square tests were performed.

The test results show that there is no difference in proportion undergoing dialysis as a treatment between genders (H_0: There is no difference in proportion of individuals undergoing dialysis as treatment between male and female, p-value=0.8928 > α = 0.05). And also, comparisons of years shows that there is no difference among years.

When it comes to the difference on proportion between regions (areas), the results somehow fit one of the research hypotheses (residents in northern Taiwan suffering more from renal failure than those in other areas). Furthermore, in multiple comparison of areas (note that this test is comparing AREAS, not REGIONS), the test result indicates that there is enough evidence to reject null hypothesis (H_0: The proportion of individuals undergoing dialysis as treatment is no different among the areas in a decade); in other words, at least one pair of the areas has different proportion of individuals undergoing dialysis as a treatment (p-value = 1.222e-05 < α = 0.05).

If examine closely, it could be found that the areas in southern Taiwan (Southern Medical Area and Kaohsiung and Pingtung Medical Area) have relatively larger proportions than every other areas. These facts strongly support one of the research hypotheses, giving answers to our research question (do people in southern Taiwan suffered more from renal failure). However, there is no statistical difference in proportions between regions among Southern Medical Areas (p-value = 0.6077> α = 0.05).

On the other hands, the Chi-Square test for different age groups shows that the individuals in older group has larger proportion of suffering from renal failure.

3-3 Models

As mentioned before, the LASSO logistic regression method was used to select important variables. In this case, 21 out of 103 variables were selected and to be considered as the most important ones. The procedure of variable selection is shown as figure 3.7.

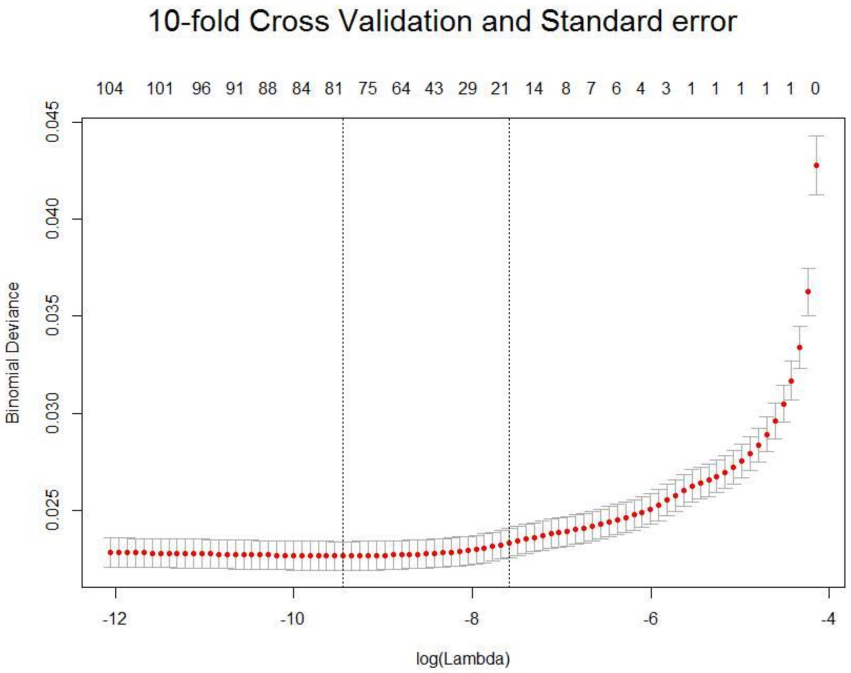

Figure 3.7

*This figure shows how the variable are selected into the model. The lower x-axis' label, *log (Lambda),* represents the scale of the constraint that lessen sum of variable's coefficients, while the upper one representing the number of variables that are already selected into the model. The x-axis represents the deviance (in this case is binomial deviance because of the binary outcome) of the model when applying the variables that already selected. The two straight lines under upper x-axis '21' and '81' are the 'elbow points' that automatically selected by the system, and they are the numbers that the system recommends. In this case, 23 variables are selected into the model.

Since the coefficients estimated by LASSO are not unbiased, generalized linear model (GLM) was used in modeling procedure to guarantee the accuracy of the estimates. The GLM is performed by adding the previous selected variables via LASSO method as variables of interest.

To build and validate the model, the data needs to be divided as training data and testing data. The stratified sampling method is shown in Figure 3.8

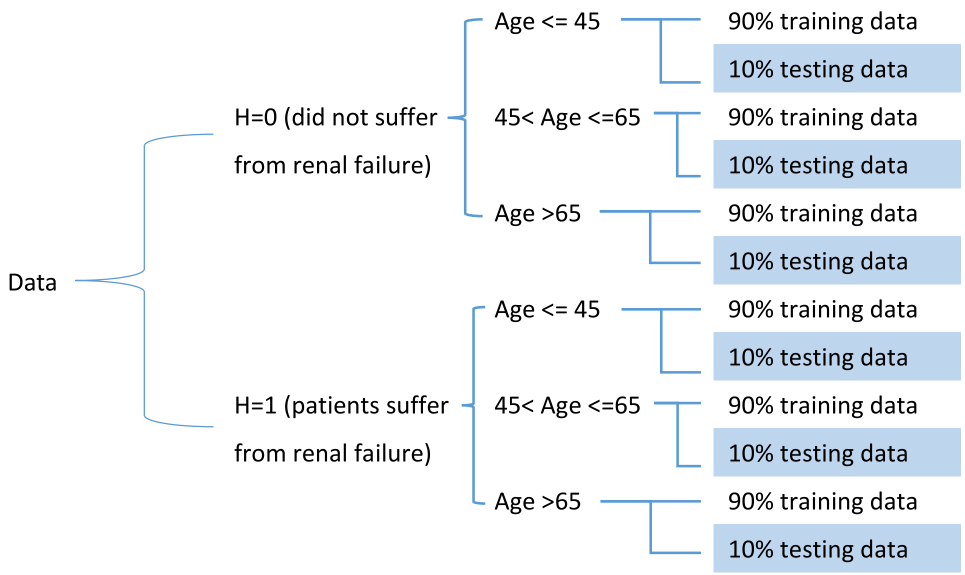

Figure 3.8

The GLM model fitted with the training data is displayed in Table 3.4

Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Intercept	-8.9311	Diabetes	0.1491	Vitreous hemorrhage	0.5606
Middle Age	0.3314	Hypertension	1.1001	Chronic pyelonephritis	0.9501
Old Age	-0.1538	hyperlipidemia	4.9083	Arterial embolism and thrombosis	0.8744
K&P Medical Area$_{*1}$	1.1229	Arthritis	-0.4773	Renal infection	0.6829
C Medical Area$_{*2}$	0.6606	Heart disease	0.3360	Haemophilia	1.3334
N Medical Area$_{*3}$	0.5595	Parkinson's disease	-0.5728	BPH	-0.4260
S Medical Area$_{*4}$	0.8086	Tuberculosis	0.5382	Urinary incontinence	-0.7418
TP Medical Area$_{*5}$	0.6027	Xeropgthalmia	-0.5942	AIDP	0.6403
Average Drug Amount	-0.0010	Retinitis	1.2242	PHP	1.4579

Table 3.4

It might be a common sense that individual in his or her higher age or taking more drugs has a higher probability of suffering from renal failure and undergoing dialysis as treatment, however, the model has converse idea. And the model also infers that individual with some particular disease – which is relate to the treatment – such as chronic pyelonephritis (trt4) and diabetes (trt1) has higher probability of suffering from renal failure.

1: K&P for Kaohsiung. 2: C for Central. 3*: N for Northern. 4*: S for Southern. 5*: TP for Taipei.

After the model is built, it still needs some validation to garentee the predicting power. Figure 3.9 in below displayed the predict result of the testing data.

It is easy to observe that the fitted probabilities which are above the sample proportion ($\widehat{\mathrm{P}}$) are color coded in red and the others black. Although it seems that there are more points above the sample proportion, it is also shown on the plot that only 4% of the predicted probabilities are above sample proportion.

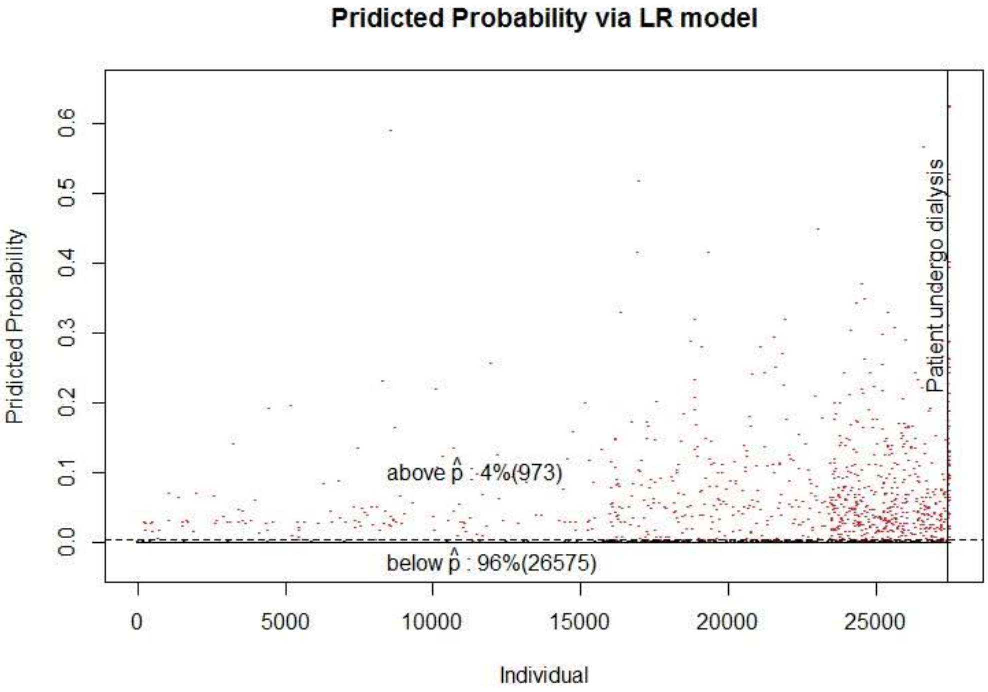

Figure 3.9

To have a closer look, Figure 3.10 performs a partial fitted probability plot. The data in the blue area indicates that the patient are undergoing dialysis. It shows that the model is really promising.

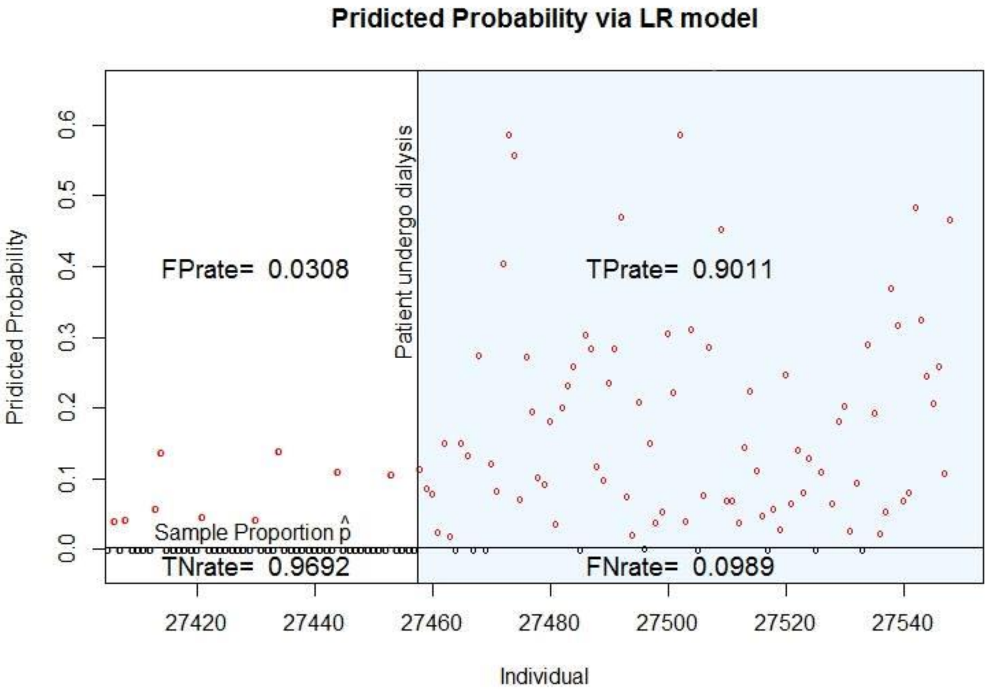

Figure 3.10

Furthermore, by observing the ROC curve (Figure 3.11), it is obvious that the area under curve is 0.9406, which means the model works very well.

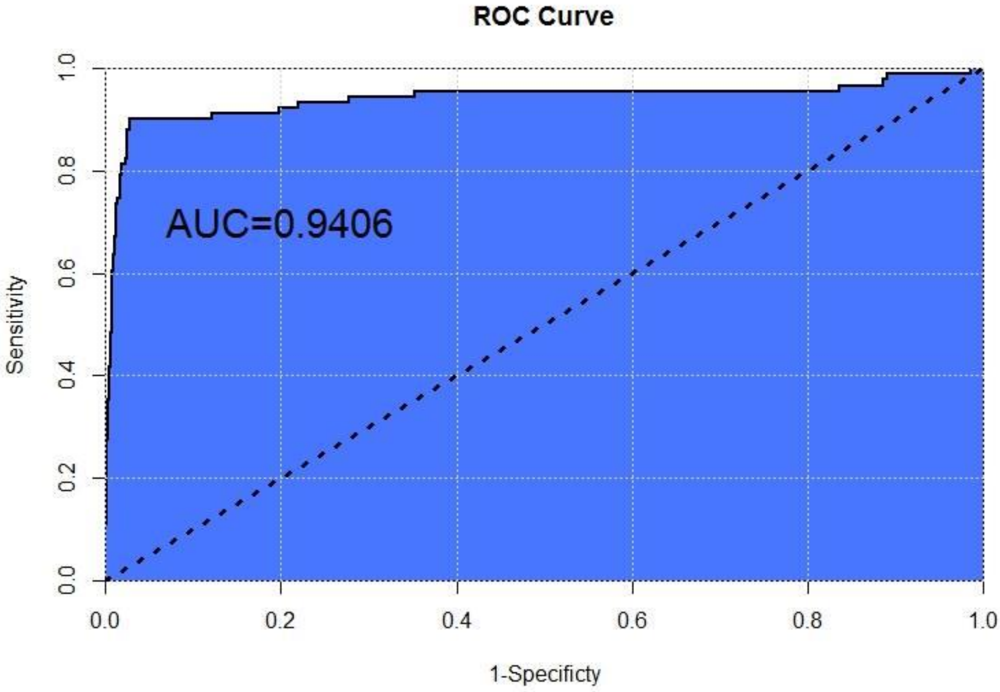

Figure 3.11

4 Conclusion

By the statistical test and the modeling above, there are some important factors that may affect the probability of whether the individual has a greater chance of undergoing a dialysis treatment or not.

First of all is the medical region factor. Several tests were conducted and shown that people who live in south part of Taiwan (coded as Southern Medical Area & Kaohsiung and Pingtung Medical Area) have a greater chance of undergoing dialysis treatment (which means suffering from a severe renal failure.)

Second, Age group also shows an important effect. The result of the test showed that the older the individual, the greater chance the person will undergo dialysis. But in the meanwhile, the coefficients of the model tells differently. It shows that people with age45 -65 has higher chance than that of age over 65. This contradiction may result from the relationships that were not observable between coefficients within the model.

At last, whether an individual is having another medical problem is crucial too. Diabetes, Hypertension, etc. these diseases already have been proved that will lead a higher chance of undergoing dialysis treatment in medical field. And that fact supports the result of the model's coefficients.

In addition, the model could have been better if using Group LASSO Logistic Regression rather than LASSO Logistic Regression.

5 Reference

Bureau of Health Insurance (2010). Executive Yuan of Republic of China

Chang, J. (June, 2014). Business Week

Chien, C., Wang, J., Sun, Y., Sun, D., Sheu, M., Weng, S., Chu, C., Chen, H., Chio, C., Hwang, J., Lu, Y., Wang, Y., and Kan, W. (2012). Long-Term Survival and Predictors for Mortality among Dialysis Patients in an Endemic Area for Chronic Liver Disease: A National Cohort Study in Taiwan

NTNU Holistic Education Program

National Kidney Foundation

National Health Insurance Research Database (2010)

Taiwan Healthcare Reform Foundation (2015)

Zhao, X. (2008). Lasso and Its Applications, University of Minnesota Duluth Graduate Students' Research Topics & Publications

6 Appendix

6-1 Tests

6-1.1 Test for proportions of different genders

H_0: There is no difference in proportion of individuals undergoing dialysis as treatment between Male and Female.

H_a: There is difference in proportion of individuals undergoing dialysis as treatment between Male and Female.

α = 0.05

Fisher's Exact Test for Count Data p-value = 0.8928 > α = 0.05	Test result: There is no enough evidence to reject null hypothesis; thus, it could be concluded that there is no difference in proportion of individuals undergoing dialysis as treatment between Male and Female.

6-1.2 Test for different areas the decade (2002~2011)

To find out whether the proportions of individuals undergoing dialysis is statistically different, such tests are conducted.

6-1.2.1 Overall area comparison

H_0: The proportion of individuals undergoing dialysis as treatment is no different among the areas in a decade.

H_a: The proportion of individuals undergoing dialysis as treatment is different among the areas in a decade.

α = 0.05

Fisher's Exact Test for Count Data p-value = 1.222e-05 < α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is different among areas in a decade.

6-1.2.2 Southern Medical Area versus Taipei Medical Area

H_0: The proportion of individuals undergoing dialysis as treatment in Southern Medical Area is not different from that in Taipei Medical Area.

H_a: The proportion of individuals undergoing dialysis as treatment in Southern Medical Area is higher than that in Taipei Medical Area.

α = 0.05

Fisher's Exact Test for Count Data p-value = 2.7e-05 < α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is higher in Southern Medical Area than that in Taipei Medical Area.

6-1.2.3 Kaohsiung and Pingtung Medical Area versus Taipei Medical Area

H_0: The proportion of individuals undergoing dialysis as treatment in Kaohsiung and Pingtung Medical Area is not different from that in Taipei Medical Area.

H_a: The proportion of individuals undergoing dialysis as treatment in Kaohsiung and Pingtung Medical Area is higher than that in Taipei Medical Area.

$\alpha = 0.05$

Fisher's Exact Test for Count Data p-value = 0.0007757 < α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is higher in Kaohsiung and Pingtung Medical Area than that in Taipei Medical Area.

6-1.2.4 Central Medical Area versus Northern Medical Area

H_0: The proportion of individuals undergoing dialysis as treatment in Central Medical Area is not different from that in Northern Medical Area.

H_a: The proportion of individuals undergoing dialysis as treatment in Central Medical Area is higher than that in Northern Medical Area.

$\alpha = 0.05$

Fisher's Exact Test for Count Data p-value = 0.03601< α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is higher in Central Medical Area than that in Northern Medical Area.

6-1.2.5 Southern Medical Area versus Northern Medical Area

H_0: The proportion of individuals undergoing dialysis as treatment in Southern Medical Area is not different from that in Northern Medical Area.

H_a: The proportion of individuals undergoing dialysis as treatment in Southern Medical Area is higher than that in Northern Medical Area.

$\alpha = 0.05$

Fisher's Exact Test for Count Data p-value = 8.031e-06< α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is higher in Southern Medical Area than that in Northern Medical Area.

6-1.2.6 Kaohsiung and Pingtung Medical Area versus Northern Medical Area

H_0: The proportion of individuals undergoing dialysis as treatment in Kaohsiung and Pingtung Medical Area is not different from that in Northern Medical Area.

H_a: The proportion of individuals undergoing dialysis as treatment in Kaohsiung and Pingtung Medical Area is higher than that in Northern Medical Area.

α = 0.05

Fisher's Exact Test for Count Data p-value = 0.0001816< α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is higher in Kaohsiung and Pingtung Medical Area than that in Northern Medical Area.

6-1.2.7 Southern Medical Area versus Central Medical Area

H_0: The proportion of individuals undergoing dialysis as treatment in Southern Medical Area is not different from that in Central Medical Area.

H_a: The proportion of individuals undergoing dialysis as treatment in Southern Medical Area is higher than that in Central Medical Area.

α = 0.05

Fisher's Exact Test for Count Data p-value = 0.003216< α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is higher in Southern Medical Area than that in Central Medical Area.

6-1.2.8 Kaohsiung and Pingtung Medical Area versus Central Medical Area

H_0: The proportion of individuals undergoing dialysis as treatment in Kaohsiung and Pingtung Medical Area is not different from that in Central Medical Area.

H_a: The proportion of individuals undergoing dialysis as treatment in Kaohsiung and Pingtung Medical Area is higher than that in Central Medical Area.

α = 0.05

Fisher's Exact Test for Count Data p-value = 0.0291< α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is higher in Kaohsiung and Pingtung Medical Area than that in Central Medical Area.

6-1.2.9 Southern Medical Area versus Eastern Medical Area

H_0: The proportion of individuals undergoing dialysis as treatment in Southern Medical Area is not different from that in Eastern Medical Area.

H_a: The proportion of individuals undergoing dialysis as treatment in Southern Medical Area is higher than that in Eastern Medical Area.

$\alpha = 0.05$

Fisher's Exact Test for Count Data p-value = 0.04783< α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is higher in Southern Medical Area than that in Eastern Medical Area.

6-2 Test for different years of the decade

H_0: The proportion of individuals undergoing dialysis as treatment is no difference among every year.

H_a: The proportion of individuals undergoing dialysis as treatment is difference among every year.

$\alpha = 0.05$

Pearson's Chi-squared test
X-squared = 16.0974, df = 9
p-value = 0.06488 >α = 0.05
	Test result:
	There is no enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is no difference among every year.

6-3 Test for regions among Tainan

H_0: The proportion of individuals undergoing dialysis as treatment is no difference among four blocks in Tainan.

H_a: The proportion of individuals undergoing dialysis as treatment is difference among four blocks in Tainan.

$\alpha = 0.05$

Fisher's Exact Test for Count Data p-value = 0.6077> α = 0.05	Test result: There isn't enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is no difference among four blocks in Tainan.

6-4 Test for different age groups

6-4.1.1 Overall comparison

These three age groups are "<45", "45-65", and ">65", and we define these three groups are "under middle age", "middle age", and "old age" separately.

H_0: The proportion of individuals undergoing dialysis as treatment is no difference among these three age groups.

H_a: The proportion of individuals undergoing dialysis as treatment is difference among these three age groups.

α = 0.05

Pearson's Chi-squared test X-squared = 974.3324, df = 2 p-value < 2.2e-16< α = 0.05	Test result: There is enough evidence to reject H_0; thus, it could be concluded that the proportion of individuals undergoing dialysis as treatment is difference among these three age groups.