The Association Between Cancer and Hypertension

Course Title: Analysis and Application of Healthcare Data

Teacher: Yu-Hsiu Lin

Student: 609530005 Pei-Yu Lee

609556006 Yen-Chu Huang

609556012 Yu-Hsuan

Date: 2021/1/7

Abstract

In Taiwan, due to the rapid aging of the population and the unhealthy lifestyles of the people, the number of people suffering from cancer continues to rise, and chronic diseases are no longer the privilege of middle-aged and elderly people. Hypertension is a common risk factor for many major chronic diseases. According to statistics, the top 10 causes of death in Taiwan include cancer and hypertension. These two diseases kill more than tens of thousands of people every year. Therefore, we want to explore whether there is a association between cancer and hypertension, use the chi-square test to analyze whether there is a association between the two. Further explore whether age and gender respectively affect people with cancer, people with hypertension, and people with both cancer and hypertension, or whether they are association or not.

Kew words: Cancer, Hypertension, Association, Age, Gender

I. INTRODUCTION

In Taiwan, due to the rapid aging of the population and the unhealthy lifestyles of the people, the number of cancer patients continues to rise. According to statistics, there were 111,684 new cancers in 2017, an increase of 5,852 from 2016（衛生福利部癌症防治組，2020）. This means that 305.4 people out of every 100,000 people will have cancer, an increase of 8.7 compared to 2016（盧映慈，2020）. The Taiwan Cancer Foundation estimates based on the incidence of domestic cancers that by this year (2020), there will be as many as 110,000 new cancer patients a year. Cancer has always been the top 10 cause of death in Taiwan, killing more than 40,000 people every year. Among them, colorectal cancer, lung cancer and female breast cancer are the top three cancers（衛生福利部癌症防治組，2020）. The risk of men suffering from esophageal cancer and oral cancer is also high. The main causes of cancer are the harm of tobacco, alcohol and betel nut. Women are more likely to develop breast cancer（盧映慈，2020）, because breast screening has become more and more popular in recent years, it can be found in the early stages of cancer. Each type of cancer has a different age group. For colorectal cancer, which has the highest incidence of cancer in the country, the National Health Administration of the Ministry of Health and Welfare has conducted research on people aged 50-69, divided into screening groups and uninfected groups. Screening group, the results show that regular screening can reduce the risk of death by up to 44%（沈彥君，2020）. Cancer can be regarded as a kind of chronic disease in modern times. The World Health Organization (HWO) pointed out that major risk factors such as tobacco, alcohol, unhealthy diet, lack of physical activity, and obesity account for 30% of cancer deaths（盧映慈，2020）. Most of the causes of cancer are related to living habits and diet, such as smoking, chewing betel nut, drinking, etc. The age group exposed to these bad behaviors has been decreasing, leading to many people get cancer

at a young age. In addition, many people in modern times have the wrong concept of having no flesh and unpleasantness and relying on diet to reduce stress. In addition to insufficient intake of fruits and vegetables, more and more people will be obese if exercise is not combined. These are the causes of cancer. It is also the main reason for the high number of cancers.

In addition to cancer, the top ten causes of death in Taiwan also include hypertension. According to statistics from the Ministry of Health and Welfare, cardiovascular diseases caused by hypertension caused 38,290 deaths in 2019, and one out of every 4.5 deaths was due to a disease directly related to hypertension（常春月刊，2020）. The number of direct deaths due to hypertensive diseases in 2019 was also 3.1% higher than in 2018. In the past, people believed that hypertension was a chronic disease that only middle-aged and elderly people suffered. However, in modern times, more and more young people suffer from hypertension due to unhealthy lifestyles. According to the 2016-2019 National Health, Nutrition and Health Status Survey, the prevalence rate of hypertension among people over 18 years old is as high as 25.82%. It is estimated that 5.08 million people suffer from hypertension, and the prevalence rate increases with age（常春月刊，2020）. This is not the only misconception. Many people think it is normal to be older and have high blood pressure, but this is a very wrong rumor. Dr. Wang Zongdao said that blood pressure readings of any age should be controlled at systolic/diastolic blood pressure of 120/80mmHg, and hypertension at any age will increase the risk of death（盧映慈，2019）. Hypertension is a common risk factor for major chronic diseases such as cardiovascular disease, stroke and diabetes, and it is currently ranked first in the global burden of disease（衛生福利部國民健康署，2019）. According to the Global Health Report released by the World Health Organization (HWO) in 2009, hypertension ranks first among the risk factors for death in the world.

Hypertension can be divided into two categories, one is essential hypertension and the other is secondary hypertension. Essential hypertension is related to heredity, lifestyle, and stress, accounting for more than 90%. Secondary hypertension is caused by an abnormality in a certain part of the body, which may be caused by renal vascular disease, cardiovascular disease, hypertension, etc. As long as this abnormal problem is treated, the problem of increased blood pressure can be solved（岡醫衛教中心，2018）. The causes of hypertension include many aspects, including the environment, overweight, excessive intake of salt or alcohol, etc（常春月刊，2020）. Modern people have a high probability of eating out（衛生福利部國民健康署，2019）. The food they eat is high in oil, high salt, high sugar, and lack of exercise. It is the cause of the increase in blood pressure.

Cancer and hypertension are common diseases in Taiwan people in modern times, and they are also one of the top ten causes of death in Taiwan. According to the top ten causes of death by the Ministry of Health and Welfare in 2019, cancer and hypertension ranked first and eighth respectively. The number of deaths due to cancer accounted for 28.6% of all deaths, and the mortality rate of hypertensive diseases also increased. 3.1%（衛生福利部統計處，2020）. These two diseases caused tens of thousands of deaths every year. We want to know whether there are common behaviors or factors that can cause these two diseases at the same time. After that, the two variables of age and gender are further added to explore people with cancer, people with hypertension, and people with both cancer and hypertension, and observe whether these two factors affect them or whether they have association.

II. METHOD

The study used Ambulatory Care Expenditures by Visits file from a National Health Insurance Research Databases (NHIRD) to analyze the association between cancer and hypertension. Figure 1 shows an overview of this process, for more detailed information, please see the following sections.

A. Dataset

The data sets used in this study were provided by the National Health Insurance Research Databases (NHIRD). The association analysis used the 2014 Ambulatory Care Expenditures by Visits file which is a record of 1,350,054 patient visits in 2014, with a total of 90,423 patient data, and 2014 Registry for Beneficiaries file, with total of 100,000 patients, as shown in Table 1.

B. Experimental Environment

We used SAS 9.4, an advanced statistical analysis system, as a tool for data pre-processing and statistical analysis, which has complete data management and data analysis functions.

C. Association Analysis

In addition to analyzing the association between cancer and hypertension, we also considered the association of age and gender with two disease.

According to different age group analysis of cancer patient data from the American Cancer Society (ACS), North American Central Cancer Registry (NAACCR) and the National Center for Health Statistics (NHCS) (Kimberly et al, 2020), they found that cancer prevalence among young adults has increased in the past 10 years. Furthermore, the Health Promotion Administration, Ministry of Health and Welfare 's (MOHW) survey on changes in the people' nutritional health status from 103 to 106 showed that the prevalence of hypertension among aged between 18-39 in Taiwan is 4.7%, and it is

estimated that more than 360,000 young people between 18-39 years old suffer from hypertension（衛生福利部國民健康署，2019）. It is clear that cancer and hypertension are not only found in the elderly, so we added age to the association analysis.

According to the 2017 Health Promotion Administration's Cancer Registry Annual Report, the gender ratio of cancer for men and women was 1:1.2（衛生福利部國民健康署第三科 癌症診療及照護，2021）, and men were more likely to have cancer than women, so we also considered the gender.

1. Data Pre-processing

   Data Pre-processing required to select the age and gender of patients who had cancer and hypertension then delete the missing values, this section contains five steps:

   (1) Since each file of Ambulatory Care Expenditures by Visits file only contains one month data, it is necessary to merge the 12-month files to obtain complete data.

   (2) The patient is marked with hypertension or cancer by selecting the CURE_ITEM_NO1-CURE_ITEM_NO4 columns in the Ambulatory Care Expenditures by Visits file. The flag for cancer patients is 1 and the others are 0. The flag for hypertension patients is 1 and the others is 0. In addition, we also mark the patient who have both cancer and hypertension. The flag for patients who have both hypertension and cancer is 1 and others is 0.

   (3) As ages older than 85 were recorded as 85+ and ages less than one year were recorded in days, neither of which had an accurate age. In order to get the patient’s age, we use 2014 minus the patient's birth year. In this step, we merge the Registry for Beneficiaries file to Ambulatory Care Expenditures by Visits file to get the patient’s birth year. We based on the age grouping of the Ministry of Health and Welfare Statistics, the age groups were divided

into 9 groups. There are 0 years, 1-14 years, 15-24 years, 25-44 years, 45-64 years, 65-74 years, 75-84 years, 85-94 years and over 95 years（衛生福利部統計處，2017）.

(4) In the Gender, 1 is male, 2 is female, and 9 is unknown, so we have deleted the data of unknown gender.

(5) We accumulate the number of visits for patients with cancer and hypertension, and use outpt_ca_count to represent the number of cancer visits, outpt_hy_count represents the number of hypertension visits, then keeping the last one. The number of last visits for each patient was replaced by 1. After delete the duplicate data, we also mark the patient who have both cancer and hypertension with the flag ca_hy_flag. We mark patients who have both cancer and hypertension 1 and the others are 0. We presented the distribution of the cancer and hypertension patient in a contingency table, as shown in Table 2.

2. Statistical Analysis

In order to identify the association between cancer and hypertension, we conducted seven analysis based on four variables: age, gender, cancer and hypertension. The analysis we do are the association between cancer and hypertension, the association between age and hypertension, the association between gender and hypertension, the association between cancer and hypertension, the association between gender and cancer, the association between age and both have cancer and hypertension and the association between gender and both have cancer and hypertension.

We use two common methods, Chi-squared test (Pearson, 1900) and T-test (Gosset, 1908). The age was using independent sample Test, the others were using chi-square test, since the T-test made by SAS is divided into two types of results, with

the same variances and different variances, so we use the F-test (Fisher, 1920) to check whether the variances are the same to determine the result. The following are the null hypothesis and alternative hypothesis for each analysis and the results will be shown in Section 3.

(1) Whether cancer is associated with hypertension.

H0: There is no association between cancer and hypertension.

H1: There is association between cancer and hypertension.

(2) Whether age is associated with hypertension.

H0: There is no association between age and hypertension.

H1: There is association between age and hypertension.

(3) Whether gender is associated with hypertension.

H0: There is no association between gender and hypertension.

H1: There is association between gender and hypertension.

(4) Whether age is associated with cancer.

H0: There is no association between age and cancer.

H1: There is association between age and cancer.

(5) Whether gender is associated with cancer.

H0: There is no association between gender and cancer.

H1: There is association between gender and cancer.

(6) Whether age is associated with both who have cancer and hypertension.

H0: There is no association between age and both who have cancer and hypertension.

H1: There is association between age and both who have cancer and hypertension.

(7) Whether gender is associated with both have cancer and hypertension.

H0: There is no association between gender and both who have cancer and

hypertension.

H1: There is association between gender and both who have cancer and hypertension.

III. RESULTS

A. Descriptive Statistics of Cancer, Hypertension, Cancer & Hypertension, Gender and Age

The following is a descriptive statistical table for this study. The Figure 2 is statistics on whether the patient has cancer or not. Patients with cancer and those without cancer are 1,713 and 86,872, respectively, and the proportions are 1.93 percent and 98.07 percent. The Figure 3 is statistics on whether the patient has hypertension or not. Patients with hypertension and those without hypertension are 12,367 and 76,218, respectively, and the proportions are 13.96 percent and 86.04 percent. The Figure 4 is statistics on whether the patient has cancer and hypertension or not. Patients with cancer and hypertension and those without cancer and hypertension are 603 and 87,982, respectively, and the proportions are 0.68 percent and 99.32 percent. The Figure 5 is statistics of the gender of the research subjects. The numbers of male patients and female patients are 42,429 and 46,156, respectively, and the proportions are 47.90 percent and 52.10 percent. The Figure 6 is statistics of the number of age group. Group 1 is patients whose age is 0 years old and the number of patients is 23. Group 2 is patients whose age is from 1 to 14 years old and the number of patients is 12,570. Group 3 is patients whose age is from 15 to 24 years old and the number of patients is 11,521. Group 4 is patients whose age is from 25 to 44 years old and the number of patients is 26,496. Group 5 is patients whose age is from 45 to 64 years old and the number of patients is 25,846. Group 6 is patients whose age is from 65 to 74 years old and the

number of patients is 6,514. Group 7 is patients whose age is from 75 to 84 years old and the number of patients is 4,061. Group 8 is patients whose age is from 85 to 93 years old and the number of patients is 1,384. Group 9 is patients whose age is over 94 years old and the number of patients is 170.

B. The Relationship Between Cancer and Hypertension

We can find the relationship between cancer and hypertension in Figure 7. The P value of chi-square is less than 0.001. We reject the null hypothesis since the P value is less than 0.05. It is show that cancer and hypertension are not independent.

C. Age

The following is the relationship between three different groups of people and age group.

1. The relationship between cancer and age

   We can find the relationship between cancer and age in Figure 8. From the figure of equality of variances, it can be found that the P value is less than 0.0001. The null hypothesis is rejected because the P value is less than 0.05. Look at the second to last figure in Figure 8. As you can see, the P value is less than 0.0001 in the row of Satterthwaite. Therefore, we reject the null hypothesis again. It is show that there is a significant difference between cancer and age. As shown in Figure 9, the number of patients without cancer is concentrated in group 4 (between 25 and 44 years old) and group 5 (between 45 and 64 years old). The number of patients with cancer is concentrated in group 5 (between 45 and 64 years old).

2. The relationship between hypertension and age

   We can find the relationship between hypertension and age in Figure 10. From the figure of equality of variances, it can be found that the P value is less than 0.0001. The null hypothesis is rejected because the P value is less than 0.05. Look at the second to last figure in Figure 10. As you can see, the P value is less than 0.0001

in the row of Satterthwaite. Therefore, we reject the null hypothesis again. It is show that there is a significant difference between hypertension and age. As shown in Figure 11, the number of patients without hypertension is concentrated in group 4 (between 25 and 44 years old). The number of patients with hypertension is concentrated in group 5 (between 45 and 64 years old).

3. The relationship between cancer & hypertension and age

We can find the relationship between cancer & hypertension and age in Figure 12. From the figure of equality of variances, it can be found that the P value is less than 0.0001. The null hypothesis is rejected because the P value is less than 0.05. Look at the second to last figure in Figure 12. As you can see, the P value is less than 0.0001 in the row of Satterthwaite. Therefore, we reject the null hypothesis again. It is show that there is a significant difference between cancer & hypertension and age. As shown in Figure 13, the number of patients without cancer & hypertension is concentrated in group 4 (between 25 and 44 years old) and group 5 (between 45 and 64 years old). The number of patients with cancer & hypertension is concentrated in group 5 (between 45 and 64 years old), group 6 (between 65 and 74 years old) and group 7 (between 75 and 84 years old).

D. Gender

The following is the relationship between three different groups of people and gender.

1. The relationship between cancer and gender

We can find the relationship between cancer and gender in Figure 14. From the second figure, there is show that the P value of chi-square is 0.0014. The null hypothesis is rejected because the P value is less than 0.05. It is show that cancer and gender are not independent.

2. The relationship between hypertension and gender

We can find the relationship between hypertension and gender in Figure 15. From

the second figure, there is show that the P value of chi-square less than 0.0001. The null hypothesis is rejected because the P value is less than 0.05. It is show that hypertension and gender are not independent.

3. The relationship between cancer & hypertension and gender

   We can find the relationship between cancer & hypertension and gender in Figure 16. From the second figure, there is show that the P value of chi-square is 0.1855. The null hypothesis is accepted because the P value is greater than 0.05. It is show that cancer & hypertension and gender are independent.

IV. DISCUSSION

A. Limitations and Suggestions

There are some limitations to this study, which we describe in detail below. In addition, we make some suggestions for further research.

1. Limited Research

   Because the data in the simulated health insurance database are not in chronological sequence, our study is cross-sectional research and the data are mainly from 2014. The longitudinal research related to the time sequence could not be conducted.

2. Further Research

   This study uses cancer, hypertension, cancer & hypertension, gender and age as variables and tries to explore the relationship between each other. There are many other variables in the file named Ambulatory Care Expenditures By Visits and Registry for Beneficiaries. You can use other variables to do association research. You also can make a variety of combinations by following our method.

B. Conclusion

Cancer has been listed as one of the top 10 causes of death in Taiwan for many years. However, hypertension is also a frequently diagnosed disease in an aging society. In this study, we use National Health Insurance Database for analysis. As we know, hypertension is a risk factor for cancer. In a study showing that hypertension is associated with an increasing risk of mortality from cancer, particularly renal cell carcinoma (Grossman et al., 2002). In medical research, age and gender are one of the variables that researchers often choose. Therefore, we conduct research on the association between cancer, hypertension, age, and gender.

The results of the association analysis showed that there was a association between the number of cancer patients and the number of hypertensive patients. People who did not have hypertension had a higher rate of having cancer than those who had hypertension. Those who had cancer had higher rates of having hypertension than those who did not have cancer. In addition, the highest precision in the classification model was 0.742 and Recall was 0.787. In the five machine learning models, K Nearest Neighbor was found to perform well in the unbalanced model. On the other hand, Random Forest performed better in the balanced model.

From age of the results (see III.results C), we can find that in patients with diseases, whether it is cancer or hypertension, they are distributed in group 5. Therefore, we can say that people aged 45 to 64 are more likely to get diseases, especially cancer and high blood pressure. However, the distribution of people with both hypertension and cancer, it is found that group 6 and group 7 increase significantly. This result confirms a common phenomenon that people will suffer more diseases when they get older. From gender of the results (see III.results D), we can find that age and people with both cancer and hypertension are independent of each other. In other words, there is no association

between these two variables.

In our research, we find that most of cancer and hypertension occur between the ages of 45 to 64. When we get older, the risk of suffering from other diseases will increase. Early prevention is the first step to a better life in the future.

V. REFERENCES

沈彥君（2020 年 4 月 15 日）。癌症發生時間軸 人生走到這個階段你該注意什麼？臺北市北投健康管理醫院。

https://www.tpehealth.com/seminar_article181.html

岡醫衛教中心（2018 年 11 月 16 日）。認識高血壓。秀傳醫療體系-岡醫衛教中心。

https://www.scmh.org.tw/internet/blog/Dept_Preview.aspx?blog=S1NfSEU=&Type=0&Dept=fc9dfc1c-b0dc-4dbb-a707-02bdd9a4113f

常春月刊（2020 年 10 月 23 日）。全台約 508 萬人罹患高血壓！ 十大死因 3 種跟它有關，這些人最好天天都要量血壓。元氣網。

https://health.udn.com/health/story/5977/4958690

衛生福利部國民健康署（2019 年 7 月 12 日）。高血壓。衛生福利部國民健康署。

https://www.hpa.gov.tw/Pages/List.aspx?nodeid=1463

衛生福利部國民健康署（2019 年 9 月 25 日）。高血壓不是中老年人專利!36 萬青壯高血壓大軍，近 24 萬自己不知道。衛生福利部國民健康署。

https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=1136&pid=3106

衛生福利部國民健康署（2019 年 10 月 15 日）。國民營養健康狀況變遷調查成果報告 2013-2016 年。衛生福利部國民健康署。

https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=3999&pid=11145

衛生福利部國民健康署第三科 癌症診療及照護（2021 年 1 月 4 日）。107 年癌症登記報告。衛生福利部國民健康署。

https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=269&pid=13498

衛生福利部統計處（2017 年 6 月 4 日）。人口組成按年齡別分。衛生福利部統計處。https://dep.mohw.gov.tw/dos/cp-1819-3969-113.html

衛生福利部統計處（2020 年 6 月 16 日）。2019 年國人死因統計結果。衛生福利

部。https://www.mohw.gov.tw/cp-16-54482-1.html

衛生福利部癌症防治組（2020 年 9 月 21 日）。衛生福利部公布癌症發生資料。衛生福利部國民健康署。https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=4141&pid=12682

盧映慈（2019 年 6 月 5 日）。血壓沒有年齡對照表！醫師：不管幾歲標準只有一個。元氣網。https://health.udn.com/health/story/5977/3853683

盧映慈（2020 年 6 月 3 日）。最新癌症登記出爐！罹癌人數大增 5.2%，乳癌、大腸癌飆升最快。Heho 健康。https://heho.com.tw/archives/85964

Grossman, E., Messerli, F. H., Boyko, V., & Goldbourt, U. (2002). Is there an association between hypertension and cancer mortality?. *The American journal of medicine, 112(6)*, 479-486.

Miller, K. D., Fidler-Benaoudia, M., Keegan, T. H., Hipp, H. S., Jemal, A., & Siegel, R. L. (2020). Cancer statistics for adolescents and young adults, 2020. *CA: A Cancer Journal for Clinicians, 70(6)*, 443-459. DOI: https://doi.org/10.3322/caac.21637

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50(302)*, 157-175. DOI:10.1080/14786440009463897

VI. FIGURE

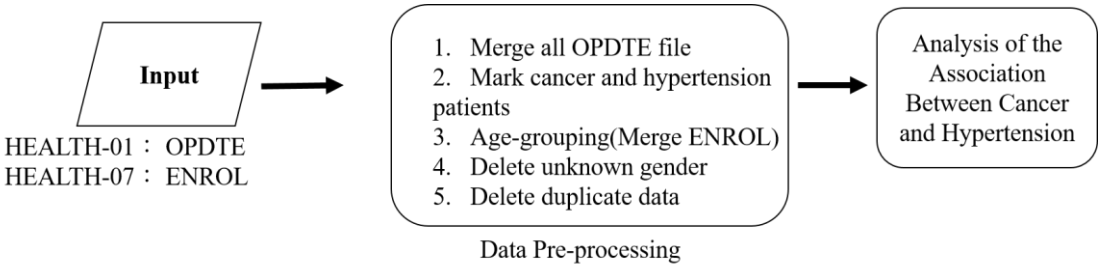


Figure 1. Flow chart of the experiment process

outpt_ca_count	次數	百分比	累積次數	累積百分比
0	86872	98.07	86872	98.07
1	1713	1.93	88585	100.00

Figure 2. Descriptive statistics with and without cancer

outpt_hy_count	次數	百分比	累積次數	累積百分比
0	76218	86.04	76218	86.04
1	12367	13.96	88585	100.00

Figure 3. Descriptive statistics with and without Hypertension

ca_hy_flag	次數	百分比	累積次數	累積百分比
0	87982	99.32	87982	99.32
1	603	0.68	88585	100.00

Figure 4. Descriptive statistics with and without Cancer & Hypertension

ID_S	次數	百分比	累積次數	累積百分比
1	42429	47.90	42429	47.90
2	46156	52.10	88585	100.00

Figure 5. Descriptive statistics of Gender

Agegroup	次數	百分比	累積次數	累積百分比
1	23	0.03	23	0.03
2	12570	14.19	12593	14.22
3	11521	13.01	24114	27.22
4	26496	29.91	50610	57.13
5	25846	29.18	76456	86.31
6	6514	7.35	82970	93.66
7	4061	4.58	87031	98.25
8	1384	1.56	88415	99.81
9	170	0.19	88585	100.00

Figure 6. Descriptive statistics of Age

cancer hyper chisq

FREQ 程序

次數 百分比 列百分比 欄百分比	outpt_ca_count * outpt_hy_count的表格		
	outpt_hy_count		
outpt_ca_count	0	1	總計
0	75108 84.79 86.46 98.54	11764 13.28 13.54 95.12	86872 98.07
1	1110 1.25 64.80 1.46	603 0.68 35.20 4.88	1713 1.93
總計	76218 86.04	12367 13.96	88585 100.00

outpt_ca_count * outpt_hy_count 之表格的統計值

統計值	DF	值	機率
卡方	1	656.1118	<.0001
概度比卡方	1	498.7431	<.0001
連續性調整卡方	1	654.3098	<.0001
Mantel-Haenszel 卡方	1	656.1044	<.0001
Phi 係數		0.0861	
列聯係數		0.0857	
Cramer V		0.0861	

Fisher 精準檢定	
儲存格 (1,1) 次數 (F)	75108
左邊 Pr <= F	1.0000
右邊 Pr >= F	<.0001
表格機率 (P)	<.0001
雙邊 Pr <= P	<.0001

樣本大小 = 88585

Figure 7. Results of chi-square test between Cancer and Hypertension

TTEST 程序

變數: Agegroup

outpt_ca_count	方法	N	平均值	標準差	標準誤差	最小值	最大值
0		86872	4.2075	1.3844	0.00470	1.0000	9.0000
1		1713	5.5698	1.1830	0.0286	2.0000	9.0000
差異 (1-2)	集區		-1.3623	1.3808	0.0337		
差異 (1-2)	Satterthwaite		-1.3623		0.0290		

outpt_ca_count	方法	平均值	95% CL 平均值		標準差	95% CL 標準差	
0		4.2075	4.1983	4.2167	1.3844	1.3779	1.3910
1		5.5698	5.5137	5.6258	1.1830	1.1447	1.2240
差異 (1-2)	集區	-1.3623	-1.4283	-1.2963	1.3808	1.3744	1.3873
差異 (1-2)	Satterthwaite	-1.3623	-1.4191	-1.3055			

方法	變異數	DF	t 值	Pr > |t|
集區	均等	88583	-40.44	<.0001
Satterthwaite	不均等	1805.7	-47.03	<.0001

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	86871	1712	1.37	<.0001

Figure 8. Results of T-test between cancer and age

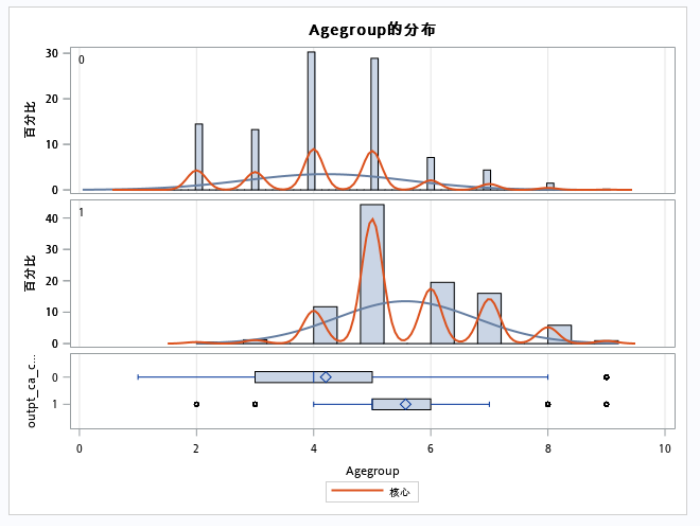


Figure 9. The relationship between cancer patients and age groups

hy age ttest

TTEST 程序

變數: Agegroup

outpt_hy_count	方法	N	平均值	標準差	標準誤差	最小值	最大值
0		76218	3.9993	1.2943	0.00469	1.0000	9.0000
1		12367	5.6795	1.0750	0.00967	2.0000	9.0000
差異 (1-2)	集區		-1.6802	1.2660	0.0123		
差異 (1-2)	Satterthwaite		-1.6802		0.0107		

outpt_hy_count	方法	平均值	95% CL 平均值		標準差	95% CL 標準差	
0		3.9993	3.9901	4.0084	1.2943	1.2878	1.3008
1		5.6795	5.6605	5.6984	1.0750	1.0617	1.0885
差異 (1-2)	集區	-1.6802	-1.7043	-1.6562	1.2660	1.2601	1.2719
差異 (1-2)	Satterthwaite	-1.6802	-1.7013	-1.6592			

方法	變異數	DF	t 值	Pr > |t|
集區	均等	88583	-136.91	<.0001
Satterthwaite	不均等	18700	-156.40	<.0001

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	76217	12366	1.45	<.0001

Figure 10. Results of T-test between hypertension and age

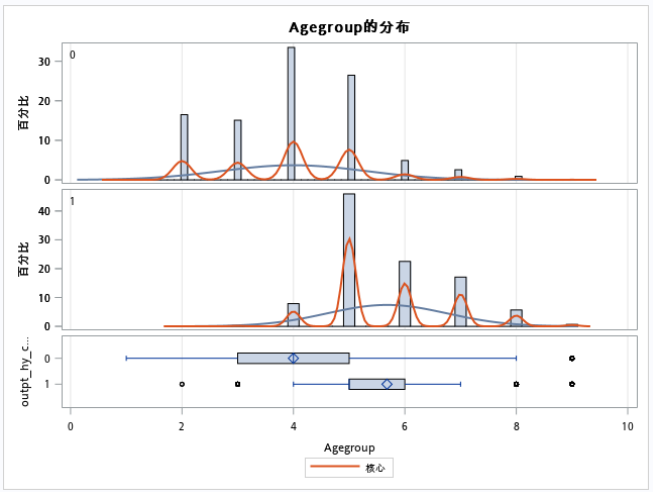


Figure 11. The relationship between hypertension patients and age groups

cancer&hypertension age ttest

TTEST 程序

變數: Agegroup

ca_hy_flag	方法	N	平均值	標準差	標準誤差	最小值	最大值
0		87982	4.2208	1.3862	0.00467	1.0000	9.0000
1		603	6.1294	1.1220	0.0457	4.0000	9.0000
差異 (1-2)	集區		-1.9085	1.3846	0.0566		
差異 (1-2)	Satterthwaite		-1.9085		0.0459		

ca_hy_flag	方法	平均值	95% CL 平均值		標準差	95% CL 標準差	
0		4.2208	4.2117	4.2300	1.3862	1.3798	1.3928
1		6.1294	6.0396	6.2191	1.1220	1.0621	1.1892
差異 (1-2)	集區	-1.9085	-2.0194	-1.7976	1.3846	1.3782	1.3911
差異 (1-2)	Satterthwaite	-1.9085	-1.9987	-1.8183			

方法	變異數	DF	t 值	Pr > |t|
集區	均等	88583	-33.73	<.0001
Satterthwaite	不均等	614.66	-41.55	<.0001

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	87981	602	1.53	<.0001

Figure 12. Results of T-test between cancer & hypertension and age

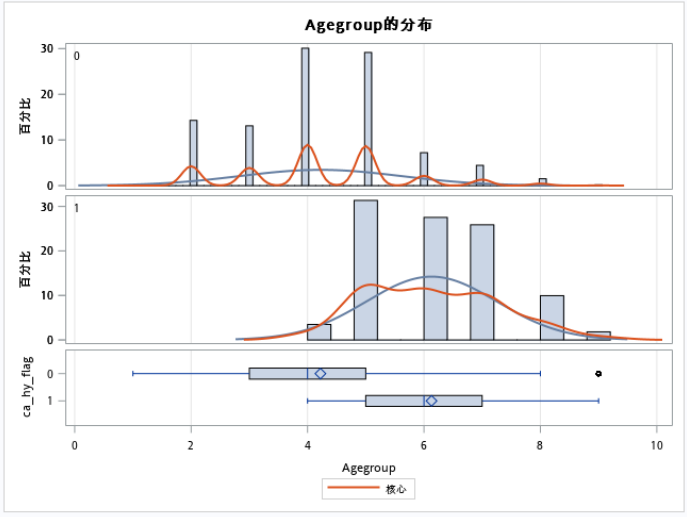
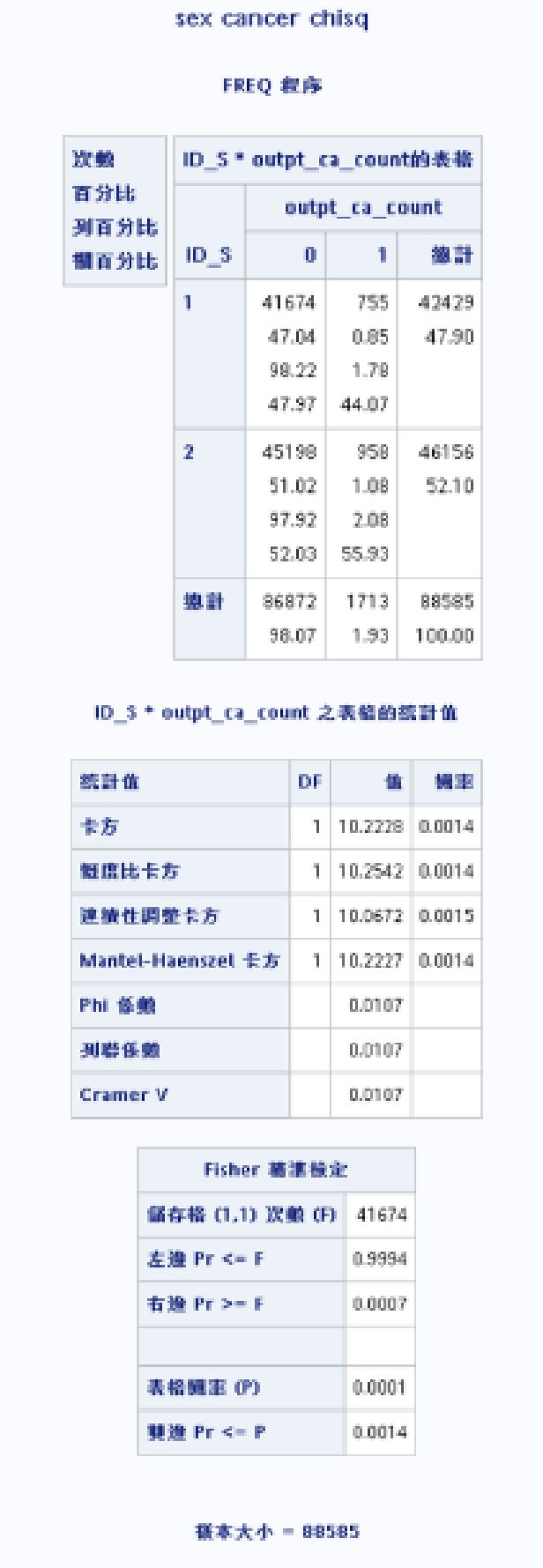


Figure 13. The relationship between cancer & hypertension patients and age groups



sex cancer chisq

FREQ 程序

次數 百分比 列百分比 欄百分比	ID_S * outpt_ca_count的表格		
	outpt_ca_count		
ID_S	0	1	總計
1	41674 47.04 98.22 47.97	755 0.85 1.78 44.07	42429 47.90
2	45198 51.02 97.92 52.03	958 1.08 2.08 55.93	46156 52.10
總計	86872 98.07	1713 1.93	88585 100.00

ID_S * outpt_ca_count 之表格的統計值

統計值	DF	值	概率
卡方	1	10.2228	0.0014
概度比卡方	1	10.2542	0.0014
連續性調整卡方	1	10.0672	0.0015
Mantel-Haenszel 卡方	1	10.2227	0.0014
Phi 係數		0.0107	
列聯係數		0.0107	
Cramer V		0.0107	

Fisher 精準檢定	
儲存格 (1,1) 次數 (F)	41674
左邊 Pr <= F	0.9994
右邊 Pr >= F	0.0007
表格概率 (P)	0.0001
雙邊 Pr <= P	0.0014

樣本大小 = 88585

Figure 14. Results of chi-square test between cancer and gender

sex hy chisq

FREQ 程序

次數 百分比 列百分比 欄百分比	ID_S * outpt_hy_count的表格		
	outpt_hy_count		
ID_S	0	1	總計
1	36262 40.93 85.47 47.58	6167 6.96 14.53 49.87	42429 47.90
2	39956 45.10 86.57 52.42	6200 7.00 13.43 50.13	46156 52.10
總計	76218 86.04	12367 13.96	88585 100.00

ID_S * outpt_hy_count 之表格的統計值

統計值	DF	值	機率
卡方	1	22.3574	<.0001
概度比卡方	1	22.3417	<.0001
連續性調整卡方	1	22.2657	<.0001
Mantel-Haenszel 卡方	1	22.3571	<.0001
Phi 係數		-0.0159	
列聯係數		0.0159	
Cramer V		-0.0159	

Fisher 精準檢定	
儲存格 (1,1) 次數 (F)	36262
左邊 Pr <= F	<.0001
右邊 Pr >= F	1.0000
表格機率 (P)	<.0001
雙邊 Pr <= P	<.0001

樣本大小 = 88585

Figure 15. Results of chi-square test between hypertension and gender

cancer&hypertension sex chisq

FREQ 程序

次數 百分比 列百分比 欄百分比	ID_S * ca_hy_flag的表格		
	ca_hy_flag		
ID_S	0	1	總計
1	42124 47.55 99.28 47.88	305 0.34 0.72 50.58	42429 47.90
2	45858 51.77 99.35 52.12	298 0.34 0.65 49.42	46156 52.10
總計	87982 99.32	603 0.68	88585 100.00

ID_S * ca_hy_flag 之表格的統計值

統計值	DF	值	機率
卡方	1	1.7527	0.1855
概度比卡方	1	1.7509	0.1858
連續性調整卡方	1	1.6460	0.1995
Mantel-Haenszel 卡方	1	1.7526	0.1855
Phi 係數		-0.0044	
列聯係數		0.0044	
Cramer V		-0.0044	

Fisher 精準檢定	
儲存格 (1,1) 次數 (F)	42124
左邊 Pr <= F	0.0998
右邊 Pr >= F	0.9138
表格機率 (P)	0.0136
雙邊 Pr <= P	0.1907

樣本大小 = 88585

Figure 16. Results of chi-square test between cancer & hypertension and gender

VII. TABLE

Table 1. Statistics of the NHIRD data

File name	File code	Number of data
Ambulatory Care Expenditures by Visits	H_NHI_OPDTE	1,350,054
Registry for Beneficiaries	H_NHI_ENROL	100,000

Table 2. Contingency table of cancer and hypertension

	Hypertension	Non-hypertension	total
Cancer	603	1110	1713
Non-cancer	11764	75108	86872
total	12367	76218	88585

VIII. SAS PROGRAMMING

```
libname TP  "E:\SAS\term paper";


data TP.opdteall;
set TP.H_nhi_opdte10301_10 TP.H_nhi_opdte10302_10
TP.H_nhi_opdte10303_10 TP.H_nhi_opdte10304_10 TP.H_nhi_opdte10305_10
TP.H_nhi_opdte10306_10 TP.H_nhi_opdte10307_10 TP.H_nhi_opdte10308_10
TP.H_nhi_opdte10309_10 TP.H_nhi_opdte10310_10 TP.H_nhi_opdte10311_10
TP.H_nhi_opdte10312_10;
run;

data TP.enrolall;
set TP.h_nhi_enrol10301 TP.h_nhi_enrol10302 TP.h_nhi_enrol10303
TP.h_nhi_enrol10304 TP.h_nhi_enrol10305 TP.h_nhi_enrol10306
TP.h_nhi_enrol10307 TP.h_nhi_enrol10308 TP.h_nhi_enrol10309
TP.h_nhi_enrol10310 TP.h_nhi_enrol10311 TP.h_nhi_enrol10312;
run;

/*標記有癌症和有高血壓的人*/
data TP.cancer_hyper;
set TP.opdteall;
cancer_flag=.;
hyper_flag=.;
if substr(CURE_ITEM_NO1,1,2) ='12' or substr(CURE_ITEM_NO2,1,2) ='12'
or  substr(CURE_ITEM_NO3,1,2) ='12' or  substr(CURE_ITEM_NO4,1,2)
='12'  then cancer_flag=1;
else cancer_flag=0;
if substr(CURE_ITEM_NO1,1,2) ='02' or substr(CURE_ITEM_NO2,1,2) ='02'
or  substr(CURE_ITEM_NO3,1,2) ='02' or  substr(CURE_ITEM_NO4,1,2)
='02'  then hyper_flag=1;
else hyper_flag=0;
run;

proc sort data= TP.cancer_hyper;
by id;
run;
```

```
data TP.cancer_hyper;
set  TP.cancer_hyper;
by id;
retain outpt_ca_count 0;
outpt_ca_count = outpt_ca_count+cancer_flag;
if first.id then outpt_ca_count=cancer_flag;

retain outpt_hy_count 0;
outpt_hy_count = outpt_hy_count+hyper_flag;
if first.id then outpt_hy_count=hyper_flag;
run;

data TP.cancer_hyper;
set  TP.cancer_hyper;
by id;
if last.id;
run;

data TP.cancer_hyper;
set TP.cancer_hyper;
if outpt_ca_count>0 then outpt_ca_count=1;
if outpt_hy_count>0 then outpt_hy_count=1;
run;

/*處理年齡*/
proc sort data=TP.cancer_hyper;
by id;
run;
proc sort data=TP.enrolall nodupkey;
by id;
run;

data TP.cancer_hyper2;
merge TP.cancer_hyper(in=a)TP.enrolall (in=b);
by id;
if a;
run;
```

```
data TP.cancer_hyper2;
set TP.cancer_hyper2;
if outpt_hy_count=. then delete;
if id_s=9 then delete;
age=2014-ID_BIRTH_Y;
run;

/*年齡分群*/
data TP.cancer_hyper2;
set TP.cancer_hyper2;
Agegroup=.;
if age=0 then Agegroup=1;
else if age>=1 and age<15 then Agegroup=2;
else if age>=15 and age<25 then Agegroup=3;
else if age>=25 and age<45 then Agegroup=4;
else if age>=45 and age<65then Agegroup=5;
else if age>=65 and age<75 then Agegroup=6;
else if age>=75 and age<85then Agegroup=7;
else if age>=85 and age<94 then Agegroup=8;
else Agegroup=9;
run;
data TP.cancer_hyper2;
set TP.cancer_hyper2;
if outpt_hy_count=1 and outpt_ca_count=1 then ca_hy_flag=1;
else ca_hy_flag=0;
run;

/*描述性統計*/
title '描述性統計outpt_ca_count outpt_hy_count ca_hy_flag id_s';
proc freq data=TP.cancer_hyper2;
table outpt_ca_count outpt_hy_count ca_hy_flag id_s Agegroup;
run;

/*癌症與高血壓相關*/
title 'cancer hyper chisq';
proc freq data=TP.cancer_hyper2;
table  outpt_ca_count*outpt_hy_count /  chisq;
run;
```

```
/*年齡與癌症相關*/
title 'cancer age ttest';
proc  ttest data=TP.cancer_hyper2;
class outpt_ca_count;
var Agegroup ;
run;

/*年齡與高血壓相關*/
title 'hy age ttest';
proc  ttest data=TP.cancer_hyper2;
class outpt_hy_count;
var Agegroup ;
run;

/*年齡與同時有高血壓癌症相關*/
title 'cancer&hypertension  age ttest';
proc  ttest data=TP.cancer_hyper2;
class ca_hy_flag;
var Agegroup ;
run;

/*性別與癌症相關*/
title 'sex cancer  chisq';
proc freq data=TP.cancer_hyper2;
table  id_s*outpt_ca_count /  chisq;
run;

/*性別與高血壓相關*/
title 'sex hy  chisq';
proc freq data=TP.cancer_hyper2;
table  id_s*outpt_hy_count /  chisq;
run;

/*性別與同時有高血壓癌症相關*/
title 'cancer&hypertension sex  chisq';
proc freq data=TP.cancer_hyper2;
table  id_s*ca_hy_flag /  chisq;
run;
```

File name	File code	Number of data
Ambulatory Care Expenditures by Visits	H_NHI_OPDTE	1,350,054
Registry for Beneficiaries	H_NHI_ENROL	100,000