

Factors affecting utilization of healthcare services in Australia

Jixin Li^{1, †}, Yuze Li^{2, *, †}

¹Virginia Tech, Blacksburg, VA 24061

²Lehigh University, 27 Memorial Dr W, Bethlehem, PA 18015

*Corresponding Author Email: yul223@lehigh.edu

†These authors contributed equally.

Abstract. The health care access rate can highly contribute to society's living quality, and further influence the general labor productivity and economic growth of a country. However, the health expenditure per person in Australia is very high and not able to decrease immediately, which can gradually impact the proportion of people visiting health care services. Thus, it is important to reveal other important factors influencing health care utilization and seek solutions via the results. This study aimed to reveal the factors associated with the healthcare service utilization in Australia. We first hypothesized that insurance coverage will be the factor with the most contribution. The data, collected from the health survey conducted by the Australian Bureau of Statistics in 1978-1978, were analyzed via regression model and exploratory data analysis. Model comparison was also performed to receive a more accurate conclusion. The overall results showed that factors influencing health conditions, such as higher age, diseases affecting daily activities appeared to have the greatest relevance with utilization rate. Insurance is also a significant factor, but has a lower contribution compared to the health-related ones. The results can act as an important indicator in introducing policies in health care system and enhance citizen's health conditions.

Keywords: health care utilization; insurance; chronic diseases.

1. Introduction

According to Maslow's Hierarchy of Needs theory, physiological needs are the most basic needs for humans that must be fulfilled to pursue other higher satisfaction [1]. Being healthy, as one of the physiological needs, is very important to every human being, which then raises the demand of healthcare services to maintain society's living quality. The access rate to health care services will also affect the community's economic performance and will further influence labor productivity [2].

Nonetheless, not everyone has the ability to use these services. Evidence has shown that one possible affecting factor is the spending on healthcare, which is expensive and unaffordable to some individuals. According to Organization for Economic Cooperation and Development's data on health spending in 2020, 27 countries have a cost of more than 3,000 dollars per person needed for healthcare, while the highest cost is up to 10,948 dollars per person [3]. The cost will not stay at a steady level, which has inclined year by year. In the 2018-2019 report of health expenditure in Australia, the decade average increase in health spending is 3.5%, and the spending per person is up to \$7,772 [4]. The increasing trend in health expenses is clearly unable to moderate at immediate time, so it is important to consider other factors and find possible solutions to increase the utilization of healthcare. As revealed in past studies, insurance coverage, gender, income level, etc., all have effects on the utilization of health care services. However, previous studies have been done only on examining a certain factor's influence, with few groups presenting a summarized result on comparing the notability and contribution of each factor. It is necessary to discover a model that shows the general impacting factors to health care utilizations in order to favor later policy decisions.

In this study, we aimed to uncover the factors other than healthcare cost that are associated with healthcare service utilizations and their significance. It was hypothesized that insurance would be highly associated with healthcare utilization, along with other minor factors. Regression models and exploratory data analysis were introduced in this study to examine our hypothesis. The data used in this study was collected from the National health survey conducted by Australian Bureau of Statistics,

which contains various categories affecting characteristics on health care utilization of surveyed Australians. Although the survey was conducted in 1977-1978, it fitted with our objective of ignoring the influence of health expenditures, with relatively low cost in healthcare in the 1980s [3].

2. Method

2.1. Data collection

The data was called Australian Health Service Utilization Data. It can be accessed from the Rdatasets website (<https://vincentarelbundock.github.io/Rdatasets/articles/data.html>), which is the collections of over 1700 datasets that were originally distributed alongside the statistical software environment R. The data originated from the 1977-1978 Australian Health Survey conducted by Australian Bureau of Statistics, which is Australia's national statistical agency providing trusted official statistics on a wide range of economic, social, population and environmental matters.

2.2. Variable measuring

The data measured 5190 observations on 12 variables including visits, gender, age, income, illness, reduced, health, private, freepoor, freerepat, nchronic, and lchronic. Visits meant the number of doctor visits in the past 2 weeks. The data contained both male and female. Age meant age in years divided by 100. Income meant the annual income in tens of thousands of dollars. Illness meant the number of illnesses in the past 2 weeks. Reduced meant the number of days of reduced activity in the past 2 weeks due to illness or injury. Health meant the general health questionnaire score using Goldberg's method. Private was a factor variable that contains yes or no, which indicated if someone has private health insurance. Similarly, variables including freepoor, freerepat, nchronic, and lchronic are all factors that contained yes or no, which separately meant if the individual has free government health insurance due to low income, the individual have free government health insurance due to old age, disability or veteran status, is there a chronic condition not limiting activity, and is there a chronic condition limiting activity. These variables were measured from people who took the Australian Health Survey from 1977-1978. All variables were collected by the Australian Bureau of Statistics.

2.3. Data analyzing

2.3.1 Data manipulation

We used a software called R studio in this data analysis. In order to make it easy to be analyzed, we manipulated the data by adding two new variables. The first one was general visit, which was a factor indicating whether an individual visited the doctor in the past 2 weeks and "1" meant visited, "0" meant did not visit. Another one was insurance, which was the factor that indicated whether an individual had one type of insurance from following category: freepoor, freerepat, or private insurance. After adding new variables, we deleted useless variables including visits, private, freepoor, and freerepat. In order to conduct the logistic regression later based on Y variable general visits, we also made all factor variables become "0" or "1", specifically, "0" represented no and "1" represented yes. "0" represented female and "1" represented male in the gender variable.

2.3.2 Exploratory data analysis

Exploratory data analysis was the way to analyze data by summarizing variables' characteristics. We calculated the mean and standard deviation for each numerical variable and the percentage of each category for each categorical variable. We used two-way tables to analyze the relationship between each categorical variable and general visit. Correlation matrix was a table used to show correlation coefficients between variables. We did the numerical correlation matrix to see the correlation between each numerical variable.

2.3.3 Logistic regression

Logistic regression was a tool used to predict the probability of a discrete outcome given an input variable. Specifically, we used logistic regression to estimate the possibility that someone visited doctors by given variables age, gender, income, illness, reduced, health, nchronic, lchronic, and insurance.

2.3.4 Model selection

Model selection was a process to determine the best fit model for the data by comparing AIC values of each model. The model with lowest AIC was the best one. We did the model selection and conducted both backward and forward selection to select the best one. After that, we did logistic regression again using the best model to compare results before and after the selection.

3. Results

3.1. Exploratory Data Analysis

Table 1 summarized the mean and standard deviation for each numerical variable, and the number and percentage of each category for each categorical variable. Each numerical variable has a small mean and standard deviation. As table 2 showed, the proportions of males and females in this data are close to equal. 59.69% of the people have chronic conditions but don't limit their activities. Only 11.66% of people have severe chronic conditions that limit their activities. As is shown in table 1, 69.58% people have insurance no matter what type it is.

Table 1. Summary of numerical variables in the data and their mean and standard deviation

Numeric Variable	N	Mean*	Sd
Age	5190	0.4064	0.2048
Income	5190	0.5832	0.3689
Illness	5190	1.4320	1.3842
Reduced	5190	0.8618	2.8876
Health	5190	1.2175	2.1243

*Age is calculated by age divided by 100. Income is in tens of thousands of dollars. Illness indicates the number of illnesses in the past two weeks. Reduced indicates the number of days of reduced activity in the past two weeks. Health is the score of general health questionnaire.

Table 2. Summary of categorical variables in the data and their proportions

Categorical Variable	N	%
Gender	5190	-
Female	2702	52.06%
Male	2488	47.94%
Nchronic	5190	-
Yes	2092	40.31%
No	3098	59.69%
Lchronic	5190	-
Yes	605	11.66%
No	4585	88.34%
Insurance	5190	-
Yes	3611	69.58%
No	1579	30.42%

Table 3. The relative frequencies for gender and Nchronic variables and the visits conditions

	Female	Male	Total		Nchronic	Without Nchronic	Total
Visitors	660	389	1049	Visitors	495	554	1049
Non-visitors	2042	2099	4141	Non-visitors	1597	2544	4141
Total	2702	2488	5190	Total	2092	3098	5190

As is shown in Table 3. 24.4% (660 is divided by 2702) female and 15.6% (389 is divided by 2488) male visited doctors, therefore, females are more likely to visit doctors than males. 23.7% (495 is divided by 2092) people with Nchronic and 17.9% (554 is divided by 3098) people without Nchronic visited doctors. Therefore, people who have chronic conditions that don't limit activities are more likely to visit doctors than healthy people.

Table 4. The relative frequencies for Insurance and Lchronic variables and the visits conditions

	Insurance	Without insurance	Total		Lchronic	Without Lchronic	Total
Visitors	825	224	1049	Visitors	212	837	1049
Non-visitors	2786	1355	4141	Non-visitors	393	3748	4141
Total	3611	1579	5190	Total	605	4585	5190

As is shown in Table 4, 22.8% (825 is divided by 3611) people with insurance and 14.2% (224 is divided by 1579) people without insurance visit doctors, therefore, people with insurance are more likely to visit doctors than people without insurance. 35% (212 is divided by 605) of people with Lchronic and 18.3% (837 is divided by 4585) people without Lchronic visited doctors. Therefore, people who have chronic conditions that limit activities are more likely to visit doctors than healthy people. One important funding is 35% people with Lchronic visited doctors, which is much higher than overall 20.2% (1049 is divided by 5190) people visited doctors in the past two weeks.

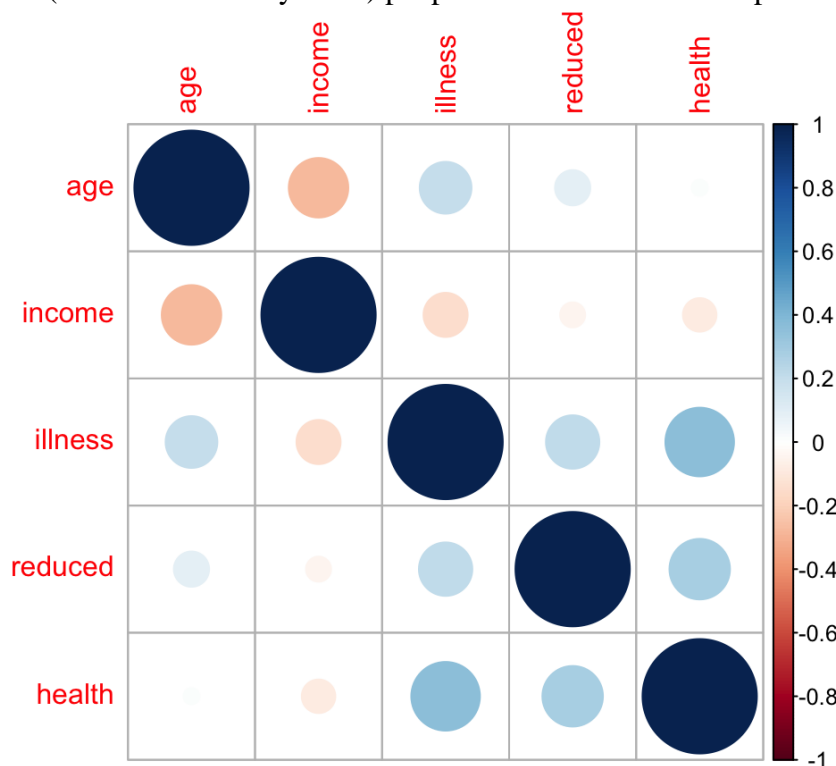


Figure 1. Relative correlation between five numerical variables

The figure 1 indicates the correlation between numerical variables. Positive correlation coefficient value represents positive association between two variables, which shows in blue color. Negative correlation coefficient value represents negative association between two variables, which shows in red color. Large size indicates strong correlation and small size means weak correlation. All variables are positively correlated with each other except income. Overall, variables health and illness have the strongest positive correlation. In contrast, the variable reduced and illness have the least negative correlation.

3.2. Logistic Regression

Table 5 showed the results of logistic regression. The coefficient estimate means the average change of the response variable associated with the increase in each predictor variable. For example, the possibility of visiting doctors would increase 0.858 with one unit increase in age. Among all variables, only increased income and male showed the decrease of possibility of visiting doctors. The standard error means the variability associated with the coefficient estimate. Most variables have the low standard error value that is close or less than 0.01. However, compared with others, variable age has a larger standard error value that is close to 0.02. P value means how well each predictor variable can predict the value of the response variable. Small value indicates great ability of prediction of the response variable and high level of significance. Variables including age, illness, reduced, health, and gender are highly significant due to their small P values which is smaller than 0.001. For those variables with P value larger than 0.001, Lchronic is the most significant followed by insurance, and Nchronic. Income is the least significant because it has the largest P value.

Table 5. The significance of all variables and relation between them and predicted response

Variables	Estimate (Standard error)	P-value
Age	0.858 (0.216)	<0.001
Income	-0.011 (0.112)	0.922
Illness	0.270 (0.029)	<0.001
Reduced	0.157 (0.012)	<0.001
Health	0.058 (0.017)	<0.001
Gender	-	-
Female	Reference	-
Male	-0.298 (0.082)	<0.001
Nchronic	-	-
No	Reference	-
Yes	0.096 (0.091)	0.292
Lchronic	-	-
No	Reference	-
Yes	0.251 (0.124)	0.043
Insurance	-	-
No	Reference	-
Yes	0.195 (0.097)	0.045

Table 6 shows logistic regression results of the best model after the selection. Comparing it with table 3, the coefficient estimates, and standard error values didn't change a lot after selection. However, the best model removed two variables with high p-value, which are Nchronic and income.

In addition, the p value of Lchronic increases from 0.043 to 0.082, which displays the decrease of significance. The p values of insurance decreases from 0.045 to 0.039 to display the increase of significance.

Table 6. The significance of selected variables and relation between them and predicted response

Variables	Estimate	(standard error)	P-value
Age	0.858	(0.202)	<0.001
Illness	0.270	(0.029)	<0.001
Reduced	0.157	(0.012)	<0.001
Health	0.057	(0.017)	<0.001
Gender	-	-	-
Female	Reference	-	-
Male	-0.297	(0.080)	<0.001
Lchronic	-	-	-
No	Reference	-	-
Yes	0.189	(0.109)	0.082
Insurance	-	-	-
No	Reference	-	-
Yes	0.200	(0.097)	0.039

4. Discussion

The goal of our study is to find out the relationships between factors and the usage of healthcare service in Australia. The results show that factors that are connected to people’s health conditions, such as age and illness, have the strongest relationship with visits to healthcare service.

Earlier research conducted by Dutton, which studied the professional and organizational factors contribution, also showed a similar conclusion with our result that the most contributing factors are highly related with health conditions [5]. In another study by Sari in 2009, showed evidence that the people with inactivity will spend more time and have a higher rate of visiting healthcare services [6].

Gender is also strongly affecting the visiting status, consistent with other research studying gender’s contribution, which states that females tend to have higher willingness to use healthcare services [7]. Women may have a higher level of mental distress than men, thus tend to visit health care services more [8]. Atella et al. also revealed that age is strongly affecting health utilization as it is also a strong factor of chronic disease [9]. Prescriptions and diagnostic tests have each increased a percentage of 26% and 27% in the decade of 2004-2014. This is consistent with our modeling result that age is a highly contributing factor, which is the older the observation, the higher healthcare utilization rate will be. Compared to the above factors, insurance and activity-limiting chronic disease have relatively low relationship with healthcare utilization. This result is also confirmed with Caltin’s study in 2016 [10]. Surprisingly, chronic disease not limiting activity contributes least to the utilization status of healthcare, since these two factors were removed from the prediction model shown in table 6. These results contrast with our hypothesis that insurance is the main affecting factor

to healthcare usage. The model also removed the income factor, which is matched with our goal of finding utilization affecting factors other than healthcare cost.

In summary, the significant factors from the final model all have shown an consistent relationship with the past studies. What we can make a final deduction from the results is that people tend to visit healthcares when they have health problems that affect their daily activities to seek for cure, no matter with their income level.

5. Conclusion

We have shown the significance of each variable in this data and figured out the best model to predict the possibility of using healthcare services, which is conducted by analyzing the relationship between variables, fitting logistic regression, and model selection. Through our analysis of Australian Health Service Utilization Data, we are trying to find if expensive cost is the main factor that affects one's use of healthcare services and what other factors contribute to it. In fact, we found that income level or high cost of healthcare services is not the main reason to determine if someone visits the doctor or not. However, some physiological factors including age, health condition, and even gender contribute to the use of healthcare utilization more than income level. In addition, the social factor whether someone has insurance or not affects it more than income as well, but its significance is still lower than other physiological factors.

The significance of physiological factors and insurance helps us point out that the Australian government or citizens should pay more attention to improving their health or insurance conditions instead of pursuing higher income to guarantee the cost of medical services. There are many ways to improve it such as increasing the number of health clubs in the community, taking the annual physical exam, and formulating policies that provide insurance for people who are disabled, unemployed, or aged. Our finding also indicates that income level is negatively correlated with other numerical physiological variables, which means people with low income are more likely to have health problems. Therefore, the population with low income should be concerned, and Australian society should safeguard them have access to healthcare services and keep healthy.

There are still some limitations in this study. Firstly, the data is collected from 1977 to 1978, which only reflects the pattern of Australian health utilization in the 1970s and may not completely coincide with the modern population. Secondly, the data lost many important variables such as race and ethnicity, level of urbanization, and residents' sociodemographic characteristics due to the location conducted the survey.

In conclusion, the model developed in this study can detect the health utilization of a country over decades to summarize health conditions in each generation. And the uncovered factors that contribute to health service utilization is crucial to determine the citizen's need for health service and point out the direction of improving the quality of medical service.

References

- [1] Maslow, A. (1954). *Motivation and personality*. New York, NY: Harper.
- [2] Raghupathi, V. and Raghupathi, W., 2020. Healthcare Expenditure and Economic Performance: Insights From the United States Data. *Frontiers in Public Health*, 8.
- [3] OECD (2020), *Health spending* (indicator). doi: 10.1787/8643de7e-en (Accessed on 31 March 2022)
- [4] Australian Institute of Health and Welfare (2020), *Health expenditure Australia*. doi: 10.25816/ysem-1g13 (Accessed on 19 April, 2022)
- [5] Dutton, D., 1986. Financial, organizational and professional factors affecting health care utilization. *Social Science & Medicine*, 23(7), pp.721-735.

- [6] Sari, N., 2009. Physical inactivity and its impact on healthcare utilization. *Health Economics*, 18(8), pp.885-901.
- [7] Vaidya, V., Partha, G. and Karmakar, M., 2012. Gender Differences in Utilization of Preventive Care Services in the United States. *Journal of Women's Health*, 21(2), pp.140-145.
- [8] Koopmans, G. and Lamers, L., 2007. Gender and health care utilization: The role of mental distress and help-seeking propensity. *Social Science & Medicine*, 64(6), pp.1216-1230.
- [9] Atella, V., Piano Mortari, A., Kopinska, J., Belotti, F., Lapi, F., Cricelli, C. and Fontana, L., 2018. Trends in age-related disease burden and healthcare utilization. *Aging Cell*, 18(1), p.e12861.
- [10] Caitlin M. Farrell and Aaron Gottlieb, 2020: The Effect of Health Insurance on Health Care Utilization in the Justice-Involved Population: United States, 2014–2016, *American Journal of Public Health* 110, S78_S84