

## ANALYSIS OF FACTORS INFLUENCING THE DETERMINATION OF MEDICAL EXPENDITURE FOR THE ELDERLY IN KOREA

CHANGSOO LEE<sup>1</sup> AND YOUNGBIN PARK<sup>2</sup>

<sup>1</sup>Department of Industrial and Management Engineering  
Gangneung-Wonju National University  
150 Namwo-ro, Heungeup-myeon, Wonju-si, Gangwon-do 26403, Korea  
cslee@gwnu.ac.kr

<sup>2</sup>Department of Information Statistics  
Gangneung-Wonju National University  
7 Jukheon-gil, Gangneung-si, Gangwon-do 25457, Korea  
dudqls1201@gwnu.ac.kr

Received June 2021; accepted August 2021

**ABSTRACT.** *A demand for medical services arises from the increased satisfaction that patients get when their health improves after receiving medical services. On average, the lower the health level, the higher the demand for medical services and the higher the medical expenses. Therefore, it can be predicted that as the percentage of the aging population increases, the public's medical expenses will increase. This thesis aims to find out what factors influence an individual's decision regarding their medical expenses. Among the demographic/socioeconomic/health-related factors that can affect medical expenses, health-related factors were found to have the greatest influence. Among them, physical activity restriction and the number of chronic diseases had the greatest influence on determination of medical expenditure, and the amount of medical expenses was also significant.*

**Keywords:** Aging society, Medical expenditure, Healthcare industry

1. **Introduction.** South Korea's aging population, citizens aged 65 years or older, is predicted to reach 37.0 percent by 2045. Projections show South Korea's elderly population will grow at the fastest pace in the world, going from 14.9 percent by 2019 to 46.5 percent by 2067 [1].

Commensurate to South Korea's growing elderly population, medical expenditure for the elderly is increasing every year. Higher medical expenditure causes older adults to become financially vulnerable. According to the Health Insurance Review & Assessment Service, as of 2019, in Korea, the average medical cost per person aged 65 and over in Korea was 4.91 million won per year. This is about 2.9 times the average medical cost per capita namely, 1.68 million won per year. An increase in the aging population leads to an increase in the cost of medical care for the elderly. In 2019, the medical cost for older adults was 35.79 trillion won, a 1.6-fold increase from 2015. As such, medical expenses are expected to increase further in the future due to a rapidly aging society, increasing life expectancy, and rise in chronic diseases [2]. Consequently, rapidly increasing medical expenses will be a burden on society in the long run.

The motivation for this study is as follows. We wanted to know what factors influence medical expenses within the range of individuals and started a study to convey the need for people to recognize the main factors affecting medical expenses and to spend them efficiently. Therefore, the objective of this study is to explore the factors influencing the medical expenses of the elderly and spend efficient medical expenses in consideration of the individual's condition or surrounding situation.

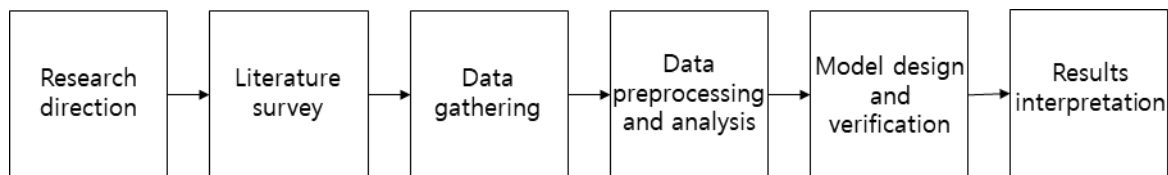


FIGURE 1. Research procedure

2. **Literature Survey.** Chadol used a data set of insurance premium prediction in Kaggle and performed a multiple linear regression analysis with R programming [3]. We used data from the Korea Health Panel Survey ([www.khp.re.kr](http://www.khp.re.kr)) in this study. Hwang used the annual integrated data of the ‘Korea Health Panel Beta Version 1.1.1’ and the ‘2009 Beta Version 1.1.1 of the Korea Health Panel Main Survey’. The data was divided by age into three groups: the Baby Boom Generation (born between 1955 and 1963), the Semi-Aged Generation (born between 1944 and 1954), and the Older Generation (born before 1943). The variables used were gender, education level, household composition, medical insurance type, economic activity, income level, residential area, home ownership status, number of chronic diseases, smoking, drinking, disability, depression experience, and medical expenses (received amount) [4].

Hwang used the ‘Korea Health Panel 2008 Beta Version 1.1.1’ and the ‘Korea Health Panel 2009 Beta Version 1.1.1’. For each factor, the number of people using medical services, frequency, and overall average medical expenses were identified. The variables used were gender, age, household composition, medical insurance type, economic activity, income level, home ownership, number of chronic diseases, smoking, drinking, disability, physical activity, subjective health status, and medical expenses (co-payment plus prescription drug prices) [5].

Jung constructed the 2008 consolidated data by combining data from Korea Health Panel’s surveys of the first half of 2008, second half of 2008, and 2009. In addition, the average medical expenses in 2008 and 2009 for each medical service were investigated. The variables used were gender, age, marital status, disability, economic activity, household income, a total of six variables. Medical expenses were defined as the sum of out-of-pocket expenses and non-benefit expenses [6].

[7] analyzed the frequency and proportion of different variables, showing the status for each type in 2016. [7] primarily analyzed health care expenditure divided into several factors. Medical expenses were divided, based on the reason for spending, into “health care services”, “medicine”, “health care products”, and “others”. Medical expenses were further divided into emergency use, hospitalization use, and outpatient use. The number of uses and expenses according to each type of medical use was analyzed. In addition, an analysis of medical expenses for chronic and severe diseases was conducted. It involved an analysis of the characteristics of medical expenses for each disease of the older population, aged 65 or older and showed the proportion of patients corresponding to each disease.

[8] analyzed the household composition, household and economic activities of 2016 by showing the trend of the number of household members and generational composition of household members’ characteristics. In “disease transfer status and prevalence status”, “disease status”, “major chronic disease prevalence status”, and “yearly prevalence trend for 8 diseases” were divided into gender, age, and number of diseases. “Medical use and medical expenditure” analyzed the medical utilization rate, number of uses, and medical expenditure of the households, and showed the medical expenses and dosage rates for each of the eight diseases. In “health form and health level”, three major health types were selected: “current smoking rate”, “high risk drinking rate”, and “physical activity with severe degree or more”, and the three behaviors were displayed by gender and year. It also analyzed complex health risk behaviors, indicating the proportion of individuals

TABLE 1. Comparison of literature survey (KHP = Korea Health Panel  
ER = Emergency room, I = Inpatient, O = Outpatient)

	Chadol [1]	Hwang 2011 [4]	Hwang 2011 [5]	Jung 2011 [6]	KHPS 2018 [7]	KHPS 2018 [8]	This study
<b>Data</b>	Kaggle	KHP	KHP	KHP	KHP	KHP	KHP
<b>Programming language</b>	R	NA	NA	NA	NA	NA	R
<b>Age</b>	All	65 or more	65 or more	65 or more	All	All	65 or more
<b>EDA type</b>	One variable Freq./Ratio	One variable Freq./Ratio (by ER, I, O)	One variable Freq./Ratio	One variable Freq./Ratio	Different variables Freq./Ratio	Different variables Freq./Ratio	One variable Freq./Ratio (by ER, I, O)
<b>Research method</b>	Estimation in Regression Model (including EDA)	Estimation in Regression Model (including EDA)	EDA	EDA	EDA	EDA	Estimation in Regression Model (including EDA)
<b>Continuity</b>	○	○	×	×	×	×	○

who indulged in smoking, high-risk drinking, and physical activity. Table 1 shows the comparison of the above literature survey and this paper.

**3. Introduction to Target Data and Variables.** This study used the ‘Korea Health Panel 2008-2017 Annual Data (Version 1.6)’. The ‘Korean Medical Panel Survey’ aims to build panel data that can comprehensively analyze factors affecting medical use and medical expenditure, as well as information on medical use patterns and medical expenditure scale. The data and variables used in this study are summarized in Figure 2.

Medical expenses are affected by demographic factors such as gender, age, and household composition, and socioeconomic factors such as home ownership and economic activity. It is also related to health status and health behavior. As aging continues, the level of health deteriorates, increasing the risk of accidents such as falls, and the incidence of chronic diseases. Therefore, independent variables were classified into three categories: ‘demographic factors’, ‘socioeconomic factors’, and ‘health-related factors’. Through literature survey, 12 independent variables and 1 dependent variable (personal medical expenses) were used in this study.

In ER, IN, and OU, which are data sets for emergency, inpatient, and outpatient services, there are ‘total medical expenses’ and ‘prescription drug prices’ variables. The sum of the two variables was defined as personal medical expenses.

Gender (C3) and Year of Birth variables (C4\_0) were obtained from the IND data. Age was created by preprocessing the year of birth variable. In general, the household type of the elderly in Korea tends to be reduced to elderly couple or single. The Household Composition (B3) variable was used to examine the effect of these social phenomena on medical expenditure.

In socioeconomic factors, we tried to find out the effect of medical expenses spent on economic conditions. Medical Insurance Type (C11) and Economic Activity (C24) were obtained from the IND data set. Medical insurance type was divided into two categories: health insurance subscribers and medical benefit recipients, and the others were excluded. Home Ownership Status (B7) variable was obtained from HH data. In a similar study, annual household income variables were classified into 6 categories and applied to each household member. However, the income level and income activity of each household member are different, and it is not used in this study because it is judged that it is unreasonable to call the household income level the household member’s income level.

In health-related factors, Chronic Disease (CD2) was obtained from the CD data, and Smoking (S2), Drinking (S22), Depression (S44), and Physical Activity Restrictions (SH117) were obtained from the APPEN data. For the smoking variable, the variable (S1) used in similar studies was tried to be used, but it disappeared after 2009. Therefore,

the variable (S2) for whether the person currently smokes was used. For drinking status, we used a variable (S22) that defined excessive drinking if more than 7 glasses of soju (5 cans of beer) for men and 6 glasses of soju (4 cans of beer) for women in one drinking session during the last month were used. For Depression (S44), we used a question asking whether the person felt very sad or unhappy enough to interfere with daily life for more than two weeks in a row in the past year. Subjective health status variables were used in similar studies, but after 2013, the variables were changed to categorical rather than numerical. Therefore, we did not use it because it was judged that sufficient information was not provided. While exploring other variables, we thought that there would be a difference in medical expenses depending on whether or not there was a disability. In addition, since the Disability variable was used in a similar study, the Disability (C14) was used.

File name	Content	The variables used
T17IND (Member of household)	Information on members of household and economic activity	Gender (C3), Year of Birth (C4.0), Medical Insurance Type (C11), Economic Activity (C24), Disability (C14)
T17HH (Household)	Household and housing, Income and expenses, Health expenditure information	Household Composition (B3), Home Ownership Status (B7)
T17APPEN (Additional survey on adult household members)	Smoking, Drinking, Mental/Physical health related information	Smoking (S2), Depression (S44), Drinking (S22), Physical Activity Restrictions (SH117)
T17CD (Chronic disease)	Chronic disease, Information on the use of medicines	Chronic Disease (CD2)
T17ER (Use of emergency services)	Information about access to emergency services	Total Medical Expenses (ER26_5), Prescription Drug Price (ER33)
T17IN (Use of inpatient services)	Information on the use of inpatient services	Total Medical Expenses (IN35_6), Prescription Drug Price (IN37)
T17OU (Use of outpatient services)	Information on the use of outpatient services	Total Medical Expenses (OU29_7), Prescription Drug Price (OU35)

FIGURE 2. Target data and variables

#### 4. Target Data Quality Diagnosis and Data Preprocessing.

**4.1. Data quality diagnosis.** In this study, we examined the quality of the target data to confirm the organization of the variables of each data set and the existence of missing values. The number of missing values for economic activity was  $-6$  (under 14 years old), and there was a total of 2,150 missing values. The number of missing values for disability was  $-1$  (non-disabled), and there was a total of 16,163 missing values. Smoking, drinking and depression had 3, 4, and 3 missing values, respectively; while the missing values for physical activity restrictions were  $-1$  (non-applicable), a total for 5,827, and  $-6$  (under 65), a total of 7,155. Missing values of total medical expenses and prescription drug prices in emergency, hospitalization, and outpatient data were excluded or replaced with other values.

**4.2. Data preprocessing.** The age variable was created using the year of birth, and the study was conducted for older adults, aged 65 years and older. In the emergency/hospitalization/outpatient data set, only Household Member Identification Number (PIDWON) and medical service use frequency variables (ERCOUNT, INCOUNT, OUCOUNT) were extracted and set as re\_er, re\_in, and re\_ou data, respectively.

From the chronic disease data, only Household Member Identification Number (PIDWON), Chronic Disease Code (cd1\_1), and Chronic Disease (CD2) were extracted. To remove duplicate rows and determine the number of chronic diseases per individual, re\_cd data was created through the following preprocessing process.

PIDWON	cd1_1	CD2 → chg_cd			Count
1000102	disease1	1 (yes)	mutate() + ifelse()	1	group_by() + summarize() + sum()
1000102	disease2	2 (no)		0	
1000102	disease3	4 (completely cured)		0	

FIGURE 3. Chronic disease data (CD) preprocessing process

IND, HH, APPEM, and re\_cd data were merged to create b1, a collection of independent variables. Gender was categorized as (1) for men and (0) for women. Household composition was divided into four categories: single, married, married with children, and other households. Economic activity, home ownership status, smoking, drinking, disability, physical activity restriction, and depression were divided into yes (1) and no (0). The number of chronic diseases was categorized from 0 to 5 or more. Medical insurance type was divided into health insurance subscribers (1) and medical benefit recipients (0), and those who did not include both were processed as missing values. Personal medical expenditure was defined as the sum of total medical expenses and prescription drug prices. Data representing annual emergencies, hospitalization, and outpatient medical expenses for individual were set as dt\_er, dt\_in, and dt\_ou, respectively. Missing values of each variable were processed as follows.

Total medical expenses variables		Prescription drug prices variables	
Before preprocessing (Existing questions)	After preprocessing	Before preprocessing (Existing questions)	After preprocessing
Effectiveness	Same as before	Effectiveness	Same as before
Not applicable (-1)	NA → median	For free (91)	0
Don't know/No answer (-9)	Amount received in the same row	Medical beneficiary/Health insurance free checkup (0)	Same as before
		Not applicable (-1)	0
		Don't know/No answer (-9)	NA → median

FIGURE 4. Process of deriving missing values of personal expenditure medical expenses

**4.3. Number and rate of medical use by medical service for the elderly over 65 years of age.** IND, re\_er, re\_in, and re\_ou were merged to examine the medical service use rate. The elderly who used the emergency room only once a year accounted for 73.59% of the total, and the maximum number of emergency room visits was 13 times a year. The use of inpatient medical care is the highest for the elderly who were hospitalized once a year (59.82%), with a maximum of 19 times a year. The elderly who used outpatient medical care 14 times a year accounted for 3.22% of the total, while the maximum number of uses was 322 times a year.

**4.4. Status of medical use and average medical expenses for each factor of the elderly over 65 years of age.** By merging the independent variable data b1 and the dependent variable data (dt\_er, dt\_in, dt\_ou), we analyzed the status of medical use and average medical expenditure by emergency, hospitalization, and outpatient use. As for the household composition variable, we predicted that households where a married couple

and their children live together will have the highest number of medical uses and medical expenses spent. However, households with a married couple and their children had the lowest number of medical uses for all three medical services, and the highest medical expenses, similar to those of married couples. Health insurance subscribers spent more on health care than medical benefit recipients. More than half of those aged 65 and over did not have economic activity and elderly people who do not have economic activity pay more for medical care. There are many older people who own homes, and they spend more on medical expenses. These results suggest that owning a home, which is a sign of wealth, is correlated with higher investment in health care. Interestingly, non-smokers or non-drinkers have higher medical expenses. We interpreted this as a variable representing the difference in health perception rather than medical use for the treatment of diseases caused by smoking and drinking. We predicted that physical activity restriction can result in higher medical expenses. Our study showed that physical activity restriction has a high correlation with medical use and expenses.

**5. Multiple Regression.** Regression analysis is a method of verifying how the dependent variable changes with a change in the independent variable. Since there were more than two independent variables used in this study, we analyzed the data using a multiple regression analysis method. To identify only the minimum variables that have an influence on the dependent variable, we used the stepwise selection. In regression analysis, both the dependent and independent variables must be continuous variables. If the independent variable has a categorical variable, the independent variable should be converted into a dummy variable and used for regression analysis. The household composition variable used in this study is a categorical variable. Since R recognizes this as a continuous variable, we converted to a dummy variable consisting of only 0 and 1.

The analysis was divided by three medical services. The significance level, which is the criterion for judging whether the model is appropriate, was selected as 0.05. The effect of the independent variable on the dependent variable was compared using the standardization coefficient. We confirmed that there was no correlation between the independent variables by calculating the VIF and visualizing the correlation matrix heatmap. The model's explanatory power was judged through the adjusted R-squared. In addition, since the Durbin Watson statistic is close to 2, we proved the independence of the residuals.

**6. Conclusion.** Regression analysis by emergency, hospitalization, and outpatient medical services showed that gender, age, and single-person households, which are demographic factors, have a statistically significant effect on medical expenses. Among health-related factors, physical activity restriction and the number of chronic diseases have the greatest influence of the three factors, and the medical expenses are also significant. On the other hand, although socioeconomic factors were statistically significant, their influence was small compared to the other factors. This suggests that people should thoroughly manage their own health, as physical activity restrictions and the number of chronic disease variables greatly affect medical expenses.

The limitations of this study are as follows. One limitation of our study is that it is human data. In general, human data are not regular. Therefore, the adjusted R-squared did not exceed 1%, which was insufficient to prove the explanatory power of the model. Second, we do not consider interactions between independent variables. For some variables, each feature can be combined to have a greater effect on the dependent variable. For example, a more suitable model can be obtained by considering the interaction between the number of chronic diseases and physical activity restriction variables. Third, the factors affecting medical expenses were explored by limiting the scope of individuals. Since there may be various factors such as welfare/medical policy and medical technology, the reliability of the study will be high if various factors are considered. A final limitation

is the narrow range of medical expenditures. To understand the current state of medical expenses more accurately, it is necessary to expand the range of medical expenses including nursing/transportation/ambulance expenses.

#### REFERENCES

- [1] *Almost Half of South Koreans Will Be 65 or Older by 2067*, Yonhap News, 2019.
- [2] *2019 Health Insurance Statistical Yearbook*, Health Insurance Review & Assessment Service, National Health Insurance Service, 2020.
- [3] Chadol, *R Programming Multiple Linear Regression Case with R Programming (Medical Expense Prediction Modeling) (3)*, <http://blog.naver.com/bestinall/221580078436>, 2019.
- [4] Y. Hwang, Analysis of factors affecting medical expenditure by generation, *The 3rd Korea Health Panel Conference*, Seoul, Seoul Foundation of Women & Family, 2011.
- [5] Y. Hwang, Health service utilization and expenditure of the elderly based on KHP, *Health and Welfare Policy Forum*, Korea Institute for Health and Social Affairs, no.182, pp.51-59, 2011.
- [6] Y. Jung, A report on the Korea health panel survey: Health care utilizations and out-of-pocket spending, *Health and Welfare Policy Forum*, Korea Institute for Health and Social Affairs, no.179, 2011.
- [7] *2016 Korea Medical Panel Basic Analysis Report (I) – Household-Burdened Medical Expenses, Medical Use, Private Medical Insurance, and Medical Expenses Analysis by Major Disease*, Korean Health Panel Study, 2018.
- [8] *2016 Korea Medical Panel Basic Analysis Report (II) – Contraction of a Disease, Chronic Disease, Health Behavior and Level of Health*, Korean Health Panel Study, 2018.