

Association Rules of Comorbidities in Dementia by Using Korea National Hospital Discharge In-depth Injury Survey Data

Mijung Kim

Associate Professor, Department of Health Administration, Bucheon University, Korea
lallegro@bc.ac.kr

Abstract

This study aims to find out the associative relationship between dementia and comorbidities. To conduct this study, we used KNHDIS(Korea National Hospital Discharge In-depth Injury Survey) data from 2009 to 2018 provided by the KDCA(Korean Disease Control and Prevention Agency) annually. We used MySQL for data preprocessing and R for data analysis. As a result of applying the Apriori algorithm criteria of support(≥ 0.01), confidence(≥ 0.6), and lift(> 1), seventeen rules related to dementia were discovered. The diseases associated with dementia were diabetes mellitus, hypertension, disorders of lipoprotein metabolism, glomerular disorders in diabetes mellitus, renal diseases, cardiovascular disease, cerebrovascular disease, and other urinary system disorders. This study can be utilized as primary data for the care of patients with dementia and provides implications for improving effective dementia prevention policies.

Keywords: Association Rules, Comorbidities, Data Mining, Dementia

1. Introduction

Korea is the fastest aging country in the world. In 2000, it moved from an ageing society with more than 7% of the population aged 65 and over to an aged society with more than 14% in 2017. The proportion of the elderly population is expected to exceed 20% in 2025(post-aged society), 30% in 2035, and 40% in 2050 [1]. Korea's life expectancy at birth was 83.3 years as of 2019, 2.3 years longer than the OECD average, continuously increasing [2]. As life expectancy increases and the proportion of the elderly population increases, geriatric diseases and dementia have emerged as social problems. In particular, dementia is a condition in which cognitive function is generally and continuously deteriorated as brain function is impaired, resulting in significant disturbances in daily life. Dementia makes it difficult for the patient to maintain human dignity and imposes a tremendous psychological and economic burden on their family members, adding to the socioeconomic burden. According to the CDC(Central Dementia Center), the estimated prevalence of dementia over 65 years of age in 2020 is 10.33%, and the cost of managing dementia from 15.7 trillion won in 2018 is projected to increase nearly tenfold 105.7 trillion won in 2060 [3]. Accordingly, national efforts to

Manuscript Received: February. 25, 2022 / Revised: February. 28, 2022 / Accepted: March. 4, 2022
Corresponding Author: lallegro@bc.ac.kr
Tel: +82-32-610-0832, Fax: +82-32-610-0899
Assistant Professor, Department of Health Administration, Bucheon University, Korea

prevent dementia are continuing, and the Dementia Management Act was enacted in 2011, and a comprehensive dementia management plan was established and implemented every five years. More than 70% of dementia patients have Alzheimer's dementia, followed by other dementias and vascular dementia [3]. For the past 40 years, senile dementia and Alzheimer's disease have been used almost synonymously. Alzheimer's disease is a chronic brain disease in which various cognitive functions, including memory, are gradually deteriorated due to the degeneration of brain cells, leading to disturbances in daily life. Among the top 10 causes of death announced by the National Statistical Office, Alzheimer's disease rose to ninth place in 2018 and seventh place in 2019, maintaining seventh place in 2020 [4]. It was the first time in 2018 that Alzheimer's was included in the top 10 causes of death in Korea since statistics were compiled in 1983.

Additionally, according to the AHA(American Heart Association), the incidence of brain disease worldwide is increasing faster than heart disease. Over the past decade, mortality from Alzheimer's disease and other types of dementia has more than doubled compared to deaths from heart disease [5]. Since dementia is difficult to completely recover, prevention is best, and early detection and active management can slow its progression. Nevertheless, the symptoms of dementia are often exacerbated because proper diagnosis and management are not carried out on time. According to a study, 61% of Alzheimer's dementia patients had 3 or more comorbidities, and the more comorbidities, the greater the impairment in cognition and self-management [6]. That suggests that deterioration of cognitive function in Alzheimer's patients can be prevented by managing comorbid diseases. Comorbidities have significant implications for the care of people with dementia.

Data mining is the process of extracting valuable information by systematically and automatically analyzing statistical rules or patterns from large-scale stored data. Data mining techniques include tracking patterns, classification, association rule analysis, outlier detection, clustering, regression, and prediction. Association rule analysis is a technique for finding patterns between items hidden in large amounts of transaction data, also called market basket analysis. The process of creating association rules consists of two steps. The first step is to create a set of candidate rules based on the frequent itemset. The most common rule generation algorithm is the Apriori algorithm. The second step is to select rules from these candidate rules that show a strong association between items. In this case, support, confidence, and lift are used to evaluate the uncertainty of the rule [7]. Support is the ratio of transactions that contain both A and B in one transaction item. Confidence is the probability that both A and B are included in the transaction included in item A. The lift used to determine whether the generated rule has utility is the ratio of how many times A and B co-occur compared to when A and B are independent.

In the dementia-related factor analysis using NHI(National Health Insurance) data, the risk factors for dementia were age, gender (female), eating habits, drinking, smoking, obesity, blood pressure, disability type, diabetes, hypertension, heart disease, stroke, depression, intracranial injury, and mild cognitive impairment. In men, the diseases that increased the incidence of dementia were hypertension, enlarged prostate, and stroke, and in women, hypertension, stroke, diabetes, and depression [8]. The association analysis between dementia and other diseases using NHI data using the Apriori algorithm, cerebral infarction, chronic kidney disease, sleep disturbance, essential hypertension, gastritis, intervertebral disc disorder, and osteoporosis were strongly correlated with dementia. It was confirmed that it shows [9]. However, it was mentioned as a limitation of the study that the accuracy of the disease code in the health insurance claim data was 70%. The risk factors for dementia were hypertension, diabetes, hyperlipidemia, heart disease, and cerebrovascular disease through multivariate analysis targeting people over 60 years of age [10].

This study aims to discover the relationship between dementia and comorbidities using KNHDIS data from 2009 to 2018. The composition of the thesis is as follows. In Section 1, the theoretical background and previous studies are reviewed. Section 2 describes the analysis methods, and Section 3 describes the results. Section 4

describes the discussion and Section 5 describes the conclusion.

2. Experiments

2.1 Subjects

The subjects were patients with dementia among inpatients. Data from 2009 to 2018 were used among the KNHDIS data provided by the KDCA. The KNHDIS is a nationally approved statistic. The KNHDIS consists of patients discharged from general hospitals with more than 100 beds as the population and is data obtained by extracting sample hospitals and sample patients according to the size of the beds and the survey method [11]. The reason for using this data is that the health care information manager, a disease classification expert, collects the disease code through discharge analysis, so it is more accurate than the health insurance claim data.

2.2 Tools

MySQL, a relational database management tool, was used for data preprocessing, and data analysis was performed in RStudio (R 4.1.2 for windows) to use R. The Arules package was utilized for the association analysis. The ArulesViz package was used to visualize the association analysis.

2.3 Data Preprocessing

The raw data consists of a single table structure composed of various variables for each discharged patient. For efficient data analysis, one table was normalized into a basic table, a diagnostic code table, and a medical practice code table. In order to analyze the association between dementia and comorbid diseases, target data were first extracted. The target data are discharged patients whose dementia was entered either in the principal diagnosis code or other diagnosis code. The data format of the diagnostic code uses the detailed classification code of the KCD-7(the 7th revision of Korean Classification of Causes of Disease) based on the ICD-10(the 10th revision of the International Classification of Causes of Disease). The KCD-7 is a classification system with a hierarchical structure expressed as IS-A. As shown in Table 1, ‘dementia in late-onset Alzheimer’s disease (F00.1)’ corresponds to ‘dementia in Alzheimer’s disease (F00)’, which is ‘organic mental disorder including symptomatic (F00-F09)’. It is a subcategory of ‘mental and behavioral disorders (F00-F99)’ and the major category [12].

Table 1. Example of the basic structure of the KCD-7

Division	Code	Description
Chapter	F00-F99	Mental and behavioral disorders
Blocks of categories	F00-F09	Organic, including symptomatic mental disorders
Three character categories	F00	Dementia in Alzheimer’s disease
Four-character subcategories	F00.1	Dementia in Alzheimer’s disease with late onset

The KCD-7 codes for dementia are F00.0 to F00.9 for dementia in Alzheimer’s disease, F01.0 to F01.9 for vascular dementia, and F03 for unspecified dementia code. A total of 2,263,724 cases of raw data from 2009 to 2018 were received, of which 10,230 were discharged patients with dementia. In the case of dementia patients with cancer, only anatomical codes such as gastric cancer (C16.99) were left, and morphological codes such as adenocarcinomas (M8140/3) were deleted. In addition, type 2 diabetes and unspecified diabetes were treated as one disease code. After extracting the target data, all detailed classification codes were reduced to three-digit sub-classification codes from the front. If it is processed with a detailed classification code, the minimum support may not be satisfied due to the low frequency of transactions. Reducing the data to a subclass

level produces the same subclass code for one patient. In this case, only one code remained, and the rest of the same code was deleted.

2.4 Data Transformation

In order to perform the association analysis algorithm, the itemset must be converted into a list form based on the transaction. That is, on the basis of the discharged patient, the set of sickness codes of the discharged patient should be converted into a list form. The number of transactions converted into list format, that is, discharged patients is 10,172, and the number of disease codes is 50,541 in total. Excluding duplicates, the number of Corporal and Infantry Codes is 917. The transaction data defined in the list format was converted into a matrix in which dementia subjects and comorbidities were treated in rows and columns, respectively, and matrix values were filled with binary values according to the presence or absence of diseases. At this time, in order to increase memory efficiency, a sparse matrix form that does not summarize the number of cases that do not occur is used. The matrix density was 0.54%, and comorbidity occurred in 0.54% of $10,172 \times 917$ cells. The number of comorbid diseases including dementia, was 1,930 cases with four diseases, followed by 1,883 cases with three diseases, 1,730 cases with five diseases, 1,218 cases with six diseases, and 1,088 cases with two diseases. The figure 1 summarizes the data in matrix form.

```
> summary(dementia.trans)
transactions as itemMatrix in sparse format with
10172 rows (elements/itemsets/transactions) and
917 columns (items) and a density of 0.005418364

most frequent items:
  F03    I10    E14    J18    I69 (Other)
10172  4250  2676  1557   994  30892

element (itemset/transaction) length distribution:
sizes
  1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19
150 1088 1883 1930 1730 1218  813  507  330  209   92   80   45   32   25   13    8   11    2
 20   21
   5    1

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  3.000   5.000  4.969  6.000  21.000
```

Figure 1. Summary of the data in matrix

Hypertension (I10), diabetes (E14), pneumonia (J18), sequelae of cerebrovascular disease (I69), urinary tract infection (N39), and heart failure(I50) were the most frequent diseases excluding dementia. Figure 2 shows the top 20 diseases in order of frequency.

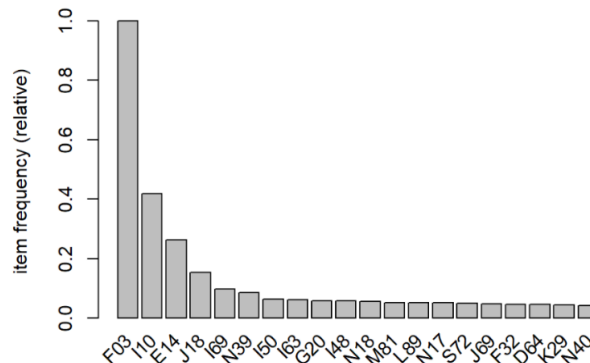


Figure 2. Top 20 diseases in order of frequency

3. Result

The frequent itemsets and association rules were found out using the Apriori algorithm criteria of the minimum support of 0.01 and the minimum confidence of 0.6. One hundred eighty-eight rules were generated, which included 74 diseases. There were 17 of 188 rules with a lift of 1 or greater. Table 3 shows the 17 rules related to dementia order by support. The rule with the strongest association is the rule 1 that diabetes (E14) and dementia (F03) accompany hypertension (I10), with support of 0.1640, reliability of 0.6236, and improvement of 1.4927. That is, the probability that diabetes, dementia, and hypertension will co-occur in all dementia patients is 16.4%, and the probability that dementia patients with diabetes will develop hypertension is 62.4%. In 12 of the 17 rules in the table, a correlation was found with dementia, hypertension, and diabetes. The association between dementia and hypertension was found in 15 rules, and between dementia and diabetes was confirmed in 14 rules, excluding three rules.

Table 3. Association rules related to dementia

No	Rules	Support	Confidence	Lift	Count
1	{E14, F03} => {I10}	0.1640	0.6236	1.4927	1,669
2	{E78, F03} => {I10}	0.0238	0.6377	1.5265	243
3	{E14, F03, I69} => {I10}	0.0205	0.7084	1.6956	209
4	{F03, N08} => {E14}	0.0197	0.9804	3.7270	201
5	{E14, F03, N39} => {I10}	0.0183	0.6515	1.5594	187
6	{E14, F03, N18} => {I10}	0.0174	0.6666	1.5956	178
7	{F03, I20} => {I10}	0.0131	0.6090	1.4578	134
8	{E14, F03, I63} => {I10}	0.0125	0.7111	1.7019	128
9	{F03, N08} => {I10}	0.0122	0.6097	1.4593	125
10	{E14, F03, N08} => {I10}	0.0120	0.6119	1.4646	123
11	{F03, I10, N08} => {E14}	0.0120	0.9840	3.7403	123
12	{E14, F03, N17} => {I10}	0.0120	0.6758	1.6175	123
13	{E14, F03, I48} => {I10}	0.0118	0.6722	1.6089	121
14	{F03, N08, N18} => {E14}	0.0115	0.9915	3.7689	117
15	{E14, E78, F03} => {I10}	0.0112	0.7450	1.7833	114
16	{E14, F03, I50} => {I10}	0.0101	0.6242	1.4940	103
17	{E14, F03, G20} => {I10}	0.0100	0.6666	1.5956	102

4. Discussion

This result supports the latest findings from the AHA. A meta-analysis of 139 studies by the AHA found that middle-aged people with high blood pressure were five times more likely to experience cognitive impairment overall and were twice as likely to experience reduced executive function and dementia and Alzheimer’s disease. Said to be high. Additionally, a meta-analysis of 14 studies found that women with diabetes had a 62% higher risk of dementia and a 58% higher risk for men [5]. Stroke and sequelae of cerebrovascular disease, atrial fibrillation, myocardial infarction, heart failure, urinary tract infection, and hyperlipidemia were associated. It is appropriate to treat those patients with vascular risk factors that meet the criteria for lipid-lowering therapy for the prevention of cardiocerebrovascular accidents [13]. Even in the results of neurocognitive analysis prospectively evaluated in the United States, atrial fibrillation is related to the prevalence of cognitive decline and dementia. As a result of a meta-analysis of 10 prospective studies, coronary artery disease is associated with dementia, cognitive impairment, or cognitive decline. It has been

found to increase the disease risk by 40% [5]. It has been reported that the use of anticoagulants in patients with atrial fibrillation reduces the risk of dementia by 48% [14].

The original result was that the renal complication of diabetes was related to dementia. People with severe diabetes and kidney complications may be at a higher risk of dementia. Further research on this is needed in the future. The limitation of the study was that the number of cases of dementia patients was small to find the rules because department hospitals, nursing hospitals, geriatric hospitals, military hospitals, and rehabilitation hospitals were excluded from the population when the sample hospital was selected. This study is meaningful because the association between dementia and comorbid diseases was identified based on the accurate KNHDIS data of 10 years of disease code. Recently, big data analysis using public data has been actively conducted. Data in the medical field can be used as primary data for human life, life extension, and healthy life, and are very important and valuable data directly related to people's health. For public data to be usefully used for public health, the collection of detailed and accurate data and data quality management should be prioritized above all else.

5. Conclusion

We analyzed the relationship between dementia and comorbidities using the KNHDIS data from 2009 to 2018. This study attempted to discover knowledge about interactions between diseases that were not discovered or discovered in previous studies by using public health data, not epidemiological methods mainly used in the health field. As a result of the analysis, it was identified that there is a strong association between dementia, diabetes, and hypertension. Stroke and sequelae of cerebrovascular disease, atrial fibrillation, myocardial infarction, heart failure, urinary tract infection, and hyperlipidemia were associated. Diabetic kidney disease and renal failure were also associated with dementia. This study may be developed the dementia prevention national health policy so that people can prevent and manage diseases related to dementia and dementia at an early stage on their own. Ultimately, it is expected to contribute to the improvement of national health.

Acknowledgement

This work was supported by 2021 BUCHEON UNIVERSITY Research Grant.

References

- [1] National Statistics Office, *Funeral Population Estimate: 2020-2070*, National Statistics Office, pp.12, 2021.
- [2] MOHW and KIHS, *OECD Health Statistics 2021(Summary) booklet*, MOHW, pp. 10, 2021.
- [3] Central Dementia Center, *Dementia Today*, https://www.nid.or.kr/info/today_list.aspx#a.
- [4] National Statistics Office, *Statistics on Causes of Death*, https://www.kostat.go.kr/portal/korea/kor_nw/1/6/2/index.board.
- [5] C. W. Tsao, A. W. Aday, and Z. I. Almarzooq et al., "Heart Disease and Stroke Statistics-2022 Update A Report From the American Heart Association," *Circulation*, Vol. 145, No. 8, pp. e153–e639, Feb 2022.
DOI: <https://doi.org/10.1161/CIR.000000000001052>Circulation.
- [6] P. M. Doraiswamy, J. Leon, and J. L. Cummings et al., "Prevalence and Impact of Medical Comorbidity in Alzheimer's Disease," *The Journals of Gerontology: Series A*, Vol. 57, No. 3, pp. 173-177, Mar 2002.
DOI: <https://doi.org/10.1093/gerona/57.3.M173>.
- [7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques 3/E*, Morgan Kaufmann Publishers, pp. 184-185, 2011.
- [8] J.H. Kang, *The Factors influencing Dementia using National Health Insurance Claims Data*, MDH. Thesis.

University of Yonsei, Seoul, Korea, 2013.

- [9] J.H. Lee, S.L. Kim, and W.S. Jo, Analysis of Associations between Diseases using Healthcare Public Data Focusing on Dementia,” *Database Research*, vol(33), No(2), pp. 137-148, 2017.
- [10] L. Jia, Y. Du, and L. Chu et al., “COAST Group. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study,” *Lancet Public Health*. Vol. 5, No. 12, pp. e661-e671. Dec 2020.
DOI: [https://doi.org/10.1016/S2468-2667\(20\)30185-7](https://doi.org/10.1016/S2468-2667(20)30185-7).
- [11] Korean Disease Control and Prevention Agency, *2018 Discharge Injury Statistics*, Korean Disease Control and Prevention Agency, pp. 4-5, 2021.
DOI: <https://doi.org/10.1042/CS20160382>.
- [12] National Statistics Office, *Korean Standard Classification of Diseases 7th revision: Volume 1*, National Statistics Office, pp. 295-296, 2015.
- [13] J. P. Appleton, P. Scutt, and N. Sprigg et al., “Hypercholesterolaemia and vascular dementia,” *Clin Sci (Lond)*. Vol. 131, No. 14, pp. 1561-1578. Jun 2017.
DOI: <https://doi.org/10.1042/CS20160382>.
- [14] V. Hachinski, “Dementia: new vistas and opportunities,” *Neurol Sci*. Vol. 40, No. 4, pp. 763-767, Apr 2019.
DOI: <https://doi.org/10.1007/s10072-019-3714-1>.