Computers and Electrical Engineering 70 (2018) 12–20

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Data mining the comorbid associations between dementia and various kinds of illnesses using a medicine database☆

Pei Hua Chen[a], Fu-Yi Yang[b], Ding-Dar Lee[c], Meng-Han Yang[a,*]

[a] Department of Computer Science and Information Engineering, National Kaohsiung University of Applied Sciences, No.415, Jiangong Rd., Sanmin Dist., Kaohsiung 807, Taiwan ROC
[b] Department of Neurology, Taipei Tzu Chi General Hospital, No. 289, Jianguo Rd., Xindian District, New Taipei 23142, Taiwan ROC
[c] Department of Dermatology, Taipei Veterans General Hospital, and Faculty of Medicine, National Yang-Ming University, No. 155, Sec.2, Linong Street, Taipei 11221, Taiwan ROC

ARTICLE INFO

Keywords:
Comorbidity
Data mining
Dementia
Medication
National health insurance program

ABSTRACT

As people's average life expectancy has significantly improved, the prevalence of dementia among elderly people has increased. The symptoms of dementia might influence patients' interpersonal relationships and their ability to work. This study was based on the nationwide dataset applied from the National Health Insurance program of Taiwan, and two types of medication ingredients, which are constantly prescribed for dementia patients, were used for case selection. This study included 5041 case patients and 13,902 control patients who were matched with age, gender, and index date. Using logistic regression, the current study has identified the comorbid associations between dementia and various kinds of illnesses. Advanced stratification analyses revealed that some comorbid illnesses associated with dementia only exists in subgroups of patients with specific age, gender, or prescription for dementia drugs. The mined characteristics would be helpful in managing patients with dementia.

1. Introduction

Today, as people's average life expectancy has improved significantly, the prevalence of dementia among elderly people has increased. Dementia is a syndrome with the following set of symptoms: memory loss and functional degradation in sense of space, judgment, computing ability, abstract thinking, and attention. The patients may have complex behaviors in response to inner stimulus, personality changes, delusions, or hallucinations [1,2]. The severity of such symptoms might influence patients' relationships and their ability to work. The prevalence of dementia increases with age, ranging from 1.43% (65–69 years old) to 12.51% (85 years old) [2]. The WHO points out that there are 47.5 million people with dementia worldwide, growing at an annual rate of 7.7 million people. Therefore, the number of people with dementia doubles every 20 years and would be more than 135.5 million in 2050 [1,3]. The mortality rate for dementia is generally high, with a rate of 20–40% in 2 years. Hence, dementia is not considered a "benign disease" [4], and patients who receive insufficient care have higher mortality risk.

The most common type of dementia is Alzheimer's disease; other common types include vascular dementia, Lewy body dementia, and frontotemporal dementia. A number of factors can help decrease an individual's risk of developing dementia, including early education; being active; not smoking; treating high blood pressure and/or depression; preventing obesity, diabetes, and/or social isolation, etc [5]. Nevertheless, no medications have been shown to cure dementia [6,7]. Psychological therapies for dementia such as

☆ Reviews processed and recommended for publication to the editor-in-chief by guest editor Dr. J-S Sheu.

* Corresponding author.
E-mail address: menghanyang@cc.kuas.edu.tw (M.-H. Yang).

https://doi.org/10.1016/j.compeleceng.2018.05.014
Received 14 September 2017; Received in revised form 14 May 2018; Accepted 14 May 2018

cognitive and behavioral interventions may be appropriate [1]. Exercise programs can also improve dementia patients' ability to perform activities of daily living [8]. However, education and emotional support are important, since comorbid health problems accompanied with dementia carry a substantial, persistent burden of pain [1,9]. Therefore, based on a nationwide population-based medication database with the statistical support, this study aimed to identify various comorbid illnesses associated with dementia that have not been fully studied yet. Those comorbid illnesses include benign prostate hypertrophy, bronchopneumonia, chronic airways obstruction, dyspepsia, functional disorder of the stomach, gastroenteritis or colitis, hematuria, gastrointestinal hemorrhage, pneumonia, pneumonitis, and urinary retention. Advanced stratification studies revealed that some comorbid illnesses associated with dementia only occur in subgroups with specific age, gender, or prescription for dementia drugs. Hence, the mined characteristics would be helpful in managing patients with dementia.

The remainder of this paper is organized as follows. Section 2 describes the materials and methods that were used in this study. Section 3 provides the evaluation outcomes, and Section 4 contains the discussion of the analysis results. Finally, Section 5 draws the conclusion of this paper.

2. Materials and methods

2.1. Data source

The National Health Insurance (NHI) program in Taiwan was launched in 1995. About 23,786,000 of the total population in Taiwan are currently enrolled in this program. In addition, about 28,110 medical hospitals and clinics in Taiwan have joined the program to provide services [10]. The NHI Administration is continuously collecting information of claims documents and data of medical services. In order to mine valuable knowledge from these data, the NHI Administration has set up the "Applied Health Research Data Integration Service" recently, which provides academic researchers to apply for aggregated information. Patients' personal identifiable data have been deleted. The data mainly consist of disease diagnoses, prescription orders, results of medical examinations or laboratory tests, etc. The data (IRB number: 2013-02-007AC) from 1997 to 2010, known as the study period, were obtained in this study. In the NHI program, disease diagnoses are coded by the International Classification of Disease, 9th Revision, Clinical Modification (ICD-9 CM).

2.2. The definitions of case and control patients

Based on the NHI data source, individuals who were prescribed with specific medications for dementia within the study period were categorized as the case patients. Two types of medication ingredients, which were approved by the US Food and drug administration and constantly used as treatment for the cognitive symptoms of dementia, were chosen in this study. The first type of ingredient is the cholinesterase inhibitor (ChEI), which includes donepezil, galantamine, and rivastigmine. Another type of ingredient is N-methyl-D-aspartate, primarily memantine. Using online search to identify the medications contracted with the NHI program [10], 42 different drugs that contain any of the two types of ingredients with 76 updating identification numbers were employed for selecting the case patients. For each of the qualified case patients, the date of the first prescription order was defined as the patient's index date. The accumulated dosages for each specific drug type, with defined daily dose (DDD) as the unit of measurement recommended by the WHO, from all prescription orders within the study period were computed for each case patient.

After selecting the case patients who were prescribed with medications for dementia, their outpatient and/or inpatient visiting records within the study period (1997–2010) were further studied. Case patients were classified into the "first category" if all of their prescription orders for dementia drugs were accompanied with a dementia diagnosis (ICD-9 CM codes: 290.0, 290.1x, 290.2x, 290.3, 290.8, 290.9, and 331.0), "second category" if the prescription orders were not accompanied with any dementia diagnosis, and "third category" if only parts of the prescription orders were accompanied with a dementia diagnosis.

For each of the case patients, three control patients were sampled from the identical data source and matched by age, gender, and year and month of the index dates. All control samples had neither records of dementia diagnosis nor prescription orders for dementia drugs. The case and control patients constituted the study cohort.

2.3. Logistic regression

To measure the occurrence tendency of the target event, logistic regression was formulated as follows:

$$f(z) = \frac{1}{1 + e^{-z}}$$

where z represents the set of independent variables and $f(z)$ as the occurrence probability of the target event. z is defined as the linear combination of k independent variables, which is expressed in the following formula:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k,$$

where β_0 is the intercept and $\beta_1, \beta_2, \ldots, \beta_k$ as the corresponding regression coefficients of independent variables $X_1, X_2, \ldots, X_k$.

If the occurrence probabilities of target events are known in the risk factor group and the control group, respectively, the "odds ratio (OR)" can be formulated as follows: