Osong Public Health and Research Perspectives

Original Article

Changing Disease Trends in the Northern Gyeonggi-do Province of South Korea from 2002 to 2013: A Big Data Study Using National Health Information Database Cohort

Young Soo Kim [a], Dong-Hee Lee [a,*], Hiun Suk Chae [a], Kyungdo Han [b]

[a] *Epidemiology Study Cluster of Uijeongbu St. Mary's Hospital, Uijeongbu St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Uijeongbu, Korea*
[b] *Department of Biostatistics, College of Medicine, The Catholic University of Korea, Seoul, Korea*

ABSTRACT

Objectives: To investigate the chronological patterns of diseases in Northern Gyeonggi-do province, South Korea, and compare these with national data.
Methods: A National Health Insurance cohort based on the National Health Information Database (NHID Cohort 2002-2013) was used to perform a retrospective, population-based study (46,605,433 of the target population, of which 1,025,340 were randomly sampled) to identify disease patterns from 2002 to 2013. Common diseases including malaria, cancer (uterine cervix, urinary bladder, colon), diabetes mellitus, psychiatric disorders, hypertension, intracranial hemorrhage, bronchitis/bronchiolitis, peptic ulcer, and end stage renal disease were evaluated.
Results: Uterine cervix cancer, urinary bladder cancer and colon cancer had the greatest rate of increase in Northern Gyeonggi-do province compared with the rest of the country, but by 2013 the incidence of these cancers had dropped dramatically. Acute myocardial infarction and end stage renal disease also increased over the study period. Psychiatric disorders, diabetes mellitus, hypertension and peptic ulcers showed a gradual increase over time. No obvious differences were found for intracranial hemorrhage or bronchitis/bronchiolitis between the Northern Gyeonggi-do province and the remaining South Korean provinces. Malaria showed a unique time trend, only observed in the Northern Gyeonggi province, peaking in 2004, 2007 and 2009 to 2010.
Conclusion: This study showed that the Northern Gyeonggi-do province population had a different disease profile over time, compared with collated data for the remaining provinces in South Korea. "Big data" studies using the National Health Insurance cohort database can provide insight into the healthcare environment for healthcare providers, stakeholders and policymakers.

Introduction

Understanding the regional trends of major diseases is important because health data for a specific region can aid establishment of effective healthcare policies. Large-scale longitudinal studies are the best available tool to extrapolate data from large districts (urban and rural areas) over time and improve the understanding of various diseases in public health. Although a large cohort study may be the best tool to study regional trends, there is a high proportion of patients lost to follow-up and this is an important consideration. In addition, large cohort studies only include those patients who receive treatment at specialist hospitals. In contrast, the National Health Insurance (NHI) claims records combine the

*Corresponding author: Dong-Hee Lee
Department of Otolaryngology-Head and Neck Surgery, Uijeongbu St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 271 Chunbo-ro, Uijeongbu, Korea
E-mail: leedh0814@catholic.ac.kr

advantage of extended follow-up and the absence of selection bias. Claims data can be analyzed to measure the prevalence of diseases, patterns of healthcare use, clinical outcomes, accessibility of health services, duration of treatment, cost of care, and adherence to good practice guidelines [1-4]. Like the NHI claims database, "big data" is defined as large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of information.

In South Korea, the NHI system includes the health insurance system, which is financed by mandatory contributions, as well as medical aid, a social assistance scheme for the very poor, which is financed by general taxation. Over 95% of all residents in South Korea are covered by the NHI whose claims database includes information about diagnostics, treatments, health service providers, and associated costs, which may be used to study the prevalence of various diseases. However, access to the NHI data of all Korean patients is restricted and only a few studies using this data have been carried out to date [5-15].

The aim of the current study was to investigate chronological patterns of diseases in the Northern Gyeonggi-do province of South Korea between 2002 to 2013, and to evaluate the regional differences in disease patterns between the Northern Gyeonggi-do province and South Korean provinces as a whole.

Materials and Methods

1. Study design and participants

This study was based on NHI data from the National Screening Program. In South Korea, the NHI provides mandatory universal health insurance to nearly all Koreans, whilst the remaining are covered by a public assistance plan (i.e., Medicaid). The Ministry of Health and Welfare entrusts the handling of claims to the NHI, and so the data of Medicaid enrollees are also managed by the NHI. This study excluded the insured employee group from the study population because both employers and employees may incur penalties if they do not undergo health checkups. Because the NHI provided data without identification codes for the dependents of the insured employee group at the beginning of our study, data from these dependents were not analyzed in this study.

This study focused on an area of surveillance consisting of the Northern Gyeonggi-do province of South Korea, including 5 districts, 2 urban (Uijeongbu, Dongducheon) and 3 rural [Pocheon, Yangju, Yeoncheon (Figure 1)]. Using the NHI Cohort Database based on the National Health Information Database (NHID Cohort 2002-2013), a retrospective, population-based study to investigate year-to-year trends of disease patterns between 2002 and 2013 was conducted and differences were
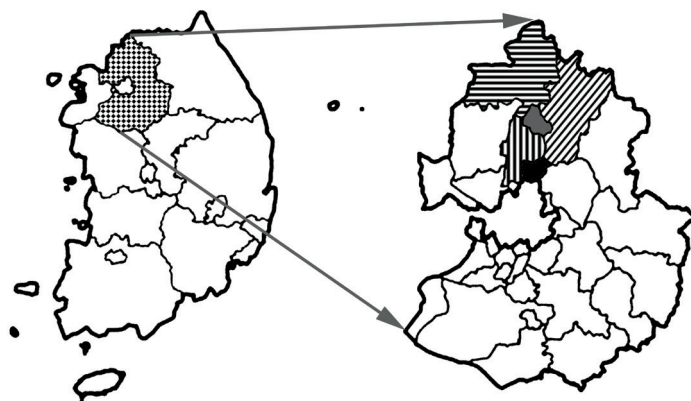

Figure 1. Areas of surveillance included in this study. The left map shows all of South Korea and the dotted area represents the Gyeonggi-do province. In the right map of the Gyeonggi-do province, areas with grayscale shading and cross-hatching represent urban and rural areas, respectively. The solid black area is Uijeongbu, the solid gray shows Dongducheon, vertical stripes show Yangju, the diagonal stripe area is Pocheon, and horizontal stripes show the Yeoncheon district.

evaluated. The NHI provided a cohort of participants who were in health screening programs, called the National Health Insurance Service-Health Screening Cohort. To construct this database, a sample cohort was first selected from the 2002 and 2003 health screening participants, who were aged between 40 and 79 in 2002 and followed up through to 2013.

This study searched the statistical data using the 3-character categories of the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM). ICD-10 codes of the disease used for searching the NHID-Cohort 2002-2013 were listed in Table 1. From an epidemiological and public health perspective, these 12 diseases were selected based on the database of 1 regional center hospital of the Northern Gyeonggi-do province because they were the most common.

2. Data sources

This study used the NHI Cohort Database based on the NHID-Cohort 2002-2013 of the Health Insurance Review and Assessment Service (HIRA) for the period from 2002 to 2013. From the NHID for the year 2001, 46,614,378 NHI claims data were extracted and 1,410,287 NHI claims data were also extracted from the medical aid database for the year 2002. Duplicated data, data of foreigners, data associated with erroneous resident registration numbers and data of patients in the 0 quintile of household income were excluded and a total of 46,605,433 NHI claims data were selected for the sampling population (Figure 2). From this population, using

Table 1. 10th revision, clinical modification codes of the diseases used to search the NHID-Cohort 2002-2013.

| 10th revision, clinical modification codes	Diseases
Certain infections and parasitic diseases	
B50.* to B54.*	Malaria
Neoplasms	
C53.*	Uterine cervix cancer
C67.*	Urinary bladder cancer
C18.*	Colon cancer
Endocrine, nutritional and metabolic diseases	
E10.* to E14.*	Diabetes mellitus
Mental and behavioral disorders	
F00.* to F99.*	Psychiatric disorders
Diseases of the circulatory system	
I10.* to I15.*	Hypertension
I21.* to I22.*	Acute myocardial infarction
I60.* to I62.*	Intracranial hemorrhage
Diseases of the respiratory system	
J20.* to J22.*	Bronchitis/ bronchiolitis
Diseases of the digestive system	
K25.* to K27.*	Peptic ulcer
Diseases of the genitourinary system	
N18.6	End stage renal disease

NHID = National Health Information Database.
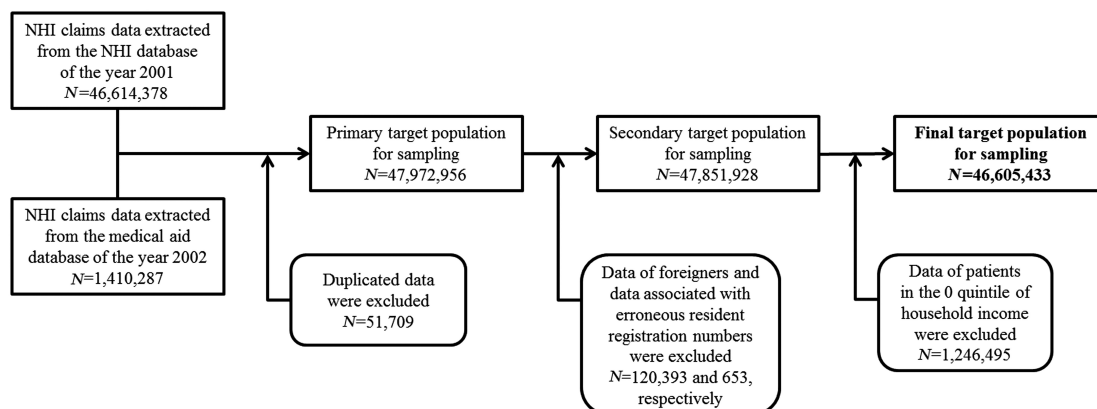
Figure 2. Schematic summary of the construction of the NHID-Cohort 2002-2013 database.
NHID = National Health Information Database; NHI = National Health Insurance.
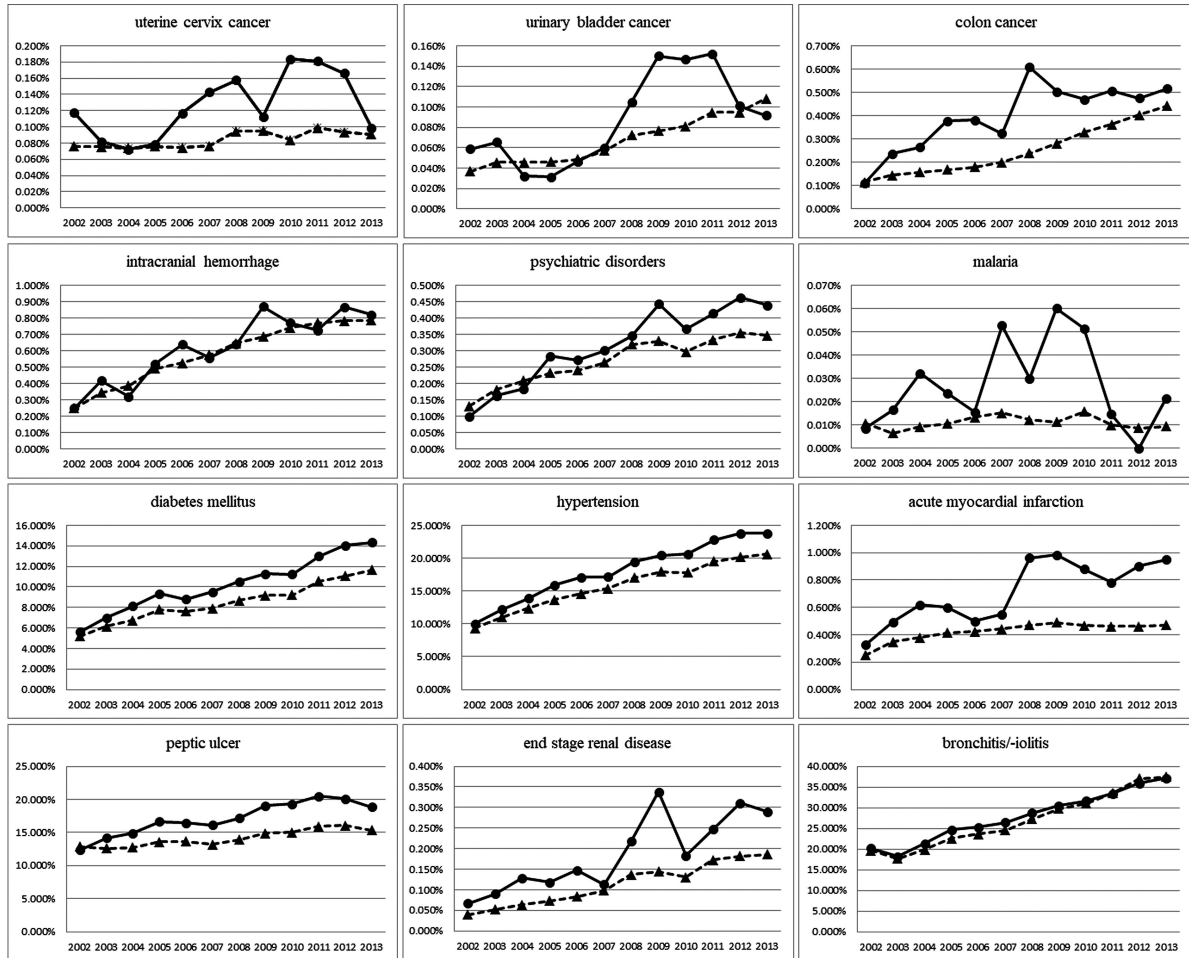
Figure 3. Time trends in the prevalence of 12 diseases using data collated by the NHI Cohort Database based on the National Health Information Database (NHID Cohort 2002-2013), South Korea, 2002 to 2013. Solid lines with circles represent data from the Northern Gyeonggi-do province and dotted lines with triangles show the averages of all other provinces in South Korea.
NHI = National Health Insurance.

proportional quota stratified random sampling, 1,025,340 NHI claims data were sampled to create the NHI cohort database. The NHI claims data were provided by the Korean HIRA, an independent body established to review the claims data and assess the quality of health care in South Korea. As NHI coverage is mandatory, the HIRA-run database contains information concerning all submitted claims and prescriptions for entire beneficiaries of health insurance and medical aid.

3. Statistical analysis

The NHI database includes information on almost the entire population of South Korea, so the assumption was made that sampling errors could be excluded. Accordingly, differences in percentages and rates were calculated without a *p* value.

This study analyzed the prevalence of diseases using the personal health registry data obtained from the NHI and compared the disease prevalence of the Northern Gyeonggi-

do province with the national averages. The prevalence of the diseases among the 5 districts of the Northern Gyeonggi-do province was also compared.

Frequencies and percentile distributions were used to describe categorical variables. Continuous variables were presented as mean ± standard deviation.

4. Ethical consideration

The study protocol was approved by the institutional review board (IRB) of Uijeongbu St. Mary's Hospital, the Catholic University of Korea (IRB No. UC15EISI0003). Informed consent was waived by the IRB.

Results

The time trends of 12 diseases in a cohort group from 2002

Osong Public Health Res Perspect 2018;9(5):248−254

to 2013 are shown in Figure 3. All 12 diseases were diagnosed with increasing frequency in the Northern Gyeonggi-do province compared with the rest of South Korea. There were also several trends unique to individual diseases.

During the 12-year study period, there was a greatly increased incidence in newly-diagnosed cases of uterine cervix cancer, urinary bladder cancer and colon cancer in the Northern Gyeonggi-do province compared with the remaining provinces in South Korea as a whole. However, by 2013, newly-diagnosed cancer cases had dropped markedly, showing similar incidence rates as the rest of the country. Acute myocardial infarction and end-stage renal disease showed variable trends, with a sharp increase in disease prevalence in 2007. Furthermore, as time progressed, the gap between disease prevalence in the Northern Gyeonggi-do province and the rest of South Korea broadened.

More gradual increasing trends over time were seen for psychiatric disorders, diabetes mellitus, hypertension and peptic ulcer.

For intracranial hemorrhage and bronchitis/bronchiolitis, no obvious differences were found between the rates of disease prevalence in Northern Gyeonggi-do province and the rest of South Korea.

In contrast, malaria showed a unique time trend. While in other provinces in South Korea there was no rise in malaria cases, in the Northern Gyeonggi-do province, cases of malaria peaked in 2004, 2007 and 2009 to 2010.

Discussion

A wide variation in health outcomes often exists across different regions of a nation. Rapid growth in healthcare spending and wide regional variations in healthcare expenditure cause the government or healthcare policymakers to consider setting a target healthcare expenditure level for each province [2,16-19]. Therefore, local government and its healthcare policymakers must consider the best strategy to decide which diseases to target when only limited healthcare resources are available. Evaluation of regional differences in the incidence and distribution of disease over time will aid these decisions. This is thought to be the first study that has used "big data" to investigate time trends of diseases nationally, as well as regionally in Northern Gyeonggi-do province using the NHI Cohort Database based on the National Health Information Database (NHID Cohort 2002-2013) during a 12-year period.

By definition, the term "big data" in healthcare refers to electronic health datasets so large and complex, that they are difficult (or impossible) to manage with traditional software and/or hardware, nor can they be easily managed with traditional or common data management tools and methods.

The unique properties of big data are volume, velocity, variety and veracity [20]. Healthcare institutes have generated large amounts of data, driven by record-keeping, compliance and regulatory requirements, and patient care. Whilst most data in the past was stored in hard copy form, the current trend is towards rapid digitization of these large amounts of data. In South Korea, all healthcare institutes have processed medical fees by electronic data interchange since 1998, and these claims data have been recorded as digitized data in the database of the National Health Insurance Service. For more than 10 years, most healthcare institutes have operated the electronic medical record systems, connecting the electronic data interchange system.

It has been reported that big data analytics in healthcare has several advantages in: 1) clinical operations, 2) research and development, 3) public health, 4) evidence-based medicine, 5) genomic analytics, 6) pre-adjudication fraud analysis, 7) device/remote monitoring, and 8) patient profile analytics [21]. The analysis of disease patterns, tracking of disease outbreaks and transmission, can be particularly helpful in the establishment of public healthcare policies, which identify needs, provide services, and predict and prevent crises, especially for the benefit of populations.

This study demonstrated that people in the Northern Gyeonggi-do province have unique patterns of the 12 diseases analyzed compared with the other provinces of South Korea. These results can help healthcare providers or policymakers in establishing healthcare policies that are optimized for the population in the Northern Gyeonggi-do province. The main aim of the study was to provide a comprehensive view of complex usage, requirements, and outcome trends of healthcare, at the local or regional level to governments and healthcare providers. This information can help governments and healthcare providers allocate resources proactively and achieve the best efficacy in outcomes. To do this well, the first step was to aggregate the healthcare-related data comprehensively and analyze them at the level of large populations. These data can be used to reduce waste, target healthcare services more directly to the areas of most need, and redirect spending to effective interventions.

The emergence of malaria in the Northern Gyeonggi-do province was much higher than the national average. This study showed that malaria rates peaked in 2004, 2007 and 2009 to 2010 in the Northern Gyeonggi-do province and also suggested that public healthcare providers should review the reasons for these increases in malaria infections. Korea Centers for Disease Control and Prevention demonstrated that the spatial distribution of malaria cases during 2001 to 2010 was uneven, with the vast majority of cases being recorded in the northern provinces of South Korea, often very close to the South-North Korean border. Many of the malaria cases detected south of
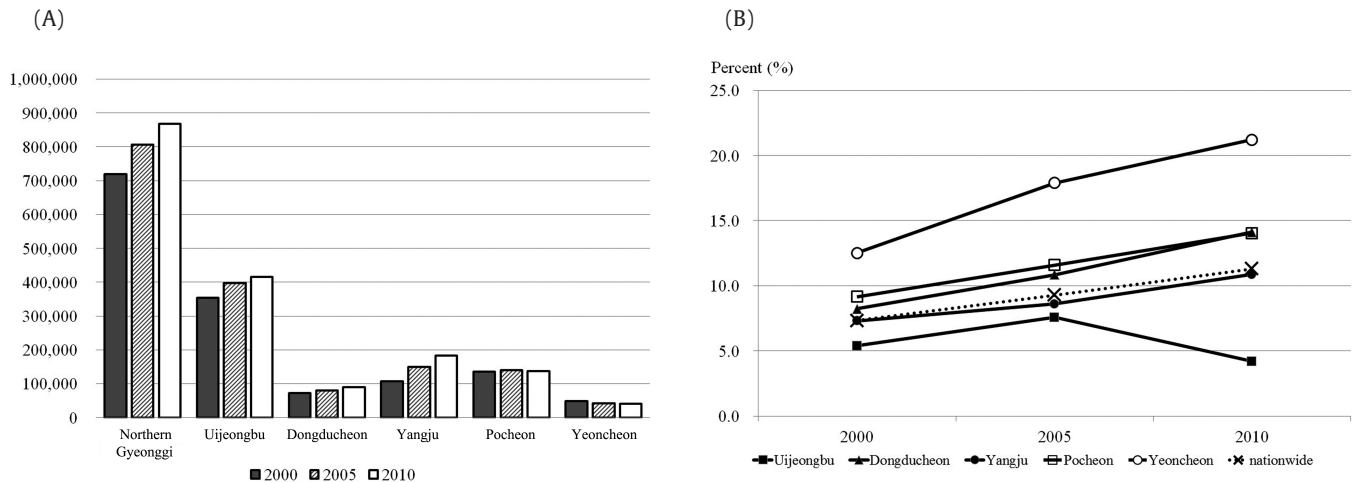
(A)

(B)

Figure 4. Population datasheets for the years 2000, 2005, and 2010. (A) The numbers of people in each district of the Northern Gyeonggi-do province. The population of South Korea was 45,985,289 in 2000, 47,041,434 in 2005, and 47,990,761 in 2010. (B) The percentage of the population that was 65 years and older in each district of the Northern Gyeonggi-do province. This percentage has increased nationally as well as in the Northern Gyeonggi-do province, except for Uijeongbu city. The data were based on the statistical database of Statistics Korea (KOSTAT).

Seoul were attributed to military veterans, < 2 years after separation from the Republic of Korea military. In these cases, malaria developed from exposure near to the demilitarized zone. Approximately 60% of the cases of malaria were due to exposure 6-18 months prior to the onset of symptoms [22-24]. In another study, results indicated that a large percentage of civilians (non-veterans) that were reported to have contracted malaria south of Seoul/Gyeonggi province, were also exposed near to the demilitarized zone. Therefore, some scientists claim that the re-emergence of malaria in South Korea can be attributed to the disease initially being spread from North Korea, and a majority of the recorded cases were from military personnel, who were mainly located close to the South-North Korean border [25]. However, it has recently been reported that the registered number of malaria cases has also increased in the civilian population who live far from the South-North Korean border [24,26]. The results from these studies indicate that the government and healthcare policymakers should focus on prevention and control of malaria in the areas of outbreak. Malaria cases since 2004, are most likely due to environmental factors such as moderate rains that increase *Anopheles* and other mosquito populations, or intense flooding that wash larvae from breeding sites, or semi-drought conditions that result in drying of larval habitats. This suggests that annual trends between the climatic variables and malaria prevalence should be analyzed and collaboration between health and climate governance is imperative.

Generally, there are many diseases that are more prevalent in an aging population. The aging population has increased in the Northern Gyeonggi-do province as well as nationwide. Yeoncheon, Pocheon, and Dongducheon districts have larger aging populations than the national average. However, the increasing rate of the aging population was similar between the Yangju district and the national average. The aging population of the Uijeongbu district was lower than the national average and decreased more in the year 2010 compared with 2005, which is in contrast with other districts and the nationwide trend (Figure 4). Considering these trends, additional factors may be influencing the regional disparity of the disease prevalence in Northern Gyeonggi-do. Therefore, more detailed studies of specific diseases will be needed in the future.

A limitation of our study was that it evaluated the regional patterns of prevalence in the Northern Gyeonggi-do province, and did not explain the exact reasons for the differences in the prevalence. This study used the NHI Cohort Database, therefore, it was impossible to evaluate the incidence or prevalence of the disease in the Northern Gyeonggi-do province. Even though it was based on the database of a regional center hospital of the Northern Gyeonggi-do province, this study investigated the 12 most common diseases treated. This may not necessarily be representative of diseases of the Northern Gyeonggi-do province. Further studies will be needed in the future.

In conclusion, this study revealed that several diseases of the Northern Gyeonggi-do province showed unique and differentiated trends over time compared to other provinces in South Korea. It also demonstrated that a "big data" study using the NHI Cohort Database based on the NHID Cohort 2002-2013 can provide useful insight into the healthcare environment for

healthcare providers, stakeholders, and policymakers.

Conflicts of Interest

No potential conflicts of interest relevant to this article was reported.

Acknowledgments

References

[1] Jee K, Kim GH. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. Healthc Inform Res 2013;19(2):79-85.
[2] Yang HK, Shin DW, Hwang SS, et al. Regional factors associated with participation in the national health screening program: a multilevel analysis using national data. J Korean Med Sci 2013;28(3):348-56.
[3] Ryu S, Song TM. Big data analysis in healthcare. Healthc Inform Res 2014;20(4):247-8.
[4] Song TM, Ryu S. Big data analysis framework for healthcare and social sectors in Korea. Healthc Inform Res 2015;21(1):3-9.
[5] Kang HY, Park CS, Bang HR, et al. Effect of allergic rhinitis on the use and cost of health services by children with asthma. Yonsei Med J 2008;49(4):521-9.
[6] Lim SJ, Kim HJ, Nam CM, et al. Socioeconomic costs of stroke in Korea: estimated from the Korea national health insurance claims database. J Prev Med Public Health 2009;42(4):251-60.
[7] Kang HY, Yang KH, Kim YN, et al. Incidence and mortality of hip fracture among the elderly population in South Korea: a population-based study using the national health insurance claims data. BMC Public Health 2010;10:230.
[8] Cho YS, Choi SH, Park KH, et al. Prevalence of otolaryngologic diseases in South Korea: data from the Korea national health and nutrition examination survey 2008. Clin Exp Otorhinolaryngol 2010;3(4):183-93.
[9] Kim S, Kim J, Kim K, et al. Healthcare use and prescription patterns associated with adult asthma in Korea: analysis of the NHI claims database. Allergy 2013;68(11):1435-42.
[10] Cho SK, Sung YK, Choi CB, et al. Development of an algorithm for identifying rheumatoid arthritis in the Korean National Health Insurance claims database. Rheumatol Int 2013;33(12):2985-92.
[11] Lee T, Kim J, Kim S, et al. Risk factors for asthma-related healthcare use: longitudinal analysis using the NHI claims database in a Korean asthma cohort. PLoS One 2014;9(11):e112844.
[12] Jung HK, Kim YH, Park JY, et al. Estimating the burden of irritable bowel syndrome: analysis of a nationwide korean database. J Neurogastroenterol Motil 2014;20(2):242-52.
[13] Park KH, Lee SH, Koo JW, et al. Prevalence and associated factors of tinnitus: data from the Korean National Health and Nutrition Examination Survey 2009-2011. J Epidemiol 2014;24(5):417-26.
[14] Park RJ, Moon JD. Prevalence and risk factors of tinnitus: the Korean National Health and Nutrition Examination Survey 2010-2011, a cross-sectional study. Clin Otolaryngol 2014;39(2):89-94.
[15] Yeom H, Kang DR, Cho SK, et al. Admission route and use of invasive procedures during hospitalization for acute myocardial infarction: analysis of 2007-2011 National Health Insurance database. Epidemiol Health 2015;37:e2015022.
[16] Tsugawa Y, Hasegawa K, Hiraide A, et al. Regional health expenditure and health outcomes after out-of-hospital cardiac arrest in Japan: an observational study. BMJ Open 2015;5(8):e008374.
[17] Hong JS, Kang HC. Regional differences in treatment frequency and case-fatality rates in korean patients with acute myocardial infarction using the Korea national health insurance claims database: findings of a large retrospective cohort study. Medicine (Baltimore) 2014;93(28):e287.
[18] Wang SY, Wang R, Yu JB, et al. Understanding regional variation in Medicare expenditures for initial episodes of prostate cancer care. Med Care 2014;52(8):680-7.
[19] Lantto M, Renko M, Uhari M. Regional differences in postneonatal childhood mortality in Finland, 1985-2004. Acta Paediatr 2015;104(5):466-72.
[20] Wyber R, Vaillancourt S, Perry W, et al. Big data in global health: improving health in low- and middle-income countries. Bull World Health Organ 2015;93(3):203-8.
[21] Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3.
[22] Lee JS, Kho WG, Lee HW, et al. Current status of vivax malaria among civilians in Korea. Korean J Parasitol 1998;36(4):241–8.
[23] Park JW, Klein TA, Lee HC, et al. Vivax malaria: a continuing health threat to the Republic of Korea. Am J Trop Med Hyg 2003;69(2):159-67.
[24] Kim TS, Kim JS, Na BK, et al. Decreasing incidence of Plasmodium vivax in the Republic of Korea during 2010-2012. Malar J 2013;12:309.
[25] Kho WG. Reemergence of Malaria in Korea. J Korean Med Assoc 2007;50(11):959-66.
[26] Macnee R. Malaria in South Korea: climatic trends and future risk, Final Report, Young Scientist Support Program. APEC Climate Center; 2012.