

# Data Configuration and Publication Trends for the Korean National Health Insurance and Health Insurance Review & Assessment Database

Hae Kyung Kim<sup>1,\*</sup>, Sun Ok Song<sup>2,\*</sup>, Junghyun Noh<sup>3</sup>, In-Kyung Jeong<sup>4</sup>, Byung-Wan Lee<sup>1</sup>

<sup>1</sup>Division of Endocrinology and Metabolism, Department of Internal Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul,

<sup>2</sup>Division of Endocrinology, Department of Internal Medicine, National Health Insurance Service Ilsan Hospital, Goyang,

<sup>3</sup>Division of Endocrinology and Metabolism, Department of Internal Medicine, Inje University Ilsan Paik Hospital, Goyang,

<sup>4</sup>Department of Endocrinology and Metabolism, Kyung Hee University Hospital at Gangdong, Kyung Hee University School of Medicine, Seoul, Korea

**Background:** Big data reports related to diseases and health care for the Korean population have been published since the National Health Insurance Service (NHIS) and the Health Insurance Review & Assessment (HIRA) Service provided limited open access to their databases. Here, we reviewed the structure, content, and means of using data from the National Health Insurance (NHI) system for the benefit of Korean researchers and presented the latest publication trends in Korean healthcare data procured from the NHI and HIRA databases.

**Methods:** Since 2013, researchers have been able to obtain nationwide population-based studies using the NHI and HIRA databases of the insured. We searched publications using the NHI and the HIRA databases between 2013 and 2019 retrieved from PubMed.

**Results:** The NHI and HIRA databases provide nationwide population-based data. The total number of publications from 2014 to 2019 using NHI and HIRA databases is 2,541 and 655, respectively. A total of 5,465 endocrinology-related studies were performed during 2014 to 2019.

**Conclusion:** The NHIS and HIRA databases have provided tools for guidelines to approach world-leading population-based epidemiology and disease research.

**Keywords:** Endocrinology; Health services research; Korea; National health programs

## INTRODUCTION

Following the National Medical Insurance Act legislation in 1963, the current single insurer, the National Health Insurance Service (NHIS), originated from medical insurance societies for companies with more than 500 employees and insurance societies for public officials and private school employees in 1977. Gradually, it expanded to cover the smaller companies and universal coverage of the self-employed by medical insurance societies in rural and urban areas in 1989. Finally, it was

integrated into a single insurer, the National Health Insurance Corporation (NHIC), in 2000 to achieve equity among insurance funds in relation to the financial burden and other managerial issues. In the same year, the separation of prescribing and dispensing drugs was also implemented. The NHIC then changed its name to the NHIS in 2013 [1,2]. In 2008, the long-term care insurance (LTCI) system was introduced to alleviate the financial burden on unpaid family caregivers, helping elderly Koreans with difficulties performing activities of daily living or housework due to geriatric diseases [3].

Corresponding author: Byung-Wan Lee  <https://orcid.org/0000-0002-9899-4992>  
Division of Endocrinology and Metabolism, Department of Internal Medicine, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea  
E-mail: [bwanlee@yuhs.ac](mailto:bwanlee@yuhs.ac)

\*Hae Kyung Kim and Sun Ok Song contributed equally to this study as first authors.

Received: Aug. 18, 2020; Accepted: Sep. 9, 2020

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2020 Korean Diabetes Association <https://e-dmj.org>

In Korea, 97% of the population is obliged to enroll in the National Health Insurance (NHI) program. Patients pay approximately 5% to 30% of the total medical costs to clinics or hospitals, although some services are not covered by insurance, such as cosmetic surgery and some unproven therapies. Clinics and hospitals then submit claims to the Health Insurance Review & Assessment (HIRA) service for inpatient and outpatient care, including data on diagnoses as determined by the International Classification of Diseases, 10th revision (ICD-10), procedures, prescription records, demographic information, and direct medical costs to obtain reimbursement for the total medical costs (ranged from 70% to 95%). The remaining 3% of the population not insured by the NHI program are either covered by another medical aid (MA) program or are temporary or illegal residents [4].

In this study, we reviewed the structure, content, and means of using data procured from the NHI system and HIRA service for the benefit of Korean researchers and presented the latest publication trends for Korean healthcare data procured from the NHI and HIRA databases.

## METHODS

### NHI and HIRA databases

#### Quality of the NHI and HIRA databases

A single insurer, the NHIS, is a nonprofit organization responsible for operating a health insurance program, managing the enrollment of the insured and their dependents, collecting contributions and setting medical fee schedules. With the NHIS

performing compulsory financial collections from the insured based on the law, the data collected from this operation are called “qualification and contribution data” [2].

The HIRA service is a value-based purchasing system for medical service quality improvement providing a review of incurred medical costs and reports from healthcare providers about medical services performed for HIRA. These reports are called “health insurance claims” and contain information on the diagnosis and status of outpatients and inpatients, name, dosage, prescription date and periods, and method of administration of drugs, and laboratory and imaging tests [2].

### NHI and HIRA database contents

NHIS provides various support to accommodate the policy and academic research of health and medicine by using a sample research database, customized database, and health disease index database through the National Health Insurance Sharing Service (NHISS). The HIRA plays roles in reviewing submitted medical fees, claims from clinics and hospitals for reimbursement decisions, and quality of healthcare services to beneficiaries. The claims database of HIRA consists of four detailed categories, including general information on specification or payment specification (20T), consultation statements (30T), diagnosis statements determined by the ICD-10 (40T), and detailed statements about prescriptions (60T) (Table 1). Each data table is designated by the letters 20T, 30T, 40T, and 60T. Each consists of unique number-table delimiter, and T refers to “table.” Payment specification (20T) describes the personal identification, health and medical care institution, principal

**Table 1.** Data characteristics according to the National Health Insurance Service program

20T	30T	40T	60T
Payment specification	Consultation statement	Diagnosis statement	Detail statement of prescription
Personal identification	Medical examination and treatment such as:	Principal diagnosis from 1st to 9th additional diagnoses	Name of drug
Health and medical care institution			Date
Principal diagnosis	Medical care		Filled days
1st additional diagnosis	In-hospital administration of medicine		Supply
Days of medical care	Procedure		Quantity dispensed
Commencement date of medical care	Surgery		Price of each drug
No. of visiting days			
Insurer and deduction payment			

20T, 20 table, consist of unique number delimiter; 30T, 30 table, consist of unique number delimiter; 40T, 40 table, consist of unique number delimiter; 60T, 60 table, consist of unique number delimiter.

diagnosis, first additional diagnosis, days of medical care, date of medical care commencement, number of visits, and insurer and deduction payment, among others. The consultation statement (30T) provides detailed information on medical examination and treatment such as medical care, in-hospital administration of medicine, procedures, and surgery. The diagnosis statement, 40T, includes the principal diagnosis and up to nine additional diagnoses. The detailed statement of prescriptions (60T) for outpatient care provides prescription claims including the following: name of drug, date filled, days of supply, quantity dispensed, and price of each drug [4]. The data in the long-term insurance database include information on the presence of geriatric diseases and the cognitive function of the beneficiaries. Table 2 summarizes the data characteristics according to the NHIS program [2].

The sample research database refers to data standardized to a

shareable form by extracting and deidentifying the sample to improve access and use by investigators because of the large size and personal information issues. It includes a sample cohort database, medical check-up database, elderly cohort database, working women cohort database, and infant medical check-up cohort database. Each database consists of a qualification database, treatment database, medical check-up database, and clinic database. The qualification database includes sex, age, income, region, and types of qualification (employee health insurance, self-employed health insurance, or MA program). The treatment database includes 20T, 30T, 40T, and 60T [5]. The medical check-up database contains major results from medical check-ups and behavior and habitual data from questionnaires including primary general medical check-up data and transition period check-up data [6]. Clinic database is composed of total 10 variables including status, facility, equipment, and personnel

**Table 2.** Health insurance coverage for the Korean population by the National Health Insurance program

Characteristic	Qualification and contribution data	Health insurance claiming data	Health check-up data	Long-term care insurance data
<b>Demographic information</b>				
Sex	○	○	○	○
Age	○	○	○	○
Region	○			○
<b>Family information</b>				
Presence of handicap	○			○
Death	○			○
Type of qualification	○			○
Contribution amount (incomes)	○			○
<b>Medical use</b>				
Medical service use		○	○	
Medical costs		○		
<b>Disease information</b>				
Chronic diseases		○	○	○
Accident/poisoning		○		
Health check-up			○	
Cognitive function			○	○
<b>Lifestyle and habits</b>				
Smoking			○	
Alcohol			○	
Obesity			○	
Exercise			○	
<b>Basic laboratory data</b>				
			○	

data of clinics by type, establishment, and location (city and state). The customized database includes health information data collected, managed, and maintained by the NHIS to be modified as requested for the purpose of policy and academic research. The health disease index database is an index of chronic disease management systematically calculated from the big data of the NHIS including both risk factors of chronic diseases and the processes of occurrence and complications [2].

### **Research samples of NHI and HIRA**

#### **(1) NHI database**

As mentioned in the method section, five types of databases are included in the sample research database. The sample cohort database is a patient sample of 2% randomly extracted from qualified NHI subscribers and MA beneficiaries. The medical check-up cohort database is composed of qualified individuals who were 40 to 79 years old in 2002 to 2003 and received a general medical check-up (approximately 510,000). The elderly cohort database is drawn from the elderly long-term nursing service insurance, targeting qualified individuals over the age of 60 as of 2002 (approximately 550,000). The sample cohort database, medical check-up cohort database, and elderly cohort database included data from 2002 to 2015, a span of 14 years. The working women cohort database consists of working women aged 15 to 64 as of 2007 (approximately 180,000) from 2007 to 2015 (9 years). The infant medical check-up cohort database is a 5% sample of each birth year from 2008 to 2012 for those who received their first or second infant medical check-up from 2008 to 2015 (8 years) [7].

#### **(2) HIRA database**

Four types of patient samples are provided by HIRA: 13% of the national inpatient sample (HIRA-NIS); 3% of the national patient sample (HIRA-NPS); 20% of the aged population sample (HIRA-APS); and 10% of the pediatric patient sample (HIRA-PPS) [8]. The patient samples from 2009 to 2015 are available, but a longitudinal study cannot be performed because of the inconsistency in data collection from individuals as each sample is cross-sectional.

### **Accessing NHI and HIRA databases and protection of personal information**

To request NHI database data, Korean researchers can access raw big data by visiting seven research centers located nationwide or by remote access. Korean researchers are required to

complete a study protocol with approval from an Institutional Review Board (IRB). The application for the data-request form is then reviewed within 25 days from the date of submission by the data provision review committee. After approval by HIRA committees, the HIRA extracts data from the data warehouse system, which is then only accessible to individual researchers for study [8]. Also through the healthcare big data open system (<http://opendata.hira.or.kr>), summary statistics related to healthcare services including prescriptions, medical conditions, and healthcare providers are publicly available.

According to the “Act on the Protection of Personal Information Maintained by Public Agencies,” the NHIS and HIRA provide data without individual identifiers by using an unidentifiable code representing each individual. There are data concerning the patient’s age, sex, diagnosis, and lists of prescribed drugs [2].

### **Journals sources and search strategies**

As we described, the NHI and HIRA databases represent the entire Korean population because every Korean resident is universally eligible for the NHI program. Because the integration of information of NHI and MA beneficiaries in HIRA and NHI databases was introduced in 2006 and the government initiative promoting “Opening and Sharing Big Data for Value Creation” has been available since 2013, we have been able to use the NHI and HIRA databases for population-based and nationwide research on various diseases since 2013 [8]. To identify articles using either the NHI or HIRA databases, we queried the literature using PubMed with various combinations of terms from January 1, 2014 to December 31, 2019. We used terms for the literature search in the dataset of NHIS that were a combination of “Korea” [affiliation] and “NHIS” or “national health insurance” [all fields]. Combinations of the words “Korea” [affiliation] and “HIRA” or “Health Insurance Review and Assessment” [all fields] were used to search literature using the dataset released from the HIRA. Then from the literature extracted, the number of papers in each field published each year was calculated using the text words “diabetes,” “metabolism,” “endocrinology,” “osteoporosis,” “thyroid,” “bone,” “pituitary,” and “adrenal.” In addition, languages were limited to English. The study protocol received ethical approval from the IRB at the Yonsei University College of Medicine (No. 4-020-0778), which waived the need for informed consent because of the descriptive nature of this study and no personal information included.

## RESULTS

### Korean population served by the National Health Insurance system

Table 3 shows 100% of the Korean population is covered by the NHI program. In 2019, the NHI system covered 97.2% ( $n=52,880,000$ ) of the population including employee insured (70.4%,  $n=37,227,000$ ) and self-employed insured (26.8%,  $n=14,164,000$ ), and the MA system covered the remaining and 2.8% ( $n=1,485,000$ ) of the population. Compared to 2013 and 2016, the proportion of health insurance coverage by NHI either employee or self-employed insured and MA beneficiaries in 2019 has not changed significantly.

### Recent publications associated with NHI and HIRA data

According to trend analysis of NHI and HIRA database research, domestic and international research-project papers are increasing (Fig. 1A). Research on various topics related to health care, public health policies, and diseases including cancer, infection, heart diseases, cerebrovascular diseases, hypertension, diabetes mellitus, and endocrine diseases has been

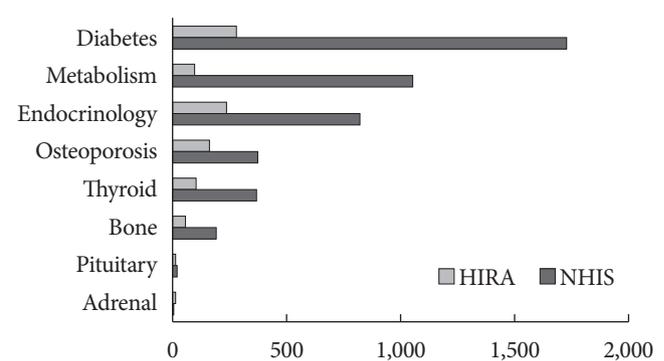
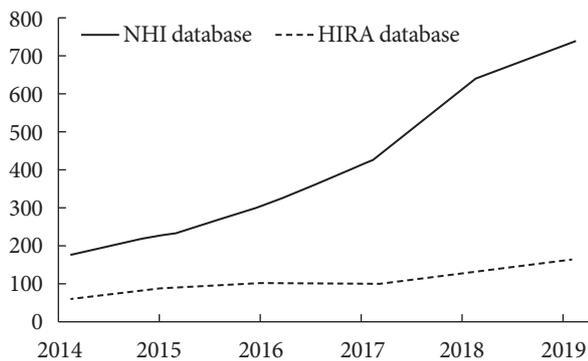
published. The total number of publications from 2014 to 2019 using the NHI and HIRA databases was 2,541 and 655, respectively. Published studies using the NHI and HIRA databases dramatically increased from 178 in 2014 to 743 in 2019 and from 60 to 165, respectively. Since 2014, publications related to endocrinology using NHI and HIRA database have also increased (Fig. 1B). A total of 5,465 studies were included according to the search method of our study using the terms of “diabetes,” “metabolism,” “endocrinology,” “osteoporosis,” “thyroid,” “bone,” “pituitary,” and “adrenal.” Of the endocrinology-related journals, 83.1% of the studies adopted the NHI database. Diabetes-related big data studies accounted for 36.7%. Metabolism-, osteoporosis-, thyroid-, bone-, pituitary-, and adrenal-related research comprised 21.0%, 9.7%, 8.5%, 4.3%, 3.5%, and 0.1%, respectively.

## DISCUSSION

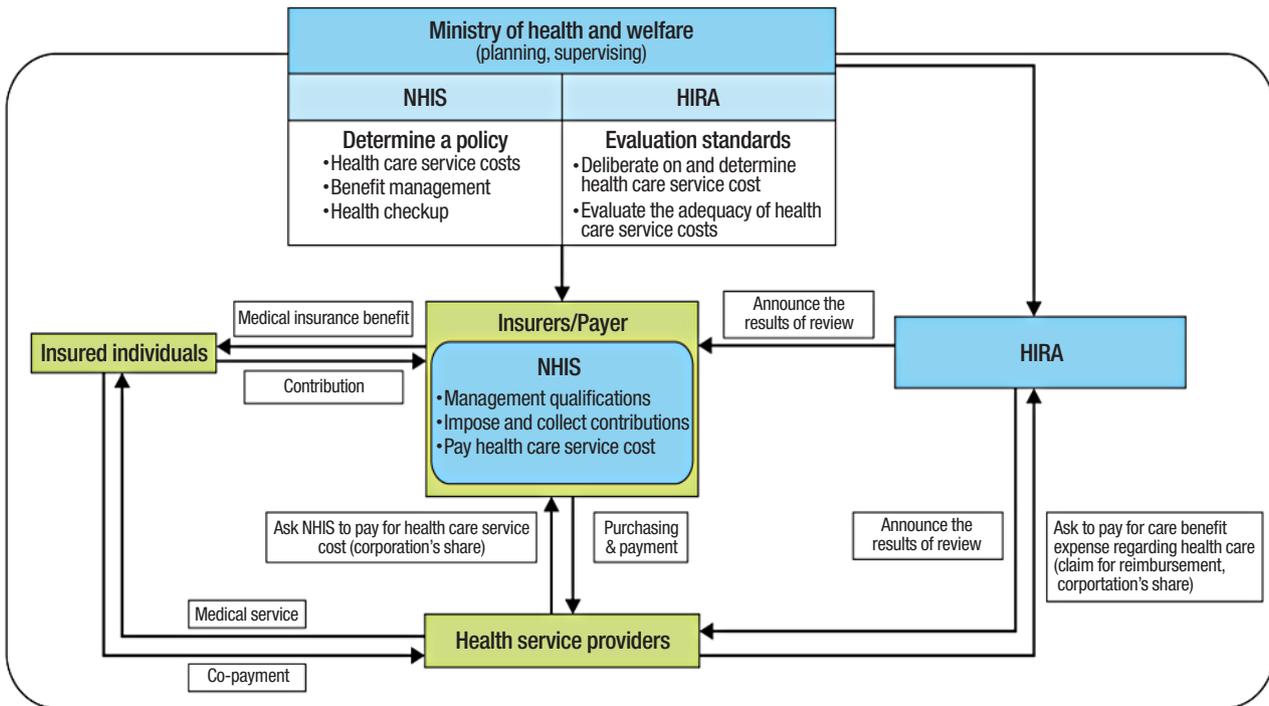
Big data studies on Korean public healthcare using the NHI and HIRA databases are becoming increasingly popular. This phenomenon could be attributed to the strengths of the com-

**Table 3.** Components of claims data in the National Health Insurance Service database

Variable	Coverage population, 1,000 persons (%)		
	2013	2016	2019
Total	51,449 (100.0)	52,273 (100.0)	52,880 (100.0)
National Health Insurance			
Employee insured	35,006 (68.0)	36,675 (70.2)	37,227 (70.4)
Self-employed insured	14,984 (29.2)	14,089 (26.9)	14,164 (26.8)
Medical aid beneficiaries	1,459 (2.8)	1,509 (2.9)	1,489 (2.8)



**Fig. 1.** (A) Year-specific number of publications using the National Health Insurance (NHI) and Health Insurance Review & Assessment (HIRA) databases (2014 to 2019). (B) Endocrinology-related search terms in the NHI and HIRA databases (2014 to 2019).



**Fig. 2.** How the National Health Insurance system works. NHIS, National Health Insurance Service; HIRA, Health Insurance Review & Assessment.

pulsory nature of the NHI system providing universal coverage for the Korean population and consequently the serial population data that are readily available. In this study, we reviewed the structure, content, and means of using data procured from the Korean NHI system by Korean researchers, especially endocrinologists, and we present the latest publication trends on Korean healthcare.

Fig. 2 illustrates the operational structure of the NHI system. Three key regulators of the healthcare system include the Ministry of Health and Welfare (MOHW), NHIS, and HIRA. The MOHW supervises the NHI program through the formulation and implementation of policies. The NHIS is a nonprofit organization and the single insurer that manages the NHI program. It is responsible for managing enrolled and insured individuals and their dependents (spouse, direct lineal ascendants or descendants, and unmarried brothers or sisters) and collecting contributions and setting medical fee schedules. The HIRA has a quality control role, evaluates healthcare performance, and reviews medical billing and claims. The HIRA also determines whether health care services are medically necessary and ensures that the services are delivered to beneficiaries at an appropriate level and cost [5]. The NHI, MA, and LTCI are the

three main health care programs for universal coverage in Korea. The NHI program classified covers the whole population as either employee or self-employed as a social insurance benefits scheme with the following features: short-term insurance and compulsory contributions made according to ability to pay. The contribution to NHI is calculated based on an employee's wage and is paid by the company's employer. In the case of the self-employed, the cost is calculated based on the household income, property, income, vehicle(s) owned, age, and gender. The MA program is managed by the Korean government. It is a public assistance scheme that secures the minimum livelihood of low-income households and assists with self-help by providing medical services. The LTCI program is based on the principle of social solidarity. The program provides benefits for at least 6 months. Although NHI covers the whole population as a compulsory scheme, not all items of healthcare are covered by the program. The aim is to cover the prevention and treatment of sickness and injury that result from daily life and childbirth. The program also covers health promotion and rehabilitation. The NHI provides the same benefits package regardless of the contributions made that are determined by the individual's ability to pay.

NHI and HIRA databases were used in various research areas including healthcare and public health policies, medical adherence, prescribing patterns, adverse events, cost-effectiveness, burden of disease, healthcare service utilization, disease incidence and prevalence, and outcomes. As availability and accessibility of NHI and HIRA data have grown in recent years, the data provision review committee has granted use to over 3,000 studies from 2014 to 2019 [9]. In this study, 83.1% of endocrinology-related big data studies used the NHI database. Despite the similarity of the two databases, their content is different, where the NHI database's main sections include healthcare use, sociodemographic variables, health screening, and mortality, whereas the HIRA research database's main sections include general demographic characteristics, healthcare use, diagnoses, and outpatient prescriptions [10]. Additionally, NHI sample databases include longitudinal cohorts, enabling researchers to perform long-term follow-up studies, whereas the HIRA sample databases include separate cohorts for each year, only suitable for conducting a cross-sectional study or short-term follow-up (less than 1 year) studies. This is because patients in the HIRA sample databases are stratified and re-sampled annually, and patient information cannot be linked across years within the HIRA sample database [11].

Concerning the strengths and limitations of the NHI and HIRA databases, the NHI database covers detailed information regarding medical examinations, therapeutics including prescription and long-term follow-up of the insured. This allows the researchers to perform longitudinal studies whereas the HIRA database cannot follow individuals. As we described earlier, HIRA data covers healthcare service records from neonates to the elderly in a full range of healthcare settings. This representative and comprehensive dataset broadens Korean research to fields that might not be easily measured using randomized controlled trials by providing demographic information. Using longitudinal data collected from NHI and MA beneficiaries, researchers can conduct cohort studies and investigate long-term outcomes. Furthermore, healthcare agencies provide information to the HIRA, which ensures the reliability of the data, and the vast sample size secures statistical power for complicated analyses [8]. However, as in the HIRA database, the discrepancy between real-world diagnosis of the insured and data collected from the NHI database, the validity of the research might be compromised. Therefore, the researchers should properly extract samples for the studies by

using appropriate definitions and carefully designed inclusion criteria to avoid misleading results. Despite the strengths of the HIRA database, it lacks information such as the severity of diseases and personal health history such as smoking or alcohol consumption. This personal information might interfere with the main outcomes of the research [10]. Also, uncovered healthcare services such as cosmetic surgeries are not included, and discrepancies between diagnosis information and actual health conditions may impose bias to compromise a study's validity [10]. Nonetheless, the benefits overcome the limitations of the NHI and HIRA databases, and these limitations could be further addressed by strategies for the optimal use of the NHI and HIRA databases.

In conclusion, the NHI and HIRA databases represent the entire Korean population and can be used as a population database. The NHIS and HIRA databases have been important resources for endocrinology research and have provided guidelines for Korean endocrinologists to conduct world-leading population-based epidemiology and disease research.

## CONFLICTS OF INTEREST

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

## AUTHOR CONTRIBUTIONS

Conception or design: B.W.L.

Acquisition, analysis, or interpretation of data: H.K.K., S.O.S.

Drafting the work or revising: J.H.N., I.K.J., B.W.L.

Final approval of the manuscript: B.W.L.

## ORCID

Hae Kyung Kim <https://orcid.org/0000-0002-5130-4245>

Sun Ok Song <https://orcid.org/0000-0003-4829-3407>

Byung-Wan Lee <https://orcid.org/0000-0002-9899-4992>

## ACKNOWLEDGMENTS

The authors thank Medical Illustration & Design, part of the Medical Research Support Services of Yonsei University College of Medicine, for all artistic support related to this work.

## REFERENCES

1. Jeong HS. Korea's National Health Insurance: lessons from the past three decades. *Health Aff (Millwood)* 2011;30:136-44.
2. Song SO, Jung CH, Song YD, Park CY, Kwon HS, Cha BS, Park JY, Lee KU, Ko KS, Lee BW. Background and data configuration process of a nationwide population-based study using the Korean national health insurance system. *Diabetes Metab J* 2014;38:395-403.
3. Kwon S. Future of long-term care financing for the elderly in Korea. *J Aging Soc Policy* 2008;20:119-36.
4. Jung CH, Seo GH, Suh S, Bae JC, Kim MK, Hwang YC, Kim JH, Lee BW. The population-based risk of need for coronary revascularization according to the presence of type 2 diabetes mellitus and history of coronary heart disease in the Korean population. *PLoS One* 2015;10:e0128627.
5. Park YT, Yoon JS, Speedie SM, Yoon H, Lee J. Health insurance claim review using information technologies. *Healthc Inform Res* 2012;18:215-24.
6. Seong SC, Kim YY, Park SK, Khang YH, Kim HC, Park JH, Kang HJ, Do CH, Song JS, Lee EJ, Ha S, Shin SA, Jeong SL. Cohort profile: the National Health Insurance Service-National Health Screening Cohort (NHIS-HEALS) in Korea. *BMJ Open* 2017;7:e016640.
7. National Health Insurance Sharing Service: Terminology. Available from: <https://nhiss.nhis.or.kr/bd/ab/bdaba012eng.do> (cited 2020 Sep 12).
8. Kim JA, Yoon S, Kim LY, Kim DS. Towards actualizing the value potential of Korea health insurance review and assessment (HIRA) data as a resource for health research: strengths, limitations, applications, and strategies for optimal use of HIRA data. *J Korean Med Sci* 2017;32:718-28.
9. Chung H, Kim SY, Kim HS. Clinical research from a health insurance database: practice and perspective. *Korean J Med* 2019;94:463-70.
10. Kim L, Kim JA, Kim S. A guide for the utilization of health insurance review and assessment service national patient samples. *Epidemiol Health* 2014;36:e2014008.
11. Kim S, Kim MS, You SH, Jung SY. Conducting and reporting a clinical research using Korean healthcare claims database. *Korean J Fam Med* 2020;41:146-52.