Brief Report

Epidemiology

dmj

DIABETES & METABOLISM JOURNAL

Data Analytic Process of a Nationwide Population-Based Study Using National Health Information Database Established by National Health Insurance Service

Yong-ho Lee[1], Kyungdo Han[2], Seung-Hyun Ko[3], Kyung Soo Ko[4], Ki-Up Lee[5]; on Behalf of the Taskforce Team of Diabetes Fact Sheet of the Korean Diabetes Association

[1]Department of Internal Medicine, Yonsei University College of Medicine, Seoul,
[2]Department of Biostatistic, The Catholic University of Korea, Seoul,
[3]Division of Endocrinology and Metabolism, Department of Internal Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Suwon,
[4]Department of Internal Medicine, Cardiovascular and Metabolic Disease Center, Inje University Sanggye Paik Hospital, Inje University College of Medicine, Seoul,
[5]Department of Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

In 2014, the National Health Insurance Service (NHIS) signed a memorandum of understanding with the Korean Diabetes Association to provide limited open access to its databases for investigating the past and current status of diabetes and its management. NHIS databases include the entire Korean population; therefore, it can be used as a population-based nationwide study for various diseases, including diabetes and its complications. This report presents how we established the analytic system of nationwide population-based studies using the NHIS database as follows: the selection of database study population and its distribution and operational definition of diabetes and patients of currently ongoing collaboration projects.

Keywords: Diabetes mellitus; Korea; National Health Insurance Service; National survey

OVERVIEW: NATIONAL HEALTH INFORMATION DATABASE ESTABLISHED BY NATIONAL HEALTH INSURANCE SERVICE

In 2014, the National Health Insurance Service (NHIS) signed a memorandum of understanding with the Korean Diabetes Association (KDA) to provide limited open access to its databases for investigating the past and current status of diabetes and its management. A previous review by Song et al. [1] described in detail regarding the history, structure, contents, and way to use data procurement in the Korean National Health Insurance (NHI) system. Briefly, the NHIS in Korea is a single-payer organization that is mandatory for all residents in Korea. Because it has adopted a fee-for-service system to pay health care providers who treat or examine Korean patients, NHIS obtains information on patient demographics, medical use/transaction information, insurers' payment coverage, and patients' deduction and claim database (diagnosis/prescriptions/consultation statements). The NHIS database represents the entire Korean population; therefore, it can be used as a population-based, nationwide study for various diseases. Recently, several epidemiologic studies with a large population using the NHIS database have been reported [2].

Corresponding author: Kyung Soo Ko http://orcid.org/0000-0002-8838-9094
Department of Internal Medicine, Cardiovascular and Metabolic Disease Center, Inje University Sanggye Paik Hospital, Inje University College of Medicine, 1342 Dongil-ro, Nowon-gu, Seoul 01757, Korea
E-mail: kskomd@paik.ac.kr

DATABASE POPULATION

A single insurer, NHI system consists of two major health care programs for universal coverage of all residents of Korea: NHI and Medical Aid (MA). Approximately 97% of the population is covered by NHI, and the remaining 3% of the population is covered by MA. Since 2006, information of MA beneficiaries has been incorporated into a single NHIS database. Therefore, the NHIS database during 2002 to 2005 included only information of NHI beneficiaries, but not MA beneficiaries, which should be considered in caution when interpreting the findings in this period. Retrospective data of individuals aged more than 30 years were extracted using the Korean NHIS database from January 2002 through December 2013.

DATABASE CONTENTS

Among sub-datasets of the NHIS database, we used Qualification DB, Claim DB, Health Check-up DB, and death information.

Qualification DB includes sex, age, income, region, and types of qualification. Using this database, we showed the distribution of study participants aged more than 30 years from the NHIS database from 2002 to 2013 by gender and age (Table 1).

Claim DB includes general information on specification (20T), consultation statement (30T), diagnosis statements defined by the International Classification of Diseases 10th revision (ICD-10; 40T), and detailed statements about prescriptions (60T). Detailed information is provided by a previous

review paper [1].

Health Check-up DB generally consists of four areas: general health check-up, lifetime transition period health check-up, cancer check-up, and baby/infant health check-up [3]. Among them, we used the database from the general health check-up, which includes (1) employee subscribers and regional insurance subscribers who are a regional householder, (2) employee subscribers' dependent and household members (40 years or older), and (3) MA beneficiaries who are a householder of 19 to 64 years of age and household members of 41 to 64 years. All examinees were requested to have biannual health check-ups, except non-office workers who are employee subscribers (annual). The proportion of complete health check-ups was approximately 40% in 2002, whereas it increased up to 68% in 2013.

DEFINITION OF DIABETES

Considering the characteristics of the NHIS database, an operational definition of diabetes was applied for further analysis. For Claim DB, individuals having diabetes were defined if anti-diabetic drugs were prescribed with the presence of ICD-10 codes E11, E12, E13, or E14, as either principal diagnosis or 1st to 4th additional diagnosis at least once a year. For the Health Check-up DB, patients with fasting glucose levels ≥126 mg/dL were considered as having diabetes. This operational definition of diabetes was concluded by the following data analysis using study participants in the 2013 NHIS database.

Table 1. Distribution of Korean National Health Insurance beneficiaries aged ≥30 years (study participants) by gender and age (unit: 10,000 people)

	Year											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Total	2,656	2,728	2,808	2,872	2,934	2,996	3,078	3,145	3,211	3,276	3,340	3,390
Sex												
Male	1,302	1,336	1,376	1,407	1,436	1,465	1,506	1,539	1,572	1,604	1,636	1,659
Female	1,354	1,392	1,432	1,465	1,498	1,531	1,573	1,606	1,640	1,672	1,704	1,731
Age, yr												
30–39	873	883	888	883	878	867	858	849	842	833	827	811
40–49	771	796	820	831	842	850	871	879	878	881	882	891
50–59	446	461	486	523	554	583	618	656	701	748	777	801
60–69	345	356	364	366	373	387	401	410	419	422	433	448
70–79	160	168	181	196	211	226	240	255	267	282	302	312
≥80	60	64	68	72	77	84	90	96	103	110	118	127

Table 2. Distribution of Korean National Health Insurance beneficiaries aged ≥30 years and General Health Check-up examinees by age in 2013 (unit: 10,000 people)

	NHI beneficiaries aged ≥30 years	General Health Check-up examinees
Age, yr		
30–39	811 (23.9)	212 (20.0)
40–49	891 (26.3)	305 (28.7)
50–59	801 (23.6)	278 (26.2)
60–69	448 (13.2)	161 (15.2)
70–79	312 (9.2)	88 (8.3)
≥80	127 (3.7)	17 (1.6)
Total	3,390 (100)	1,061 (100)

Values are presented as number (%).
NHI, Korean National Health Insurance.

Table 2 shows numbers of NHI beneficiaries aged ≥30 years and General Health Check-up examinees in 2013 by age. If the prevalence of diabetes is calculated among NHI beneficiaries aged ≥30 years based on either prescription of anti-diabetic drugs (insulins, sulfonylureas, metformin, meglitinides, thiazolidinediones, dipeptidyl peptidase-4 inhibitors, and α-glucosidase inhibitors), ICD-10 codes, or both, there are substantial discrepancies by the different categories. The proportion of patients with diabetes was 8.50% (3.95%+4.55%) defined by combination of diagnosis and prescription data, whereas it was 13.15% (4.14%+0.51%+3.95%+4.55%) by ICD-10 codes alone and 8.69% (0.12%+0.07%+3.95%+4.55%) based on prescription data alone (Table 3). The Taskforce Team concluded the operational definition of diabetes as either (1) patients who had both data of diagnosis and prescription of anti-diabetic drugs or (2) patients whose fasting glucose levels from Health Check-up DB are more than 126 mg/dL. According to this definition, the prevalence of diabetes was 11.40% (2.32%+0.07%+0.51%+3.95%+4.55%).

ANALYTIC METHODS IN 16 COLLABORATION PROJECTS

Since 2014 when the KDA and NHIS signed an agreement for open access to the NHIS database, 16 subjects of collaboration projects are currently ongoing. Depending on the project objectives, different operational definitions have been applied to diagnose diseases. Overall projects can be divided into two categories: analyses of disease status and analyses of casual relationships among diseases, management, and drugs. Analyses

Table 3. Difference in number of patients with diabetes according to the criteria categories

ICD-10	Drug prescription	Fasting glucose level ≥126 mg/dL	No. (%)
No	No	No	15,134,844 (84.34)
No	No	Yes	415,763 (2.32)
No	Yes	No	22,212 (0.12)
Yes	No	No	743,248 (4.14)
No	Yes	Yes	11,747 (0.07)
Yes	No	Yes	90,878 (0.51)
Yes	Yes	No	708,795 (3.95)
Yes	Yes	Yes	816,479 (4.55)

ICD-10, International Classification of Diseases 10th revision.

of disease status demonstrated the annual prevalence and age-standardized prevalence of specific diseases, such as diabetic nephropathy. Analyses of casual relationships (e.g., effects of anti-diabetic drugs on cancer, association between diabetes, and percutaneous coronary intervention) applied the study design with washout periods and Cox-hazard regression models.

VALUE AND CHALLENGES OF NHIS DATABASE

The NHIS database represents the entire Korean population; therefore, it can be used as a population-based nationwide study for various diseases. Because it contains detailed information regarding statement of prescriptions and medical examination or treatments, such as medical care and in-hospital administration of medicine, procedures, and surgery, investigation of the trends or status of specific diseases is feasible. Furthermore, long-term follow-up of a single individual can allow us to perform longitudinal studies of casual relationships. By combining laboratory and standard questionnaire information from the Health Check-up DB, limitation of Claim DB (without having any laboratory or personal history data) can be overcome.

Despite its strengths, one of the most critical drawbacks is the discrepancy between diagnosis of individuals in real practice and that recorded in Claim DB. Generally, proportion of discrepancy in diagnosis might be more prominent in claim data from outpatient clinics, less-severe illnesses, and primary care clinics, compared with inpatient hospitalization, severe illnesses, and tertiary or general hospitals, respectively. Therefore, appropriate operational definitions should be required to mini-

mize the inconsistent and inaccurate results. Moreover, because the NHIS covers only insured benefits, uninsured payments could not be estimated from this database. Because information of MA beneficiaries was incorporated into a single NHIS database from 2006, NHIS database during 2002 to 2005 included only information of NHI beneficiaries but not MA beneficiaries, which should be considered with caution when interpreting the findings in this period. In terms of Health Check-up DB, the proportion of complete health check-up examinees was only 40% in 2002, and different intervals of health check-up among beneficiaries should be considered in the study design.

CONFLICTS OF INTEREST

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

REFERENCES

1. 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.
2. Kim NH, Lee J, Kim TJ, Kim NH, Choi KM, Baik SH, Choi DS, Pop-Busui R, Park Y, Kim SG. Body mass index and mortality in the general population and in subjects with chronic disease in Korea: a nationwide cohort study (2002-2010). PLoS One 2015;10:e0139924.
3. National Health Insurance Service: Health Checkup. Available from: http://www.nhis.or.kr/static/html/wbd/g/a/wbdga0606.html (cited 2016 Feb 11).