Cohort Profile

Cohort Profile: The National Health Insurance Service–National Sample Cohort (NHIS-NSC), South Korea

Juneyoung Lee,[1] Ji Sung Lee,[2] Sook-Hee Park,[3] Soon Ae Shin[3] and KeeWhan Kim[4]*

[1]Department of Biostatistics, College of Medicine, Korea University, Seoul, Korea, [2]Clinical Research Center, Asian Medical Center, Seoul, Korea, [3]Big Data Steering Department, National Health Insurance Service, Seoul, Korea and [4]Department of Applied Statistics, Korea University Sejong Campus, Sejong City, Korea

*Corresponding author. Department of Applied Statistics, Korea University Sejong Campus 2511 Sejong-ro, Sejong City, 30019, Korea. E-mail: korpen@korea.ac.kr

Why was the cohort set up?

The National Health Insurance Service–National Sample Cohort (NHIS-NSC) is a population-based cohort established by the National Health Insurance Service (NHIS) in South Korea. The sole purpose of constructing the cohort was to provide public health researchers and policy makers with representative, useful information regarding citizens' utilization of health insurance and health examinations. Korea's universal coverage health insurance system for all citizens was initiated in 1963, based on the National Medical Insurance Act, and was introduced for companies with over 500 employees in 1977. Universal healthcare coverage was achieved in 1989, only 12 years after its introduction, which is the fastest this has been achieved globally.[1] In 2000 a single-insurer system, the NHIS, was launched by integrating more than 366 medical insurance organizations, for efficient system operation in Korea.[1] The NHIS provides benefits for prevention, diagnosis, disease and injury treatment, as well as rehabilitation, births, deaths and health promotion. Currently the NHIS maintains and stores national records for healthcare utilization and prescriptions.

The NHIS records have garnered academic interest due to the effectiveness of the system and relevance to public health and medical research. To meet this interest, a population database has been developed, the 'National Health Information Database' (NHID)[2] containing personal information, demographics and medical treatment data for Korean citizens, who were categorized as insured employees, insured self-employed individuals or medical aid beneficiaries. The NHID was generated using participants' medical bill expenses claimed by medical service providers. Data were rearranged according to date of medical treatment rather than date of claim. However, due to limited useability of the NHID's unavoidably large volume and the lack of confidentiality regarding personal information, the NHIS decided to construct a representative sample database, the NHIS-NSC, with a substantial volume of representative information that does not require privacy regulation for research and policy development.

Who is in the cohort?

To construct the NHIS-NSC, we first built a target population of 46 605 433 individuals using 47 851 928 individuals in the 2002 NHID by excluding non-citizens and special-purpose employees with an unidentifiable income level. From the target population a representative sample cohort

of 1 025 340 participants was randomly selected, comprising 2.2% of the total eligible Korean population in 2002, and followed for 11 years until 2013 unless participants' eligibility was disqualified due to death or emigration. Systematic stratified random sampling with proportional allocation within each stratum was conducted using the individual's total annual medical expenses as a target variable for sampling.[3] First, 1476 strata were constructed by age group, sex, participant's eligibility status and income level. Specifically, strata were defined by 18 age groups (infants under 1 year, ages 1–4, 5-year age groups between 5 and 79, and 80 years and above), two groups according to sex (male, female) and 41 groups based on participant's income level (upper 20 percentiles for insured employees, lower 20 percentiles for insured self-employed individuals, and the lowest level of income for medical aid beneficiaries). Next, within each stratum, systematic sampling was conducted after sorting population data by the value of total annual medical expenses and maintaining a sampling rate of 2.2%. Stratum samples were iteratively drawn until a maximum absolute percentage error—defined as a relative percentage difference between population and sample averages of total annual medical expenses to the population average—reached a predefined value of less than 5%. This technique was used to compensate for the severely positively skewed total annual medical expenses of the entire cohort and each

stratum. During the follow-up period, the cohort was refreshed annually by adding a representative sample of newborns, sampled across 82 strata (two for sex, combined with 41 for parents' income levels) using the 2.2% sampling rate (Figure 1). Participant's residential information was not used as a stratum variable because the NHIS maintained records of workplace addresses until 2005 and residential addresses after 2006.

Although the representativeness for follow-up years is not guaranteed, using an appropriate sampling design and sufficient sample size for the initial cohort can help ensure representativeness. The sample's representativeness was, therefore, evaluated by examining whether a 95% confidence interval for the sample's average total annual medical expenses contained the population average; it was satisfied in every stratum. Further, the sample cohort was compared with the population according to residence distribution across 16 regions in Korea. Moreover, the mean and standard deviation of health insurance premiums for the sample and population for each cohort year were compared; these were not used as a stratification or target variable for sampling. The difference in the proportion of residence is negligible for 2002, and changed slightly during the follow-up years 2003–13 by 0–0.3%. The difference in average health insurance premium is also negligible during cohort years.

Figure 1. A schematic representation of the cohort data construction. DB, database.

How often have they been followed-up?

The cohort sampled in the 2002 NHID database was followed until 2013, provided that participants were still eligible for health insurance. The total numbers of participants in each of cohort years are presented in Table 1. The number of infants (age 0) in the initial cohort and those added annually are also given in the table. Currently the NHIS plans to maintain regular annual cohort updates for the NHIS-NSC.

What has been measured?

The cohort comprises four databases on participants' insurance eligibility, medical treatments, medical care institutions and general health examinations. The insurance eligibility database contains 14 variables including information on participant's identity and socioeconomic variables such as gender, residential area, type of health insurance, level of income, type and grade of disability registered, birth and death. Variables for cause of death and residential area details are provided upon request (see the 'Can I get hold of the data?'). The medical treatment database consists of 57 variables containing information about participants' medical bills claimed by medical service providers. It comprises four databases: participant's electronic medical treatment bills, bill details, details of diseases and details of prescriptions. All four databases are further classified according to type of medicine: 'medical' and 'dental & Chinese medicine' tables. A pharmacy table is also included in the first two databases. In the medical care institution database, information regarding the type of institution, establishment, location, number of beds, facilities and physicians are recorded

Table 1. Number of participants in each cohort year and number of infants added annually (unit: person)

Year	Number of participants in cohort (A)	Number of infants aged 0 in the cohort	Number of participants who took the health examination (B)	Percentage of subjects who took the health examination (B/A)
2002	1025340	9565	113641	11%
2003	1017468	9437	118758	12%
2004	1016580	9320	142281	14%
2005	1016820	8557	135475	13%
2006	1002005	7872	174625	17%
2007	1020743	9766	162829	16%
2008	1000785	9393	210960	21%
2009	998527	8616	211541	21%
2010	1002031	9032	228746	23%
2011	1006481	9694	235336	23%
2012	1011123	9851	241397	24%
2013	1014730	8825	234478	23%

under 10 variables. The general health examination database comprises information regarding nationwide health examinations conducted by the NHIS in 2002–13, including major health examination results and information about lifestyles and behaviours obtained from questionnaires. In Korea, nationwide health examinations are conducted for citizens aged 40 years and above.[4] Two types of examinations are performed, a general and a life-transition health examination. The former, initiated in 1995, is administered biennially to citizens aged 40 years or older who are dependants of the insured employee or householder/family members of insured self-employed individuals. An insured employee and a householder of the insured self-employed can receive the general health examination regardless of his/her age. For blue-collar employees, this examination is conducted annually. According to the 2013 NHIS statistics, 72.1% of eligible beneficiaries had received general health examinations.[5] The more comprehensive life-transition health examination, initiated in 2008, is given to individuals on reaching age 40 and age 66, twice in a lifetime, who are eligible for general health examinations. Both nationwide health examinations involve a screening and a confirmatory test. Examination details are summarized in Table 2. The NHIS-NSC database contains only the first-stage (screening) examination data for those who took the examination during cohort years, with two separate datasets for 2002–08 and 2009–13 because major changes were made to the content of health examinations and questionnaires in 2009 in accordance with a system reformation. Thus, the general health examination database contains 37 variables in the 2002–08 datasets and 41 in the 2009–13 datasets. The numbers of participants who received health examinations during the cohort years are presented in the fourth column in Table 1; 11% received an examination in the initial cohort year (2002), whereas this is more than doubled in 2013, reaching 23%. A detailed list of NHIS-NSC database variables is included in Appendix Table 1 (available as Supplementary data at *IJE* online).

To protect participants' privacy, the Resident Registration Number (RRN, a unique identification number in Korea) which was initially used to construct the cohort, has been replaced with a newly-assigned eight-digit personal ID. Furthermore, to prevent the possibility of identifying a participant by merging information about rare disease status, age and residence, we replaced ICD-10 codes of 114 sensitive diseases with an asterisk except for the code's initial.

What has it found? Key findings and publications

A comparison of socio-demographic variables in the NHIS-NSC database and population in 2002, as well as in

Table 2. Types and content of general health examinations provided by the NHIS

Type	Stage and purpose	Content in brief
General health examination[a]	First-stage examination as a screening test	A medical interview and postural examination, chest X-ray examination, blood test, urine test, dental screening, etc
	Second-stage examination as a confirmatory test[c]	Consultation on screening results and healthcare education for eligible individuals. Detailed examination on fasting glucose level and blood pressure
Life-transition health examination[b]	First-stage examination as a screening test	For 40-year-old individuals: an examination and consultation, hepatitis B antigen and antibody tests, etc
		For 66-year-old individuals: an examination and consultation, bone density test, elderly physical function test, etc
	Second-stage examination as a confirmatory test[d]	For 40-year-old individuals: confirmatory testing for hypertension and diabetes, evaluation of health examination results and health risk consultation, mental examination (depression), lifestyle examination (evaluation and prescriptions)
		For 66-year-old individuals: identical items to those in individuals aged 40, as well as an additional examination for mental health (cognitive dysfunctions)

[a]Eligibility: An insured employee and a householder of the insured self-employed regardless of his/her age, a dependant of the insured self-employed individual over 40 years old, or a dependant of the insured employee, over 40 years old.

[b]Eligibility: individuals aged 40 and 66. This examination was started in 2008.

[c]The second-stage examination is performed if an examinee is categorized with suspected hypertension or diabetes or if a 70- or 74-year-old examinee is classified into a high-risk cognitive impairment category from his/her first-stage examination.

[d]The second-stage examination is performed on all examinees who received the first-stage examination regardless of its result.

2013 (the most recent year of available data), are presented in Tables 3 and 4, respectively; and a comparison of health examination variables in 2002 and 2013 are presented in Appendix Table 1 (available as Supplementary data at *IJE* online) and Appendix Table 2 (available as Supplementary data at *IJE* online), respectively. A 95% confidence interval of each variable is also presented. For all demographic variables in 2002, the intervals contained the population average, indicating that the difference between the cohort and population was not significant (Table 3).

In the year 2013, after 11 years of follow-up, the cohort proportion of insured employees underestimated that of the general population, whereas the cohort proportions of self-employed insured and medical-aid beneficiaries overestimated the population proportions for both males and females; however, the differences of less than 0.3% were trivial (Table 4).

For smoking status—as a health examination variable— the cohort overestimated the proportion of male and female non-smokers compared with the general population at the time of data collection in the initiation year. The cohort included a significantly higher proportion of men who did not exercise and a lower proportion of men engaging in mild–moderate exercise, compared with the population. No statistical differences for other health variables between the cohort and population in 2002 were found (Appendix Table 2, available as Supplementary data at *IJE* online). In 2013, for males only the sample proportion of ex-smokers was 0.7% higher than that of the general population.

There were differences between the cohort and the population in frequency of exercise (intensive physical activity more than 20 minute per week and moderate exercise more than 30 minute per week, variables that were surveyed since 2009; see Appendix Table 3, available as Supplementary data at *IJE* online). This finding implies that the cohort's representativeness regarding some general health examination variables for health behaviour could be inadequate, requiring a periodic adjustment for future cohort years. We also would like to mention that the NHIS is currently preparing to build a special-purpose cohort, specific to general health examination data, using a population database of the NHID.

Providing public access to the NHIS-NSC database can support research in auxiliary fields such as sociology, economics, environment policy and industry, besides evidence-based academic research in public health and medicine. As of March 2015, 8 months after becoming publicly available in July 2014, 109 studies (99 academic and 10 political researches) are being conducted using the NHIS-NSC database. Among these, Rim *et al.* found that the risk of stroke after retinal vein occlusion (RVO) was significantly higher especially for ischaemic stroke patients.[6] They also showed that those with RVO had an approximately 2-fold higher hazard ratio among younger, compared with older, adults: suggesting that ophthalmologists need to specifically attend to this population.[6] Kwon *et al.* examined the association between bisphosphonate exposure and osteonecrosis of the jaw (ONJ) in Korean patients with osteoporosis.[7] They performed a nested case-control study using the NHIS-NSC database and found a positive relationship between the two, arguing that this relationship must be acknowledged for older adults requiring

Table 3. Comparison of socio-demographic variables between the general population and sample cohort in 2002 [number of subjects (percentage)]

Variable	Male					Female				
	Population		Sample		95% CI[a]	Population		Sample		95% CI[a]
Total	23329805	(100)	513258	(100)		23275628	(100)	512082	(100)	
Age (years)										
0	227248	(1.0)	5005	(1.0)	(0.9, 1.0)	206371	(0.9)	4560	(0.9)	(0.9, 0.9)
1–4	1240099	(5.3)	27283	(5.3)	(5.3, 5.4)	1131232	(4.9)	24886	(4.9)	(4.8, 4.9)
5–9	1801438	(7.7)	39633	(7.7)	(7.6, 7.8)	1600649	(6.9)	35215	(6.9)	(6.8, 6.9)
10-14	1717529	(7.4)	37784	(7.4)	(7.3, 7.4)	1524800	(6.6)	33547	(6.6)	(6.5, 6.6)
15–19	1666548	(7.1)	36663	(7.1)	(7.1, 7.2)	1542917	(6.6)	33948	(6.6)	(6.6, 6.7)
20–24	1950098	(8.4)	42902	(8.4)	(8.3, 8.4)	1902008	(8.2)	41841	(8.2)	(8.1, 8.2)
25–29	1975183	(8.5)	43453	(8.5)	(8.4, 8.5)	1938614	(8.3)	42651	(8.3)	(8.3, 8.4)
30–34	2277379	(9.8)	50103	(9.8)	(9.7, 9.8)	2188378	(9.4)	48143	(9.4)	(9.3, 9.5)
35–39	2083161	(8.9)	45826	(8.9)	(8.9, 9.0)	1983794	(8.5)	43645	(8.5)	(8.4, 8.6)
40–44	2196838	(9.4)	48330	(9.4)	(9.3, 9.5)	2111780	(9.1)	46458	(9.1)	(9.0, 9.2)
45–49	1696211	(7.3)	37316	(7.3)	(7.2, 7.3)	1646589	(7.1)	36223	(7.1)	(7.0, 71)
50–54	1230513	(5.3)	27069	(5.3)	(5.2, 5.3)	1213561	(5.2)	26699	(5.2)	(5.2, 5.3)
55–59	962847	(4.1)	21185	(4.1)	(4.1, 4.2)	1007404	(4.3)	22163	(4.3)	(4.3, 4.4)
60–64	920569	(3.9)	20251	(3.9)	(3.9, 4.0)	1041312	(4.5)	22912	(4.5)	(4.4, 4.5)
65–69	641017	(2.7)	14102	(2.7)	(2.7, 2.8)	830015	(3.6)	18257	(3.6)	(3.5, 3.6)
70–74	370012	(1.6)	8142	(1.6)	(1.6, 1.6)	608032	(2.6)	13374	(2.6)	(26, 2.7)
75–79	213386	(0.9)	4696	(0.9)	(0.9, 0.9)	402329	(1.7)	8852	(1.7)	(1.7, 1.8)
80+	159729	(0.7)	3515	(0.7)	(0.7, 0.7)	395843	(1.7)	8708	(1.7)	(1.7, 1.7)
Insurance type										
Employee insured										
	11275055	(48.3)	248048	(48.3)	(48.2, 48.5)	11298107	(48.5)	248549	(48.5)	(48.4, 48.7)
Self-employed insured										
	11461798	(49.1)	252161	(49.1)	(49.0, 49.3)	11174712	(48.0)	245869	(47.9)	(47.9, 48.2)
Medical-aid beneficiary										
	592952	(2.5)	13049	(2.5)	(2.5, 2.6)	802809	(3.4)	17664	(3.4)	(3.4, 3.5)

CI, confidence interval.

[a]95% confidence interval for the sample proportion.

dental integration, to ensure that the benefits and risks are evaluated and that symptoms suggestive of ONJ are monitored.[7]

What are the main strengths and weaknesses?

The NHIS-NSC database contains representative population-based cohort data, which is a major strength as it ensures its applicability in research—for example, when evaluating the effects of medical practice on health outcomes. Moreover, the data are large-scale, extensive and stable because it is constructed based on nationwide health insurance data generated by the government or public institutions' involvement. Therefore, the cohort can also be used by policy makers to create higher value-added policies.

Similar databases such as the Healthcare Cost and Utilization Project-National Inpatient Sample (NIS)[8] in the USA or the National Health Insurance Research Database (NHIRD)[9] in Taiwan, are available. Because the primary sampling unit of the NIS database, however, is the hospital, overlapping participants may introduce a selection bias. The NHIRD database uses a simple random sampling strategy; hence, the representativeness of major health-related indicators including the population's demographic characteristics may have been lost. Moreover, they may not free of inherent limitations of cross-sectional data in evaluating, for example, an effect of medical practice on a health outcome. However, since the NHIS-NSC is a cohort based on nationwide health insurance data, it is both representative of the population and overcomes the limitations of cross-sectional data.

The NHIS-NSC database has several limitations. Although the cohort comprises over one million participants, information on rare diseases may not be sufficient. Therefore, it is necessary to conduct a pre-evaluation of study size when using the NHIS-NSC database. The NHIS

Table 4. Comparison of socio-demographic variables between the population and sample cohort in 2013, number of subjects (percentage)

Variable	Male Population		Sample		95% CI[a]	Female Population		Sample		95% CI[a]
Total	25342562	(100)	507289	(100)		25385482	(100)	507441	(100)	
Age (years)										
0	214199	(0.9)	4529	(0.9)	(0.9, 0.9)	203389	(0.8)	4296	(0.8)	(0.8, 0.9)
1–4	967583	(3.8)	18956	(3.7)	(3.7, 3.8)	913452	(3.6)	17887	(3.5)	(3.5, 3.6)
5–9	1199738	(4.7)	22152	(4.4)	(4.3, 4.4)[b]	1122761	(4.4)	20664	(4.1)	(4.0, 4.1)[b]
10–14	1461815	(5.8)	28704	(5.7)	(5.6, 5.7)[b]	1340274	(5.3)	26095	(5.1)	(5.1, 5.2)[b]
15–19	1794779	(7.1)	36610	(7.2)	(7.1, 7.3)	1612700	(6.4)	32875	(6.5)	(6.4, 6.5)
20–24	1795130	(7.1)	36484	(7.2)	(7.1, 7.3)	1595580	(6.3)	32626	(6.4)	(6.4, 6.5)[b]
25–29	1640486	(6.5)	32339	(6.4)	(6.3, 6.4)[b]	1526474	(6.0)	30473	(6.0)	(5.9, 6.1)
30–34	2069275	(8.2)	40520	(8.0)	(7.9, 8.1)[b]	1988233	(7.8)	39055	(7.7)	(7.6, 7.8)
35–39	1984652	(7.8)	39006	(7.7)	(7.6, 7.8)	1909943	(7.5)	37372	(7.4)	(7.3, 7.4)[b]
40–44	2312537	(9.1)	46223	(9.1)	(9.0, 9.2)	2234987	(8.8)	44602	(8.8)	(8.7, 8.9)
45–49	2161800	(8.5)	43546	(8.6)	(8.5, 8.7)	2103145	(8.3)	42297	(8.3)	(8.3, 8.4)
50–54	2192369	(8.7)	44450	(8.8)	(8.7, 8.8)	2152880	(8.5)	43460	(8.6)	(8.5, 8.6)
55–59	1785614	(7.1)	35949	(7.1)	(7.0, 7.2)	1790479	(7.1)	35624	(7.0)	(7.0, 7.1)
60–64	1202352	(4.7)	24256	(4.8)	(4.7, 4.8)	1257816	(5.0)	25288	(5.0)	(4.9, 5.0)
65–69	931096	(3.7)	19004	(3.7)	(3.7, 3.8)	1035630	(4.1)	20788	(4.1)	(4.0, 4.2)
70–74	787789	(3.1)	16374	(3.2)	(3.2, 3.3)[b]	1005549	(4.0)	20509	(4.0)	(4.0, 4.1)
75–79	498430	(2.0)	10565	(2.1)	(2.0, 2.1)	766614	(3.0)	15628	(3.1)	(3.0, 3.1)
80+	342918	(1.4)	7622	(1.5)	(1.5, 1.5)[b]	825576	(3.3)	17902	(3.5)	(3.5, 3.6)[b]
Insurance type										
Employee insured										
	17191838	(67.8)	342918	(67.6)	(67.5, 67.7)[b]	17207170	(67.8)	342699	(67.5)	(67.4, 67.7)[b]
Self-employed insured										
	7510410	(29.6)	151202	(29.8)	(29.7, 29.9)[b]	7358849	(29.0)	147807	(29.1)	(29.0, 29.3)
Medical-aid beneficiary										
	640314	(2.5)	13169	(2.6)	(2.6, 2.6)[b]	819463	(3.2)	16935	(3.3)	(3.3, 3.4)[b]

[a]95% confidence interval for the sample proportion.

[b]The population value has not been included in the 95% confidence i interval of the sample proportion.

is currently preparing special-purpose cohort databases such as a cohort of older adults and of female workers, as well as customized databases for policy development/ evaluation and academic research. Disease codes listed in the cohort may not represent participant's true disease status because the code was created to claim health insurance serviced to participants, an inherent limitation of insurance databases. Hence, it warrants careful use by researchers. In this cohort, non-insurance benefits data such as cosmetic surgeries and information for over-the-counter drugs have not been included. Moreover, evaluating details of a participant's specific medical treatment is difficult if his/her insurance claims were made under the diagnosis-related-group (DRG) policy. In contrast to the traditional fee-for-service payment system, the DRG system reimburses a fixed amount of medical fees to all hospitalized patients, depending on the patient's illness and regardless of the type or cost of medical services provided during hospitalization.[10,11] In Korea, nearly all types of healthcare providers follow the fee-for-service payment system and the DRG is applied only to seven disease groups (for details, see Health Insurance Review & Assessment Service of Korea website) [http:// kostat.go.kr/portal/english/index.action].[12]

Can I get hold of the data? Where can I find out more?

Currently, the NHIS-NSC database consists of 156 SAS® data files, comprising 13 files—for participants' insurance eligibility (1 file), medical treatments (10 files), medical care institutions (1 file) and health examination (1 file)— for each of the 12 years of the cohort between 2002 and 2013. The total cohort file size is approximately 211 gigabytes with 2619 million cases in 2002–13.

Data can be accessed through the NHIS' National Health Insurance Data Sharing Service website [http:// nhiss.nhis.or.kr/bd/ab/bdaba021eng.do]. To gain access to NHIS-NSC data, a completed application form, a research

Figure 2. The process for accessing the NHIS-NSC database. IRB, Institutional Review Board.

proposal and the applicant's institutional review board (IRB) approval document should be submitted to and reviewed by the Review Committee of Research Support in NHIS. After granting approval, data are provided to an applicant for a fee. The data application process is described in Figure 2. Upon request, causes of death prepared by Statistics Korea[12] and information regarding participant's district of residence can be provided by the NHIS after the committee's review.

Supplementary Data

A list of variables and other NHIS-NSC data are included in the Appendix, available as Supplementary data at *IJE* online.

The NHIS-NSC profile in a nutshell

- The NHIS-NSC database is a population-based sample cohort. Its purpose is to provide representative, useful health insurance and health examination data to public health researchers and policy makers.
- A total of 1 025 340 participants of the cohort, 2.2% of the total eligible population, were randomly sampled from the 2002 Korean (nationwide) health insurance database to obtain baseline data.
- Cohort participants were followed for 11 years, until

2013. During the follow-up period, a representative sample of newborns (age 0) was added annually and deceased or emigrated participants were excluded. In 2013, the database included 1 014 730 participants.
- Information about participants' insurance eligibility, medical treatment history, healthcare provider's institution and general health examination are included.
- The NHIS-NSC database access on [http://nhiss.nhis.or.kr/bd/ab/bdaba021eng.do] requires a completed application form, a research proposal and the institutional review board's approval document.

References

1. National Health Insurance Service (NHIS), Korea. *History of the NHIS.* http://www.nhis.or.kr/static/html/wbd/g/a/wbdga0203.html Accessed (12 August 2015, date last accessed).

2. National Health Insurance Corporation. *Final Report of Academic Research Project. The Effect of National Health Examination Program on the Early Diagnosis of Diseases, Medical Utilization, and Health Outcome*. Seoul: National Health Insurance Corporation, 2011.

3. Cochran WG. *Sampling Techniques*. 3rd edn. New York, NY: Wiley, 1977.

4. National Health Insurance Service (NHIS), Korea. *Health Insurance Guide*. http://www.nhis.or.kr/static/html/wbd/g/a/wbdga0606.html (12 August 2015, date last accessed).

5. National Health Insurance Service (NHIS). *National Health Examination Statistical Yearbook*. Seoul: National Health Insurance Service, 2014.

6. Rim TH, Kim DW, Han JS, Chung EJ. Retinal vein occlusion and the risk of stroke development. A 9-year nationwide population-based study. *Ophthalmology* 2015;**122:**1187–94

7. Kwon J-W, Park E-J, Jung S-Y, Shon HS, Ryu H, Suh HS. A large national cohort study of the association between bisphosphonates and osteonecrosis of the jaw in patients with osteoporosis: A nested case-control study. *J Dent Res* 2015;**95**(Suppl 9)**:**2125–29S.

8. Healthcare Cost and Utilization Project (HCUP). *HCUP Nationwide Inpatient Sample (NIS) 2007–09*. Rockville, MD: Agency for Healthcare Research and Quality, 2007–09.

9. National Health Research Institute, Taiwan. *National Health Insurance Research Database (NHIRD)*. http://nhird.nhri.org.tw/en/ (12 August 2015, date last accessed).

10. Health Insurance Review & Assessment Service (HIRA). *Diagnosis-Related Group Payment*. https://www.hira.or.kr/eng/about/05/02/07/index.html (12 August 2015, date last accessed).

11. Fetter RB, Freeman JL. Diagnosis related groups: product line management within hospitals. *Acad Manage Rev* 1986;**11:**41–54.

12. Statistics Korea. http://kostat.go.kr/portal/english/index.action (12 August 2015, date last accessed).