

# BMJ Open Cohort Profile: National health insurance service-senior (NHIS-senior) cohort in Korea

Yong Ik Kim,<sup>1</sup> Yeon-Yong Kim,<sup>2</sup> Jong Lull Yoon,<sup>3</sup> Chang Won Won,<sup>4</sup> Seongjun Ha,<sup>2</sup> Kyu-Dong Cho,<sup>1</sup> Bo Ram Park,<sup>1</sup> Sejin Bae,<sup>1</sup> Eun-Joo Lee,<sup>2</sup> Seong Yong Park,<sup>1</sup> Jong Heon Park,<sup>2</sup> Kyeong-ran Lee,<sup>1</sup> Donghun Lee,<sup>1</sup> Seung-lyeal Jeong,<sup>2</sup> Hyung-soo Kang<sup>1</sup>

**To cite:** Kim YI, Kim Y-Y, Yoon JL, *et al.* Cohort Profile: National health insurance service-senior (NHIS-senior) cohort in Korea. *BMJ Open* 2019;**9**:e024344. doi:10.1136/bmjopen-2018-024344

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-024344>).

Received 22 May 2018

Revised 23 October 2018

Accepted 17 January 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>National Health Insurance Service, Wonju, Gangwon-do, Republic of Korea

<sup>2</sup>Big Data Steering Department, National Health Insurance Service, Wonju, Republic of Korea

<sup>3</sup>Hallym University College of Medicine, Chuncheon, Gangwon, Republic of Korea

<sup>4</sup>Kyung Hee University School of Medicine, Seoul, Republic of Korea

## Correspondence to

Mr Hyung-soo Kang;  
khs0293@nhis.or.kr

## ABSTRACT

**Purpose** The National Health Insurance Service (NHIS)-Senior was set up to provide high-quality longitudinal data that can be used to explore various aspects of changes in the socio-economical and health status of older adults, to predict risk factors and to investigate their health outcomes.

**Participants** The NHIS-Senior cohort, a Korean nationwide retrospective administrative data cohort, is composed of older adults aged 60 years and over in 2002. It consists of 558 147 people selected by 10% simple random sampling method from a total of 5.5 million subjects aged 60+ in the National Health Information Database. The cohort was followed up through 2015 for all subjects, except for those who were deceased.

**Findings to date** The healthcare utilisation and admission rates were the highest for acute upper respiratory infections and influenza (75.2%). The age-standardised (defined with reference to the world standard population) mortality rate for 10 years (through 2012) was 4333 per 100 000 person-years. Malignant neoplasms were the most common cause of death in both sexes (1032.1 per 100 000 person-years for men, 376.7 per 100 000 person-years for women). A total of 34 483 individuals applied for long-term care service in 2008, of whom 17.9% were assessed as grade 1, meaning that they were completely dependent on the help of another person to live daily life.

**Future plans** The data are provided for the purposes of policy and academic research under the Act on Promotion of the Provision and Use of Public Data in Korea. The NHIS-Senior cohort data are only available for Korean researchers at the moment, but it is possible for researchers outside the country to gain access to the data by conducting a joint study with a Korean researcher. The cohort will be maintained and continuously updated by the NHIS.

## INTRODUCTION

The Republic of Korea (hereafter ‘Korea’) is experiencing the fastest population ageing among the Organisation for Economic Co-operation and Development countries.<sup>1</sup> It is projected that the proportion of the elderly – 65 years old and over – population,

## Strengths and limitations of this study

- It provides reliable data based on a large sample (n=558 147) and a long duration of follow-up (from 2002 to 2015), and is representative of the entire elderly population.
- It is possible to analyse the pre-hospital stages of disease using information from the health screening programme and long-term care services.
- The NHIS-Senior cohort has certain limitations. First, some of the subjects did not participate in the health screening programme or receive long-term care services. Second, variables on health behaviours are limited since those data were obtained from self-reporting. Third, the disease codes might not accurately reflect patients’ medical conditions.

which was 12.8% in 2015, will reach 42.5%<sup>2</sup> in 2065 due to a dramatic increase in life expectancy and a sharp decrease in the birth rate.<sup>3</sup> Increasing population ageing is also a source of economical, health and social burdens.<sup>4</sup> As the family structure has also rapidly changed, the proportion of elderly people living with adult children decreased from 53.2% in 1998<sup>5</sup> to 28.6% in 2008.<sup>6</sup> Accordingly, the age dependency ratio, defined as the number of individuals aged 65 and over per 100 people of working age, is projected to reach 88.6 in 2065.<sup>2</sup>

In 2008, long-term care insurance (LTCI) was introduced in Korea as a form of social insurance to share caregivers’ burden due to functional loss and chronic disease complications. It was organised and planned by the Ministry of Health and Welfare, and has been implemented by the National Health Insurance Service (NHIS) based on the Act on Long-Term Care Insurance for the Senior Citizens.<sup>7</sup> Since Korea has a single public insurance system that covers medical utilisation for the entire population, it was easier

**Table 1** General characteristics of the national health insurance service-senior (NHIS-Senior) cohort

	2002		2005		2012		2015	
	N	%	N	%	N	%	N	%
Total	558 147	100.0	521 967	100.0	405 614	100.0	352 869	100.0
Sex								
Men	230 582	41.3	213 048	40.8	157 316	38.8	133 741	37.9
Women	327 565	58.7	308 919	59.2	248 298	61.2	219 128	62.1
Age, years								
60–64	196 116	35.1	85 776	16.4				
65–69	147 361	26.4	167 293	32.1				
70–74	97 657	17.5	120 415	23.1	172 606	42.6	74 382	21.1
75–79	61 217	11.0	76 092	14.6	118 931	29.3	135 748	38.5
80+	55 796	10.0	72 391	13.9	114 077	28.1	142 739	40.5
Insurance type								
Self-employed insured	234 763	42.1	182 351	34.9	113 113	27.9	92 615	26.3
Employee insured	277 958	49.8	289 184	55.4	255 864	63.1	227 504	64.5
Medical aid	45 426	8.1	50 432	9.7	36 637	9.0	32 750	9.3
Charlson comorbidity index								
0	300 329	53.8	211 637	40.6	95 206	23.5	70 891	20.1
1	120 671	21.6	126 146	24.2	94 252	23.2	79 002	22.4
2	65 648	11.8	78 037	15.0	75 680	18.7	67 630	19.2
3+	71 499	12.8	106 147	20.3	140 476	34.6	135 346	38.4
Disability								
Yes	554 477	99.3	517 318	99.1	401 148	98.9	348 376	98.7
No	3670	0.7	4649	0.9	4466	1.1	4493	1.3

to implement the insurance system using the pre-existing health insurance organisations—the NHIS with 178 branch offices nationwide—than using a tax-based long-term care programme that would need to be newly created.<sup>8</sup> LTCI provides facility or home care services for the elderly with impairments in activities of daily living. When a person applies for LTCI services, he or she is given a grade based on a comprehensive consideration of cognitive function, activities of daily living and mental and rehabilitation status. The type of service received is determined by the grade.<sup>7</sup>

In 2011, the NHIS established an administrative database for research purposes, the National Health Information Database (NHID), which stores all the records of healthcare and long-term care services.<sup>9</sup> As the NHIS also provides a national health screening programme that includes medical check-ups every 2 years, data from a self-reported questionnaire (lifestyle, past medical history and family medical history) and measured biometric information (blood pressure, anthropometry, clinical laboratory and urinalysis findings) are included in the NHID.<sup>10</sup>

The NHIS-Senior cohort, a nationwide retrospective cohort, which includes information from elderly individuals randomly sampled by the NHID, was constructed by the Big Data Steering Department of the NHIS head office

in 2016. The NHIS-Senior cohort consists of five databases: an eligibility database, a national health screening database, a healthcare utilisation database, a long-term care insurance database and a healthcare provider database. This cohort was set up to provide high-quality longitudinal data that can be used to explore various aspects of changes in the socio-economical and health status of older adults, to predict risk factors and to investigate their health outcomes.

## COHORT DESCRIPTION

### The participants of the cohort

The NHIS-Senior cohort is composed of older adults aged 60 years and over in 2002. It consists of 558 147 people selected by 10% simple random sampling method from a total of 5.5 million subjects aged 60+ in the NHID (online supplementary table 1). The cohort was followed up retrospectively through 2015 for all subjects, except for those who lost eligibility for national health insurance, a compulsory social insurance programme, due to death and emigration, in accordance with the National Health Insurance Act. Emigrants were excluded from the NHIS-Senior cohort because the rate of emigration from Korea is very low<sup>11</sup> and it is difficult to follow-up emigrants due to the frequent changes in their eligibility.

**Table 2** Healthcare utilisation of the population in the national health insurance service-senior (NHIS-Senior) cohort

	2002			2005			2012			2015		
	N	Subgroup %	Total %	N	Subgroup %	Total %	N	Subgroup %	Total %	N	Subgroup %	Total %
	Cohort population	558 147	100	100	521 967	100	100	405 614	100	100	352 869	100
<b>Inpatients</b>												
Number of inpatients	51 515	100	9.2	88 127	100	16.9	115 557	100	28.5	108 203	100	30.7
Frequency of inpatient visits												
1	37 002	71.8	55 545	63.0	48 302	44.6	55 408	48.0	48 302	44.6	48 302	44.6
2	9 625	18.7	18 632	21.1	24 911	20.5	24 911	21.6	22 202	20.5	22 202	20.5
3	2 484	4.8	6 154	7.0	9 866	8.7	9 866	8.5	9 376	8.7	9 376	8.7
4+	2 404	4.7	7 796	8.9	25 372	26.2	25 372	22.0	28 323	26.2	28 323	26.2
<b>Hospital days</b>												
0-14	35 805	69.5	55 595	63.1	62 764	51.5	62 764	54.3	55 758	51.5	55 758	51.5
15-29	9 077	17.6	16 217	18.4	20 048	16.3	20 048	17.4	17 614	16.3	17 614	16.3
30-59	4 407	8.6	9 692	11.0	12 438	10.6	12 438	10.8	11 509	10.6	11 509	10.6
61-119	1 574	3.1	3 892	4.4	6 981	6.3	6 981	6.0	6 811	6.3	6 811	6.3
120+	652	1.3	2 731	3.1	13 326	15.3	13 326	11.5	16 511	15.3	16 511	15.3
<b>Outpatients</b>												
Number of outpatients	453 969	100	81.3	478 209	100	91.6	387 096	100	95.4	335 792	100	95.2
Frequency of outpatient visits												
0-4	84 170	18.5	71 238	14.9	24 289	6.1	24 289	6.3	20 587	6.1	20 587	6.1
5-9	78 003	17.2	65 699	13.7	36 232	9.5	36 232	9.4	32 016	9.5	32 016	9.5
10-14	67 237	14.8	63 771	13.3	45 870	11.9	45 870	11.9	39 871	11.9	39 871	11.9
15-29	122 020	26.9	135 034	28.2	118 729	31.4	118 729	30.7	105 434	31.4	105 434	31.4
30-59	75 758	16.7	99 637	20.8	106 866	27.6	106 866	27.6	92 806	27.6	92 806	27.6
60+	26 781	5.9	42 830	9.0	55 110	13.4	55 110	14.2	45 078	13.4	45 078	13.4
<b>Medication</b>												
Number of long-term medications prescribed over 300 days per year	51 471	100	9.2	123 372	100	23.6	210 708	100	52.0	198 971	100	56.4
Number of active ingredients (long-term prescriptions)												
1	15 646	30.4	29 344	23.8	37 683	16.9	37 683	17.9	33 704	16.9	33 704	16.9
2	13 204	25.7	29 664	24.0	41 664	18.1	41 664	19.8	35 935	18.1	35 935	18.1
3	9 149	17.8	23 126	18.7	36 627	16.5	36 627	17.4	32 827	16.5	32 827	16.5
4	5 962	11.6	15 641	12.7	29 561	14.0	29 561	14.0	27 811	14.0	27 811	14.0
5+	7 510	14.6	25 597	20.8	65 173	34.5	65 173	30.9	68 694	34.5	68 694	34.5

**Table 3** Major variables in the national health insurance service-senior (NHIS-Senior) cohort

Domain	Health problems	Variables	Year													
			'02	'03	'04	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15
National health screenings and healthcare utilisation	Hypertension	Systolic blood pressure	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Diastolic blood pressure	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Diabetes mellitus	Fasting blood glucose	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Total cholesterol	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Dyslipidaemia	Triglyceride														
		HDL cholesterol														
	Anaemia	LDL cholesterol														
		Haemoglobin	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Kidney/urinary disease	Urine glucose	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Urine blood	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Chronic kidney disease	Urine pH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Urine protein	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Liver disease	Creatinine														
AST (SGOT)		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Frailty/motility	ALT (SGPT)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	γ-GTP	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Osteoporosis	Neurological examination of lower legs for subjects at age 40 or 66															
	Bone density for subjects at age 40 or 66															
Cognitive impairment, depression	Dental examination	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	Mental health screening	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
National health screenings and healthcare utilisation	Common and uncommon diseases	Disease diagnosis per ICD-10 codes; operation and procedure history, medication history (generic name code, dose, duration of prescription and material codes)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Vital statistics including dates and causes of deaths	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Cigarette smoking	Cigarette smoking status	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Daily smoking dose	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	All cause- and cause- specific deaths	Past daily smoking dose														
		Current daily smoking dose														
	Smoking duration (ex-smoker)	Smoking duration	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Smoking duration (ex-smoker)														
	Smoking duration (current smoker)	Smoking duration (current smoker)														
		Smoking duration (current smoker)														

Continued

Table 3 Continued

Domain	Health problems	Variables	Year													
			'02	'03	'04	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15
Alcohol	Drinking frequency		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Days of drinking per week						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount of drinking per count		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Amount of drinking per day						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obesity	Body mass index		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Waist circumference						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Physical activity	Days of activity per week		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Days of vigorous activity per week						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Days of moderate activity per week						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Days of mild activity per week						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Dental caries, etc	Dental examination		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Hypertension, diabetes mellitus, dyslipidaemia, pulmonary tuberculosis, stroke, ischaemic heart disease, etc.		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Family history	Hypertension, diabetes mellitus, stroke, ischaemic heart disease, etc.		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Date of visit, types of medical institutions (clinics/hospitals/tertiary hospitals/public health centres), types of visit (inpatient/outpatient/emergency/intensive care), length of stay, medical cost (insurer/patient)		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Healthcare utilisation	Location, type of hospitals, number of beds, medical equipment, human resources, specialities of physicians		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Age, sex, age, residential area, insurance type (the employee insured, the self-employed insured, dependents, medical aid), monthly insurance contributions (a proxy for income), types and grades of disabilities		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Long-term care insurance	Application information (medical history, family history, socio-economical status), need assessment records (activities of daily living score, cognitive function, nursing care needs, rehabilitation needs), service grade level, types of LTCI benefits (home care service, institutional care benefits, care allowance for special cases)						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

ALT, alanine aminotransferase; AST, aspartate aminotransferase; HDL, high-density lipoprotein; ICD-10, International Classification of Disease, Tenth Division; LDL, low-density lipoprotein; LTCI, long-term care insurance; SGOT, serum glutamic-oxaloacetic transaminases; SGPT, serum glutamic-pyruvic transaminases;  $\gamma$ -GTP, gamma-glutamyl transpeptidase.

**Table 4** Numbers of all-cause deaths through 2012 (10 years after baseline) and crude and age-standardised (with the 2005 Korean census and world standard populations as references) mortality rates (per 100 000 person-years) in the national health insurance service-senior (NHIS-Senior) cohort

All-cause	No. of cohort population*	No. of deaths	Men			Women			All				
			Age-standardised mortality rates			Age-standardised mortality rates			Age-standardised mortality rates				
			Crude Rate	Census	WHO	Crude rate	Census	WHO	Crude rate	Census	WHO		
Mortality rates (2003–2012)†													
2-year (2004)	558 147	33987	3093	3288	3528	16 638	3675	4426	4688	17 349	2686	2578	2819
5-year (2007)	487 460	83788	3197	3568	3821	40 886	3821	4784	5061	42 902	2767	2832	3085
10-year (2012)	405 614	1 64985	3415	4071	4333	79 357	4087	5379	5661	85 628	2963	3313	3577

\*Number of cohort population at the end of the year.

†Death cases were defined as those cases who died in 2003 to 2012.

In most studies, the elderly are defined as 65 years old and over; however, the seniors in the NHIS-Senior cohort were defined as those aged 60 and over in order to compare health status before and after 65 years old during the follow-up period. The number of subjects in five databases is presented as online supplementary table 2. In this cohort, the first selected subject was traced and no new subjects were added. Therefore, since the age of 60 years or older was selected as of 2002, the age of subjects in 2015 is 73 years or older. In the NHID, de-identified join keys replacing personal identifiers are used to secure ethical clearance.<sup>9</sup> Therefore, the researcher cannot receive informed consent from individual patients for the use of personal information. However, the use of NHID for research purposes requires approval (or exemption) from the institutional review board.

The general characteristics of the cohort are shown in table 1. The total number of the cohort was 558 147 at the beginning (2002) and 352 869 at the end of the follow-up (2015). The proportion of men, which was 41.3% in 2002, declined to 37.9% in 2015 due to higher mortality rates among men. The number of subjects with three points or more on the Charlson comorbidity index,<sup>12</sup> which was calculated from the International Classification of Disease codes in the healthcare utilisation database every year using comorbidity weights from a previous study,<sup>12</sup> increased from 12.8% in 2002 to 38.4% in 2015.

The characteristics of the cohort population regarding healthcare utilisation are presented in table 2, and these findings clearly show how the health status of the elderly became worse over the past decade as they aged. The subjects who had ever been admitted to hospital(s) during a year increased from 51 515 (9.2%) in 2002 to 108 203 (30.7%) in 2015. The proportion of inpatients hospitalised for 120 days or more per year also increased from 1.3% in 2002 to 15.3% in 2015. Of the cohort population, 95.2% utilised outpatient medical services in 2015. The number of patients who received prescriptions for over 300 days per year increased from 9.2% in 2002 to 56.4% in 2015. Among them, the number of patients whose prescriptions included more than five active ingredients increased from 14.6% in 2002 to 34.5% in 2015.

### Follow-up interval

The cohort has been followed-up through 2015 annually for eligibility information including death information and biennially for information from the health screening programme. The data are based on information collected from various sources. Information on death (date and cause of death) was collected from Statistics Korea. By law, all death certificates must be reported to Statistics Korea. Personal information regarding income deciles based on the insurance contribution imposed, residential area and disability status were collected from the Public Information Sharing System, National Tax Service and Ministry of Health and Welfare of Korea. Information on health screening results was only tracked for those who participated in a health screening programme with

scheduled check-ups at least every 2 years. The participation rate in the health screening programme was 77.7% in 2016.<sup>13</sup> LTCI information was only available for those who applied for these services, which started in July 2008. As the NHIS covers the entire population of Korea as a single public insurer, the healthcare utilisation information includes all medical services (from inpatient, outpatient and pharmacy visits) claimed by healthcare facilities in Korea. Information about the healthcare facilities has been also updated annually.

### The key variables

The key variables of the NHIS-Senior cohort, which were mainly selected from the variables of the NHID, are presented in table 3. The eligibility database included information about income-based insurance contributions (a proxy for income), demographical variables and date and cause of death. Health-related risk factors obtained using questionnaires (cigarette smoking status/daily amount/duration, frequency per week and amount per day of alcohol drinking (regardless of the type of alcohol), type and days per week of physical activity, past medical history and family history), blood pressure, anthropometry (body mass index and waist circumference) and clinical laboratory results (fasting glucose, lipid profile, haemoglobin, urine stick test results, creatinine levels and liver enzyme levels) were included in the health screening database. Some variables have had changes in their measurement methods during the follow-up period. The healthcare utilisation database was based on data collected during the process of claiming healthcare services and included information on inpatient and outpatient medical services (diagnosis, length of stay, services provided and treatment costs) and prescription records (drug codes, days prescribed and daily dosage). The healthcare provider database included information on the types, personnel and equipment of healthcare facilities. The LTCI database included information on applications for long-term care service and the utilisation of such services (activities of daily living, cognitive function, nursing care needs, rehabilitation needs, service grade and type of service).

### Patient and public involvement

This data set was drawn from a retrospective cohort based on administrative data, and separate patient recruitment procedures were not carried out. As the data were de-identified, the consent of the subject and direct contact were not applicable.

### Findings to date

Since the NHIS-Senior was launched in December 2015, several studies using the NHIS-Senior cohort database have been published. The published studies have examined topics emerging as important issues in Korea, such as the risk of dementia,<sup>14–16</sup> the risk of osteoporotic fracture and hip surgery<sup>17 18</sup> and associations of body anthropometry (body mass index and waist

circumference) with mortality.<sup>19 20</sup> Although numerous studies have not yet investigated these issues, other possible topics include functional disabilities and lifestyle modifications in the elderly population. Some studies have used the LTCI database, not the NHIS-Senior cohort, to evaluate the effectiveness of introducing long-term care services.<sup>21 22</sup>

We herein present the basic statistics of the NHIS-Senior cohort for future data users. We calculated the healthcare utilisation and mortality rates. The rates were age-standardised using the census population of Statistics Korea in 2005 and the world standard population.<sup>23</sup> The rates that were standardised using the world standard population are presented below.

The healthcare utilisation and admission rates of 10 major diseases at baseline are presented in online supplementary tables 3 and 4. The rates were the highest for acute upper respiratory infections and influenza (75.2%), followed by disorders of the teeth and supporting structures (40.8%) and other diseases of the eye and adnexa (30.8%). The mortality rates of the cohort population are presented in table 4. We calculated mortality rates using the entire sample data of the NHIS-Senior cohort from 2003 to 2012. The age-standardised (defined with reference to the world standard population) mortality rate for the first 2 years (through 2004) was 3528 per 100 000 person-years, while the rate for 5 years (through 2007) was 3821 per 100 000 person-years and the rate for 10 years (through 2012) was 4333 per 100 000 person-years. In men, the mortality rate was higher than in women (2-year mortality rates of 4688 per 100 000 person-years for men and 2819 per 100 000 person-years for women) ( $p < 0.001$ ).

The major causes of death during the follow-up period (2003 to 2015) are presented by sex in table 5. Causes of death were classified using the list of 56 causes of death used by Statistics Korea, which was derived from the list of 80 causes of death recommended by the WHO for the tabulation of mortality statistics. Malignant neoplasms were the most common cause of death in both sexes (1032.1 per 100 000 person-years for men, 376.7 per 100 000 person-years for women). Cerebrovascular diseases were the second most common cause of death in both men (386.0 per 100 000 person-years) and women (256.0 per 100 000 person-years). Heart disease was the third most common cause of death in both men (247.5 per 100 000 person-years) and women (190.8 per 100 000 person-years). Diabetes mellitus was the fourth most common cause of death in both men (143.8 per 100 000 person-years) and women (101.3 per 100 000 person-years).

Information regarding the long-term care service grade and functional impairment score is shown in online supplementary table 5. A total of 34 483 individuals applied for long-term care service in 2008, of whom 17.9% were assessed as grade 1, meaning that they were completely dependent on the help of another person to live daily life.

**Table 5** Cause-specific death rates for leading causes of death (2003 to 2015) and age-standardised (with the 2005 Korean census and world standard populations as references) mortality rates (per 100 000 person-years) in the national health insurance service-senior (NHIS-Senior) cohort

Rank	Cause of death	All						Men			Women		
		Crude rates	Age-standardised rates		Crude rates	Age-standardised rates		Crude rates	Age-standardised rates		Crude rates	Age-standardised rates	
			Census	WHO		Census	WHO		Census	WHO		Census	WHO
1	Malignant neoplasms	578.5	620.3	634.7	934.3	1003.8	1032.1	337.8	364.4	376.7	337.8	364.4	376.7
2	Cerebrovascular diseases	265.3	286.1	307.0	336.7	361.7	386.0	217.0	235.7	256.0	217.0	235.7	256.0
3	Heart disease	191.5	202.8	222.7	233.3	247.5	267.6	160.6	170.7	190.8	160.6	170.7	190.8
4	Diabetes mellitus	106.1	114.2	119.1	130.9	138.6	143.8	87.8	96.2	101.3	87.8	96.2	101.3
5	Chronic lower respiratory diseases	73.2	78.7	87.1	129.2	67.8	154.8	48.3	49.1	58.7	48.3	49.1	58.7
6	Pneumonia	64.9	67.5	78.7	96.4	101.5	117.4	47.8	50.0	57.6	47.8	50.0	57.6
7	Intentional self-harm (suicide)	51.0	53.8	54.5	84.1	88.3	90.0	41.6	43.7	49.8	41.6	43.7	49.8
8	Hypertensive diseases	47.5	49.3	56.3	67.5	67.8	68.0	27.5	29.7	30.3	27.5	29.7	30.3
9	Diseases of liver	40.3	40.7	41.0	56.3	58.5	59.0	24.0	23.8	28.9	24.0	23.8	28.9
10	Transport accidents	34.8	36.3	63.4	43.7	45.4	50.7	19.2	20.0	20.7	19.2	20.0	20.7

\*The cause of death was classified using the list of 56 causes of death provided by Statistics Korea, which originated from the list of 80 causes of death for the tabulation of mortality statistics recommended by the WHO.



## Strengths and limitations

The NHIS-Senior cohort provides nationally representative cohort data regarding the elderly population in Korea. The NHIS-Senior cohort has several strengths. First, it provides reliable data based on a large sample (n=558 147) and a long duration of follow-up (from 2002 to 2015), and is representative of the entire elderly population. Second, due to the characteristics of the national administration data, the NHIS-Senior cohort has a very low attrition rate and includes more valid and accurate information than self-reported questionnaire-based survey data, especially for socio-economical status, healthcare utilisation and death information. Third, it is possible to analyse the pre-hospital stages of disease using information from the health screening programme and long-term care services.

The NHIS-Senior cohort has certain limitations. First, some of the subjects did not participate in the health screening programme or receive long-term care services due to issues regarding service eligibility. Therefore, there is a possibility of selection bias in health screening information. Second, variables on health behaviours are limited since those data were obtained from self-reporting questionnaires in nationwide health screenings. Third, the disease codes might not accurately reflect patients' medical conditions, as they are sometimes exaggerated to receive reimbursement due to fee-for-service payment system.<sup>24</sup>

**Contributors** YIK, YYK, JLY, CWW, SJ, HK contributed to the conception of this article. YYK, JHP, BRP were involved in manuscript writing and revision. YYK, SH, KDC, SB, EL, SYP, KL, DL were involved in data analysis and interpretation. All authors read and approved the final manuscript.

**Funding** This work was supported by National Health Insurance Service (NHIS) in Korea.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

1. Organisation for Economic Co-operation and Development. *Pensions at a Glance 2015: OECD and G20 Indicators*: OECD, 2015.
2. Korea S. *Population Projections for Korea: 2015-2065*. Daejeon: Statistics Korea, 2016.
3. Organisation for Economic Cooperation and Development. *Facing the future: Korea's family, pension and health policy challenge*. Paris: OECD, 2007.
4. Peabody JW, Gertler PJ, Yang BM, et al. 12. Financing health care for the elderly: will an aging population end 'health for all' in South Korea? *The Economics of Health Care in Asia-Pacific Countries* 2002:241.
5. Chung KH. *The living profile and welfare service needs of older persons in Korea*. Seoul: Korean Institute for Health and Social Affairs, Ministry of Health and Welfare, 1998.
6. Park MH. *The living profile and welfare service needs of older persons in Korea*. Seoul: Korean Institute for Health and Social Affairs, Ministry of Health and Welfare, 2008.
7. National Health Insurance Services. '*National Health Insurance Act*' & '*Act on Long-term Care Insurance for Senior Citizens*': National Health Insurance Corporation, 2009.
8. Kwon S. Future of long-term care financing for the elderly in Korea. *J Aging Soc Policy* 2008;20:119–36.
9. Cheol Seong S, Kim YY, Khang YH, et al. Data Resource Profile: The National Health Information Database of the National Health Insurance Service in South Korea. *Int J Epidemiol* 2017;46:799–800.
10. Seong SC, Kim YY, Park SK, et al. Cohort profile: the National Health Insurance Service-National Health Screening Cohort (NHIS-HEALS) in Korea. *BMJ Open* 2017;7:e016640.
11. Ministry of Foreign Affairs. *Statistics on Korean emigration registration*. Seoul: Ministry of Foreign Affairs.
12. Christensen S, Johansen MB, Christiansen CF, et al. Comparison of Charlson comorbidity index with SAPS and APACHE scores for prediction of mortality following intensive care. *Clin Epidemiol* 2011;3:203.
13. National Health Insurance Service (NHIS). *National Health Screening Statistical Yearbook*. Seoul: National Health Insurance Service.
14. Hwang D, Kim S, Choi H, et al. Calcium-Channel Blockers and Dementia Risk in Older Adults- National Health Insurance Service - Senior Cohort (2002-2013). *Circ J* 2016;80:2336–42.
15. Kim JH, Lee Y. Dementia and death after stroke in older adults during a 10-year follow-up: results from a competing risk model. *J Nutr Health Aging* 2018;22:297–301.
16. Kim H, Chun HW, Kim S, et al. Longitudinal study-based dementia prediction for public health. *Int J Environ Res Public Health* 2017;14:983.
17. Kim SY, Kim S, Choi SE, et al. Number of daily antihypertensive drugs and the risk of osteoporotic fractures in older hypertensive adults: National health insurance service - Senior cohort. *J Cardiol* 2017;70:80–5.
18. Bu N, Kim S, Choi H, et al. Pre-discharge rehabilitation after hip surgery reduces 30-day readmissions in older adults: National Health Insurance Service-Senior Cohort (2007–2012). *Eur Geriatr Med* 2017;8:430–4.
19. Kim H, Yoon JL, Lee A, et al. Prognostic effect of body mass index to mortality in Korean older persons. *Geriatr Gerontol Int* 2018;18:538–46.
20. Lee S-H, Kim D-H, Lee C-W, et al. Retrospective Study of the Association between Waist Circumference and Mortality in the Korean Elderly. *Korean J Clin Geri* 2017;18:36–43.
21. Lee TW, Yim E, Cho E, et al. Cognitive function, behavioral problems, and physical function in long-term care insurance beneficiaries with dementia in South Korea: comparison of home care and institutional care services. *J Am Geriatr Soc* 2014;62:1467–75.
22. Jung W-S, Yim E-S. The Effect on Health Care Utilization of the Non-Use of Beneficiaries of Long-Term Care Insurance Service - around of Geriatric Hospital's Medical Cost -. *Journal of the Korea Academia-Industrial cooperation Society* 2015;16:7463–73.
23. Ahmad OB, Boschi-Pinto C, Lopez AD, et al. *Age standardization of rates: a new WHO standard*. Geneva: World Health Organization, 2001:9.
24. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol* 2004;57:131–41.