

























Developing a national surveillance system for stroke and acute myocardial infarction using claims data in the Republic of Korea: a retrospective study

Tae Jung Kim^{1,2} , Hak Seung Lee³ , Seong-Eun Kim⁴ , Jinju Park⁵ , Jun Yup Kim⁴ , Jiyeon Lee⁶ , Ji Eun Song⁶ , Jin-Hyuk Hong⁵ , Joongyub Lee⁷ , Joong-Hwa Chung⁸ , Hyeon Chang Kim⁹ , Dong-Ho Shin¹⁰ , Hae-Young Lee¹¹ , Bum Joon Kim¹² , Woo-Keun Seo¹³ , Jong-Moo Park¹⁴ , Soo Joo Lee¹⁵ , Keun-Hwa Jung² , Sun U. Kwon¹² , Yun-Chul Hong⁷ , Hyo-Soo Kim¹¹ , Hyun-Jae Kang¹¹ , Juneyoung Lee⁶ , Hee-Joon Bae⁴ 

¹Department of Critical Care Medicine, Seoul National University Hospital, Seoul, Republic of Korea

²Department of Neurology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea

³Medical AI Co., Ltd., Seoul, Republic of Korea

⁴Department of Neurology, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea

⁵Central Division of Cardio-cerebrovascular Disease Management, Seoul National University Hospital, Seoul, Republic of Korea

⁶Department of Biostatistics, Korea University College of Medicine, Seoul, Republic of Korea

⁷Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

⁸Department of Cardiology, Chosun University Hospital, Gwangju, Republic of Korea

⁹Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

¹⁰Department of Cardiology, Severance Cardiovascular Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

¹¹Department of Internal Medicine and Cardiovascular Center, Seoul National University Hospital, Seoul, Republic of Korea

¹²Department of Neurology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

¹³Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

¹⁴Department of Neurology, Uijeongbu Eulji Medical Center, Eulji University, Seoul, Republic of Korea

¹⁵Department of Neurology, Daejeon Eulji Medical Center, Eulji University, Daejeon, Republic of Korea

Received: September 3, 2023

Revised: November 30, 2023

Accepted: December 3, 2023

Corresponding author:

Hee-Joon Bae

Department of Neurology, Seoul National University Bundang Hospital, Seoul National University College of Medicine, 82 Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Republic of Korea
E-mail: braindoc@snu.ac.kr

Objectives: Limited information is available concerning the epidemiology of stroke and acute myocardial infarction (AMI) in the Republic of Korea. This study aimed to develop a national surveillance system to monitor the incidence of stroke and AMI using national claims data.

Methods: We developed and validated identification algorithms for stroke and AMI using claims data. This validation involved a 2-stage stratified sampling method with a review of medical records for sampled cases. The weighted positive predictive value (PPV) and negative predictive value (NPV) were calculated based on the sampling structure and the corresponding sampling rates. Incident cases and the incidence rates of stroke and AMI in the Republic of Korea were estimated by applying the algorithms and weighted PPV and NPV to the 2018 National Health Insurance Service claims data.

Results: In total, 2,200 cases (1,086 stroke cases and 1,114 AMI cases) were sampled from the 2018 claims database. The sensitivity and specificity of the algorithms were 94.3% and 88.6%

for stroke and 97.9% and 90.1% for AMI, respectively. The estimated number of cases, including recurrent events, was 150,837 for stroke and 40,529 for AMI in 2018. The age- and sex-standardized incidence rate for stroke and AMI was 180.2 and 46.1 cases per 100,000 person-years, respectively, in 2018.

Conclusion: This study demonstrates the feasibility of developing a national surveillance system based on claims data and identification algorithms for stroke and AMI to monitor their incidence rates.

Keywords: Incidence; Myocardial infarction; Population characteristics; Population surveillance; Stroke

Introduction

An accurate computation of national health metrics, such as the incidence rates of stroke and acute myocardial infarction (AMI), is indispensable for crafting a robust national healthcare system adept at managing cardiovascular diseases [1–8]. In Korea, the National Health Insurance Service (NHIS) operates a comprehensive claims database that contains data from healthcare providers and insurers for reimbursement purposes. Given that this database covers the vast majority of the Korean population, it contains patient-level data pertaining to diagnoses, treatments, healthcare utilization, access, outcomes, and costs [9–13]. Consequently, this extensive dataset provides a promising foundation for establishing a national surveillance system for stroke and AMI.

Previous research efforts utilizing claims data to explore the incidence of stroke and AMI have encountered limitations in case identification. These challenges primarily originate from the use of International Classification of Diseases (ICD) codes, which are often not well-validated and are challenging to use for differentiating between acute and chronic events [14–22]. In 2004, there was an attempt to construct a surveillance system based on claims data, incorporating a validated diagnostic tool, to estimate the incidence of stroke and AMI. However, this system was predominantly centered around monitoring care quality at individual hospitals, relying heavily on ICD codes [23]. Since this attempt, there have been no additional endeavors to construct a similar system with national coverage. The present study seeks to fill this gap by constructing a national surveillance system for stroke and AMI, harnessing the NHIS claims data and employing validated identification algorithms for these conditions.

Materials and Methods

Overview of the Study Process and the Organization of Study Teams

This study was launched with the aim of constructing a

HIGHLIGHTS

- This study demonstrates the feasibility of creating a national surveillance system using claims data and identification algorithms to estimate the incidence of stroke and acute myocardial infarction.
- The age- and sex-standardized incidence rates stood at 180.2 per 100,000 person-years for stroke and 46.1 per 100,000 person-years for acute myocardial infarction.
- This system facilitates ongoing monitoring of the burden of stroke and cardiovascular disease in the Republic of Korea and aids in expanding nationwide epidemiological research.

national surveillance system for the incidence of stroke and AMI as a private consignment project of the Korean Disease Control and Prevention Agency (KDCA). It required collaboration of a variety of organizations, encompassing (1) the Central Support Group of Cardiovascular Disease Management, Public Health and Medical Services, Seoul National University Hospital; (2) the Department of Biostatistics, Korea University College of Medicine; (3) the Korean Stroke Society, (4) the Korean Society for Preventive Medicine, and (5) the Korean Society of Cardiology. The structure of the study teams is illustrated in [Figure 1A](#).

The study was conducted in the following steps: first, we formulated definitions for epidemiological indices pertaining to stroke and AMI. This was achieved through interactive consultations involving the Advisory Board for Disease Definition, 3 participating societies, and other experts ([Figure 1A](#)). Second, we designed identification algorithms for stroke and AMI based on claims data. Third, the developed algorithms underwent validation via a review of hospital records for cases selected from the NHIS claims data. Finally, we calculated the incidence of AMI and stroke in the Republic of Korea by applying the weighted positive predictive values (PPVs) and negative predictive values (NPVs) to the 2018 NHIS claims database ([Figure 1B](#)).

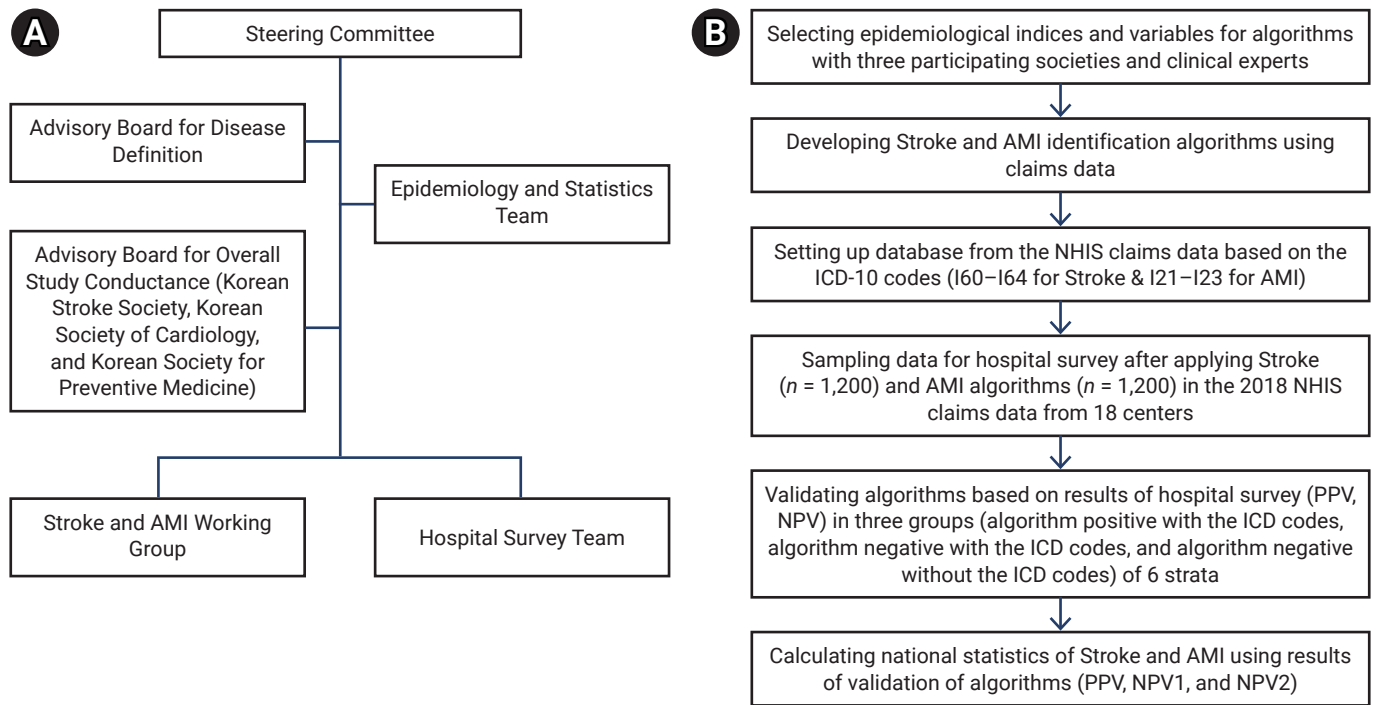


Figure 1. (A) Organization of the study teams. (B) Overview of the study process. AMI, acute myocardial infarction; ICD, International Classification of Diseases; NHIS, National Health Insurance Service; PPV, positive predictive value; NPV, negative predictive value.

Epidemiological Indices, Definitions, and Identification Algorithms

To establish the epidemiological indices for stroke and AMI, we conducted an extensive review of the pertinent literature, including reports from the Organisation for Economic Co-operation and Development (OECD), the National Institute for Health and Care Excellence, and the OECD Health Care Quality Indicators project. Based on this review, we identified the incidence rate and proportion of stroke and AMI as the epidemiological indices for our study [24–27].

We defined stroke according to the World Health Organization (WHO) criteria, which delineate stroke as the rapid development of clinical signs indicating a focal or global disturbance of cerebral function lasting more than 24 hours or resulting in death, with no apparent cause other than of vascular origin [28].

Regarding AMI, we employed the fourth universal definition, which characterizes it as the presence of AMI accompanied by clinical evidence of acute myocardial ischemia and abnormal cardiac biomarkers [29].

The subsequent step was the development of identification algorithms. We selected relevant variables from claims data, based on our understanding of AMI and stroke management within the Korean healthcare setting. We designed an identification algorithm for stroke that encompassed

both ischemic and hemorrhagic stroke cases, along with nontraumatic subdural hemorrhage. The algorithm was built around key identifiers, including ICD codes (I60–I64) for stroke and claims codes related to stroke diagnosis and treatments (Figure 2A). To apply this algorithm, we considered possible stroke admission episodes as a single claim or as combining claims if the gap between the discharge date of the first claim and the admission date of the second claim was 2 days or less. Claims with an interval of more than 2 days were deemed as separate episodes, irrespective of whether the consecutive claims were from the same hospital.

Similarly, we developed an identification algorithm for AMI. The algorithm’s key identifiers included ICD diagnosis codes (I21–I23) for AMI and claims codes related to AMI diagnosis and treatments (Figure 2B). In applying this algorithm, possible AMI admission episodes were defined as a single claim or by combining claims if the gap between the discharge date of the first claim and the admission date of the second claim was 3 days or less, or if the gap in admission dates between the first and second claims did not exceed 28 days.

The epidemiological indices and identification algorithms with their key identifiers received approval from experts from the Korean Stroke Society, the Korean Society of Cardiology, and the Korean Society for Preventive Medicine (Figure 1B).

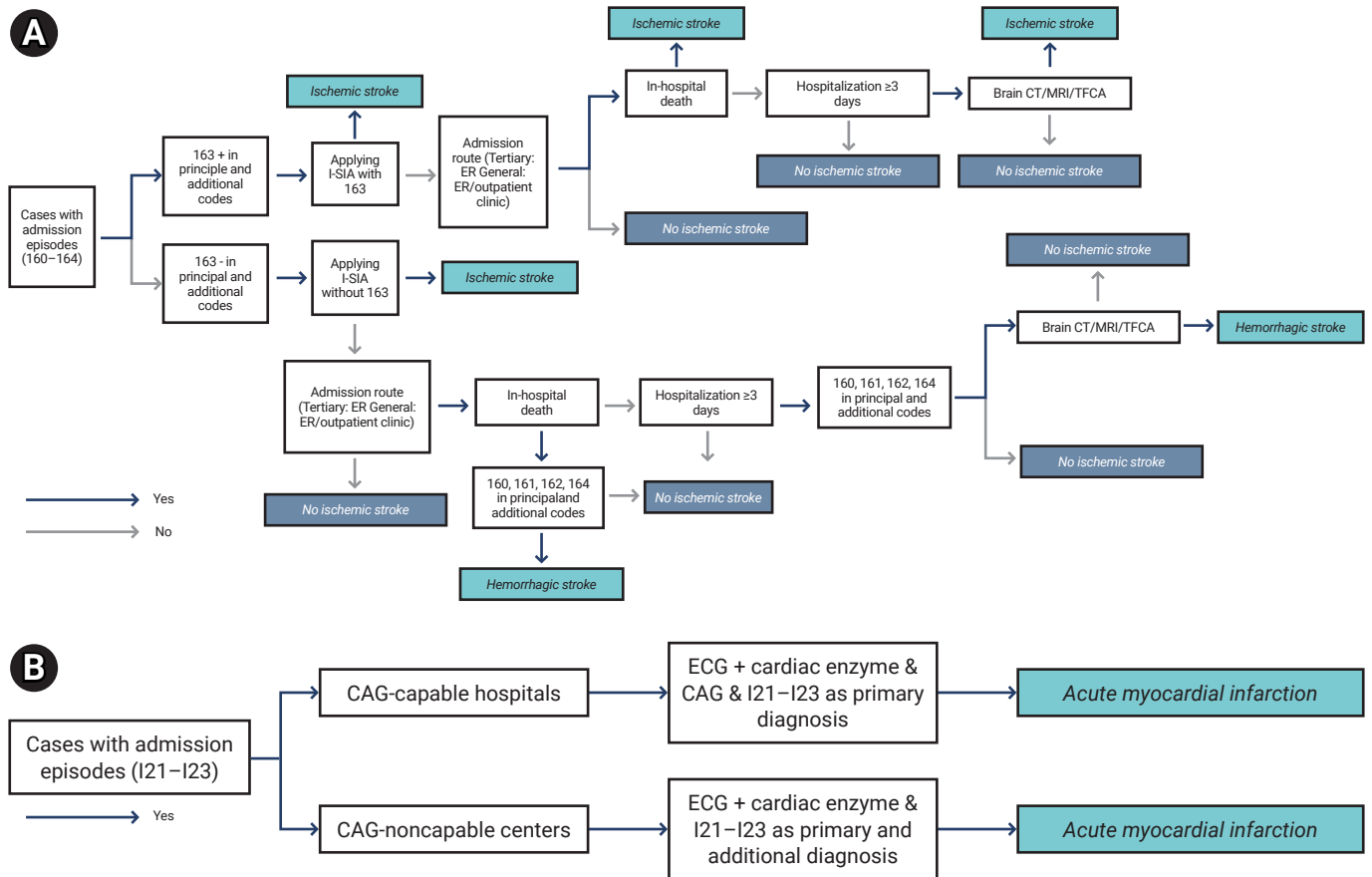


Figure 2. Identification algorithms for stroke and acute myocardial infarction (AMI).

(A) Any stroke identification algorithm (A-SIA). The stroke identification algorithm used in our study was a 2-stage process based on International Classification of Diseases (ICD) codes. In the first stage, the algorithm identified ischemic stroke cases, while the second stage focused on identifying hemorrhagic stroke cases. Each stage utilized a set of key identifiers associated with clinical practices during acute stroke management. A total of 23 key identifiers were employed in this algorithm. These identifiers included variables related to various aspects of stroke care, such as the admission route, stroke unit care procedures, brain imaging techniques (such as computed tomography [CT], CT angiography, brain magnetic resonance imaging [MRI], and transfemoral cerebral angiography [TFCA]) used for stroke diagnosis. It also included recanalization therapy (comprising intravenous thrombolysis and endovascular therapy) administered as hyperacute management post-ischemic stroke, antithrombotic therapy (including antiplatelet agents and anticoagulants), and interventional therapies (such as carotid endarterectomy or carotid and intracranial angioplasty/stenting) for secondary prevention after ischemic stroke. Surgical therapy, rehabilitation, and outcomes at discharge (such as length of stay and in-hospital mortality) were also incorporated into the key identifiers. I-SIA, ischemic stroke identification algorithm. (B) AMI identification algorithm. The AMI identification algorithm used in our study involved a classification process based on a hospital's capacity to perform coronary angiography (CAG). Subsequently, an AMI event was defined when the following conditions were met in CAG-capable hospitals: (1) diagnosis codes I21–I23 were present; (2) an electrocardiogram (ECG) was performed; (3) the serum troponin I or troponin T level was tested; and (4) CAG was performed during hospitalization. Conversely, in CAG-incapable hospitals, an AMI event was defined as follows: (1) diagnosis codes I21–I23 were present, (2) an ECG was performed, and (3) the patients underwent serum troponin I or troponin T testing during hospitalization.

Sampling

We applied 2-stage stratified sampling to select cases for the hospital survey [30,31]. Using the 2018 NHIS claims data, we defined admission episodes with and without the corresponding ICD codes for stroke (I60–I64) and AMI (I21–I23), as described earlier. These episodes were used for hospital and case selection.

In the first stage, we selected hospitals. Initially, we chose hospitals eligible for the survey from 6 administrative divisions—namely, Seoul, Gyeonggi, Daegu, Gyeongsangnam-

do, Ulsan, and Busan—considering the feasibility of the survey. Subsequently, we determined the number of hospitals to invite for participation, considering the geographic regions and case availability. We also balanced the ratio of tertiary to general hospitals at 8:10 as guided by data from the 8th Acute Stroke Quality Assessment Program (ASQAP) [32]. The chosen hospitals were required to have over 20 episodes of both stroke and AMI in the year 2018. Our initial plan was to include a total of 18 hospitals in the survey. These were distributed across the 6 strata as follows: 3 tertiary

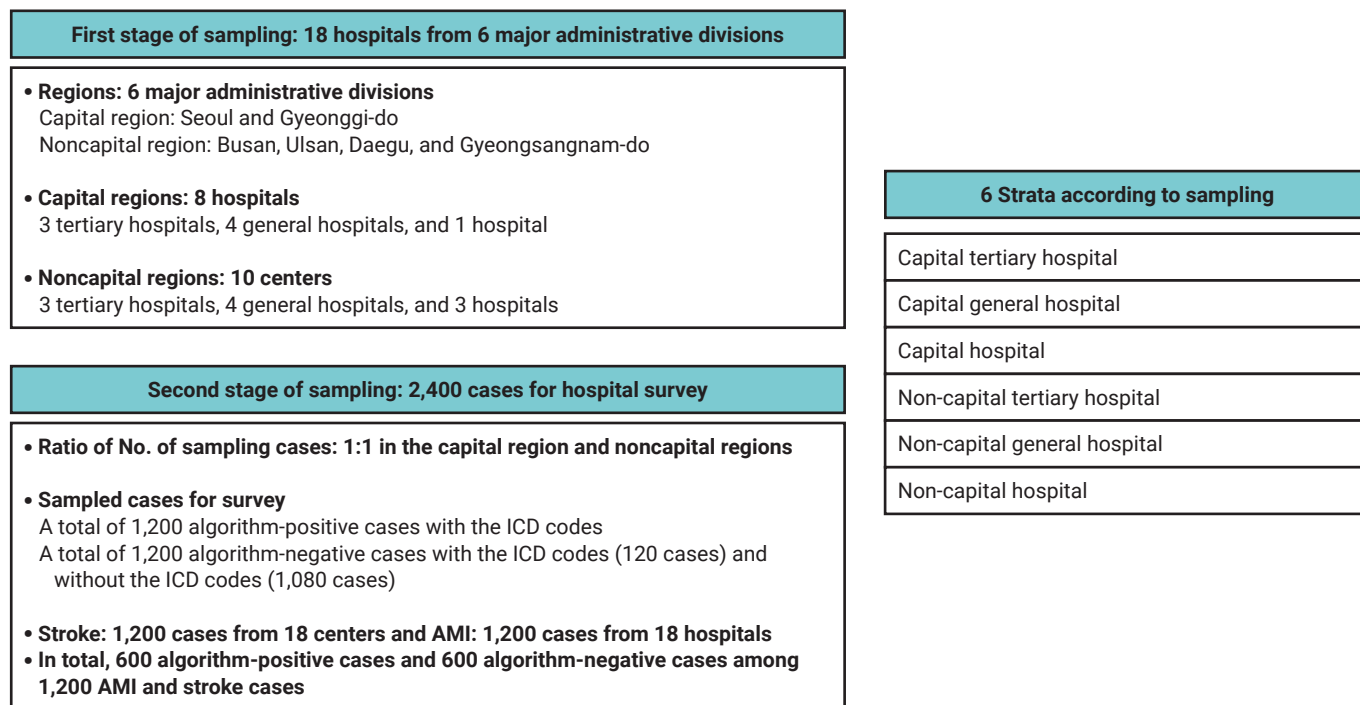


Figure 3. Two-stage stratified sampling method for the algorithm-positive and algorithm-negative groups. Initially, 6 major administrative divisions and 18 hospitals were selected based on the feasibility of the survey. In the second stage, a specific number of cases from the 18 hospitals were determined for the survey. Ultimately, 6 strata were selected for the survey. The ratio of the numbers of tertiary hospitals to general hospitals was 8:10, based on the 8th Acute Stroke Quality Assessment Program (ASQAP) data, in which the number of tertiary hospitals and general hospitals was 42 and 356, respectively, with a sampling fraction of 20% in tertiary hospitals and 3% in general hospitals. The ratio of the sampled case volume was 1:1 between capital and non-capital regions and 6.5:2.5:1 between tertiary hospitals, general hospitals, and other hospitals, drawing from the 8th ASQAP data. For sampling cases for the hospital survey, stroke and acute myocardial infarction (AMI) algorithms were applied to cases with and without the International Classification of Diseases (ICD) codes. Cases with ICD codes were divided into 2 groups: the algorithm-positive group and the algorithm-negative group. In addition, algorithm-negative group without the ICD codes were identified by applying the algorithms to the 2018 National Health Insurance Service (NHIS) claims data.

hospitals, 4 general hospitals, and 1 other hospital from the capital region, and 3 tertiary hospitals, 4 general hospitals, and 3 other hospitals from the non-capital region (Figure 3; Table S1).

In the second stage, we determined the number of sampled cases with and without the ICD codes for stroke and AMI for each hospital and stratum. We employed the optimal allocation method [33,34] and adhered to these guidelines: (1) the ratio of sampled case volume was 1:1 between capital and non-capital hospitals, and 6.5:2.5:1 between tertiary, general, and other hospitals, based on data from the 8th ASQAP (Figure 3); (2) the ratio between algorithm-positive and algorithm-negative cases was 1:1 within each hospital; and (3) among the algorithm-negative cases in each hospital, 10% had the corresponding ICD codes, while 90% did not. As a result, our goal was to sample and survey a total of 1,200 algorithm-positive cases (600 for stroke and 600 for AMI) and 1,200 algorithm-negative cases (600 for stroke, including 60 with I60–I64 and 540 without, and 600

for AMI, including 60 with I21–I23 and 540 without) from the 18 participating hospitals (Figure 3; Table S1).

Hospital Survey

We enlisted 8 qualified reviewers for the survey with the support of the KDCA. Experts from the stroke and AMI working groups established the protocols for reviewing hospital records (Figure 1A). These reviewers underwent specialized training sessions to master how to evaluate sampled cases within the hospital settings. With the consent and support of the participating hospitals, the trained reviewers carried out the survey, consulting online with stroke and AMI experts as needed. The survey findings were scrutinized and confirmed by the respective stroke and AMI experts.

Validation of the Algorithms

The stroke and AMI algorithms were subjected to rigorous validation using the results derived from the hospital survey. The sensitivity, specificity, accuracy, PPV, and NPVs

(NPV1 and NPV2) with their weighted values were estimated for each of the 6 strata by calculating a total sampling rate, taking into account the sampling structure and the corresponding sampling rates. We obtained 2 sampling rates. To calculate the first sampling rate, we divided the number of cases in selected hospitals by the total number of cases (admission episodes) from the 2018 NHIS claims data. We computed the second sampling rate by dividing the number of sampled cases from selected hospitals for the hospital survey by the number of cases in selected hospitals, which was used as a numerator to calculate the first sampling rate. The stratum-specific total sampling rate was obtained by multiplying these 2 sampling rates in each stratum (Figure 4).

During the hospital survey, the number of cases confirmed

as either stroke/AMI or not were determined within each stratum across 3 groups: algorithm-positive cases with the ICD codes (used for PPV calculation), algorithm-negative cases with ICD codes (used for NPV1 calculation), and algorithm-negative cases without ICD codes (used for NPV2 calculation). By multiplying the obtained weights (the inverse of the total sampling rates) by the number of confirmed cases in each group within each stratum, we obtained the necessary values to calculate the weighted sensitivity, specificity, accuracy, PPV, and 1-NPV.

Estimation of Stroke and AMI Incidences

To obtain the number of incident stroke and AMI cases in the entire population, as well as to break these cases down by sex, we applied the algorithms on the 2018 NHIS claims

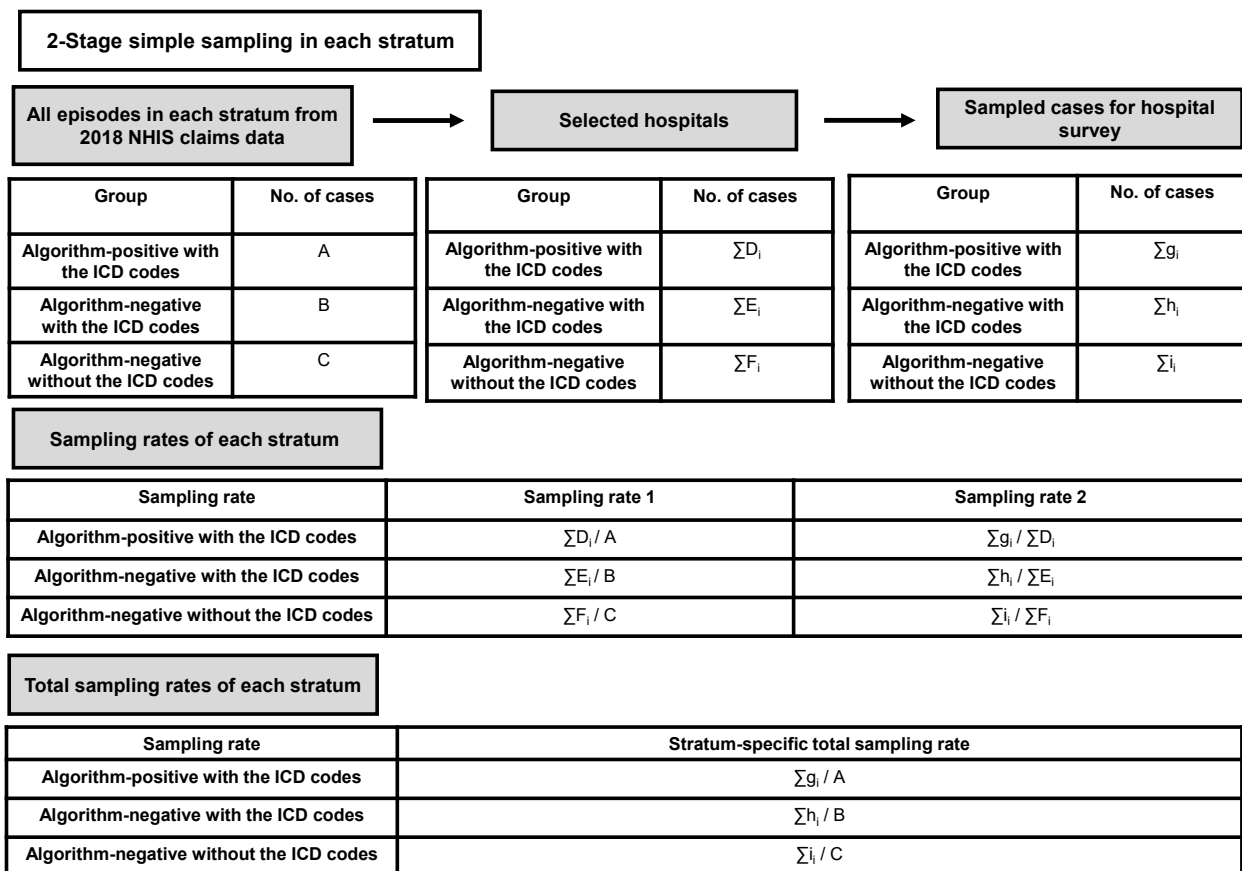


Figure 4. Calculation of total sampling rates in each stratum.

In the initial stage of our sampling process, total cases were partitioned into 3 groups: (1) algorithm-positive cases with the International Classification of Diseases (ICD) codes, (2) algorithm-negative cases with the ICD codes, and (3) algorithm-negative cases without the ICD codes. Sampling rate 1 was determined by dividing the total number of cases in selected hospitals, after applying the algorithms and ICD codes, by the total number of cases derived from applying the algorithm and ICD codes to the 2018 National Health Insurance Service (NHIS) claims data. This calculation was made for each stratum. Sampling rate 2 was calculated by dividing the number of sampled cases for the hospital survey by the total number of cases in selected hospitals after implementing the algorithms in each stratum. The total sampling rate was calculated by multiplying sampling rate 1 by sampling rate 2. This provided a comprehensive measure of the sampling efficiency for the hospital survey, accounting for both the algorithm-based selection and the actual case sampling.

data and ascertained the number of cases within individual strata. The number of stroke and AMI cases within each stratum was computed by multiplying the stratum-specific weighted PPV by the number of algorithm-positive cases with ICD codes and the 1-NPV by the number of algorithm-negative cases with the ICD codes (NPV1) and without the ICD codes (NPV2). The resulting numbers were then added within each stratum. Summing up the numbers of stroke and AMI cases across all 6 strata provided an estimate of the incident case number of stroke and AMI in 2018 (Figure 5).

The incidence rate of stroke and AMI was calculated by dividing the total number of new stroke or AMI cases in 2018, including recurrent events, by the total person-time observed in 2018, as shown in the equation below [24–27].

$$\text{Incidence rate} = \frac{\text{No. of all occurrence of stroke or AMI in 2018}}{\text{Total person - Time of observation in 2018}}$$

The incidence proportion of stroke and AMI represents the proportion of individuals who developed stroke or AMI

among the entire Korean population in 2018. For the incidence proportion, the numerator is the number of patients who experienced stroke or AMI in 2018, while the denominator is the total population at risk in mid-2018, as illustrated in the equation below [24–27]. Recurrent events were not counted.

$$\text{Incidence proportion} = \frac{\text{No. of patients with stroke or AMI in 2018}}{\text{Total population at risk in the middle of 2018}}$$

Finally, the age- and sex-standardized incidence rate and proportion, as well as the age-standardized rate and proportion, were estimated using the mid-year population of the Republic of Korea in 2005 as a benchmark. For age standardization, the WHO standard population in 2000 served as the reference.

Statistical Analysis

In this study, continuous variables were expressed as means and standard deviations, while categorical variables were presented as counts and percentages. The sensitivity,

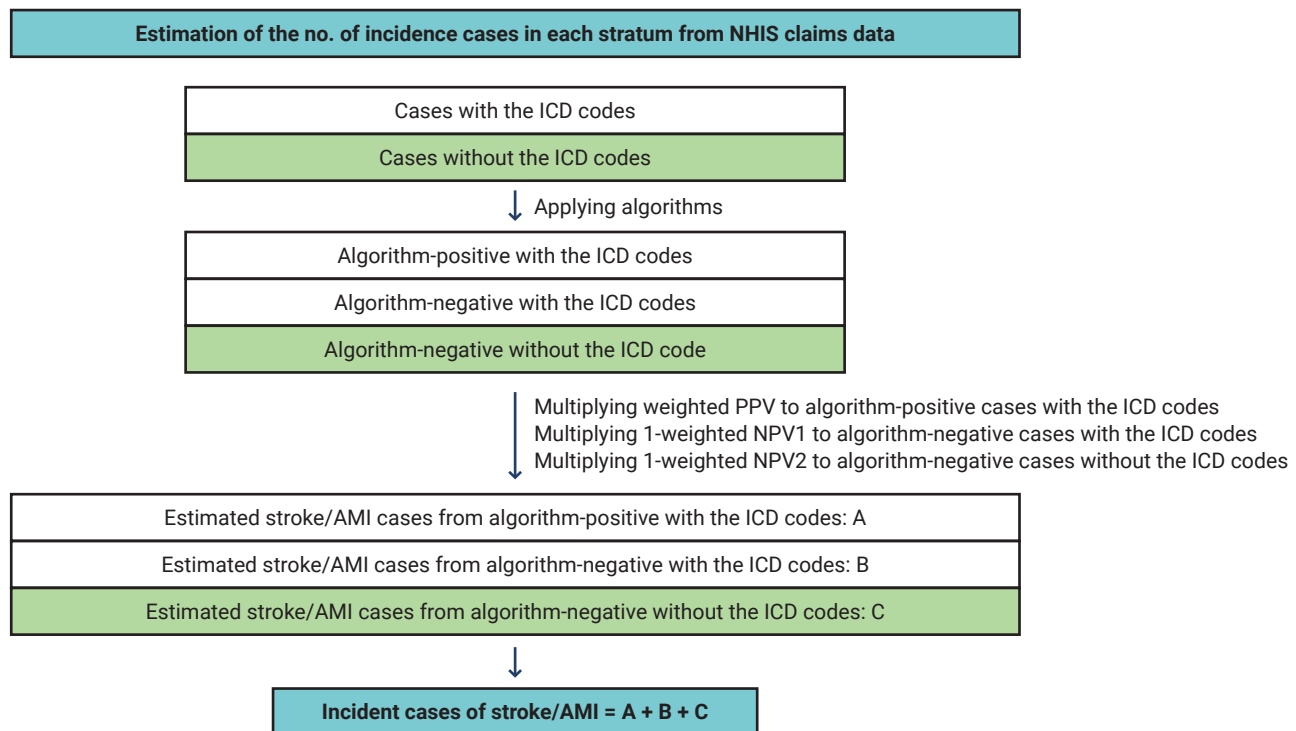


Figure 5. Estimation of the number of stroke and acute myocardial infarction (AMI) cases in each stratum. Initially, we applied the algorithms to cases both with and without the International Classification of Diseases (ICD) codes to estimate the number of algorithm-positive and algorithm-negative cases. Subsequently, we computed the weighted values for each category. For algorithm-positive cases with the ICD codes, we multiplied the weighted PPV by the number of cases. For algorithm-negative cases with the ICD codes, we multiplied 1 minus the weighted NPV1 by the number of cases. Similarly, for algorithm-negative cases without the ICD codes, we multiplied 1 minus the weighted NPV2 by the number of cases. Finally, we derived the incident cases of stroke or AMI by adding these values (A, B, and C) calculated for each category. This approach allowed us to estimate the overall number of incident cases, taking into account the performance of the algorithm and the presence or absence of ICD codes in the data. NHIS, National Health Insurance Service; PPV, positive predictive value; NPV, negative predictive value.

specificity, PPV, and NPV values, as weighted, were provided for the algorithms identifying stroke and AMI. Additionally, 95% confidence intervals (CIs) for incidence rates and incidence proportions were calculated using either the normal approximation or gamma confidence limits when appropriate. All statistical analyses were conducted using SAS ver. 9.4 (SAS Institute Inc.).

Ethics Approval

The Institutional Review Board (IRB) of Seoul National University (IRB No: E-2104-135-1213, E-2109-031-1252, and H-2106-064-1225) approved this study. Additionally, the requirement for informed consent was waived by the IRB because of the retrospective nature of this study.

Results

Study Milestones and Timeline

This study spanned 1 year, running from April 2021 through March 2022, as detailed in Figure 6. While we were on track in the early study period, securing IRB approval, recruiting hospital record reviewers, and hosting their training sessions, challenges arose when it came to case selection for the hospital survey and securing collaboration from the targeted

hospitals. These delays impacted our roster of reviewers, necessitating a fresh round of recruitment and training. Moreover, the time constraints that resulted from these delays put a squeeze on our hospital survey duration and our schedule for estimating epidemiological indices by applying the validated algorithms to the 2018 NHIS claims data. Despite these hurdles, we successfully validated our identification algorithms for stroke and AMI, and derived estimates on the epidemiological indices for 2018 in Korea.

Establishment of Epidemiological Indices, Definitions and Identification Algorithms

The stroke identification algorithm incorporated 23 key identifiers pertinent to acute stroke management, such as brain imaging, stroke unit care, reperfusion therapy, antithrombotic therapy, and rehabilitation. This algorithm was primarily based on a previously developed ischemic stroke identification algorithm (Figure 2A; Table S1) [13]. To ensure the detection of cases with hemorrhagic stroke in addition to ischemic stroke, a 2-tiered strategy was employed. Initially, the previously validated algorithm for ischemic stroke was put into action. Following this, another segment of the algorithm, which included the admission route, in-hospital mortality, length of hospitalization, and

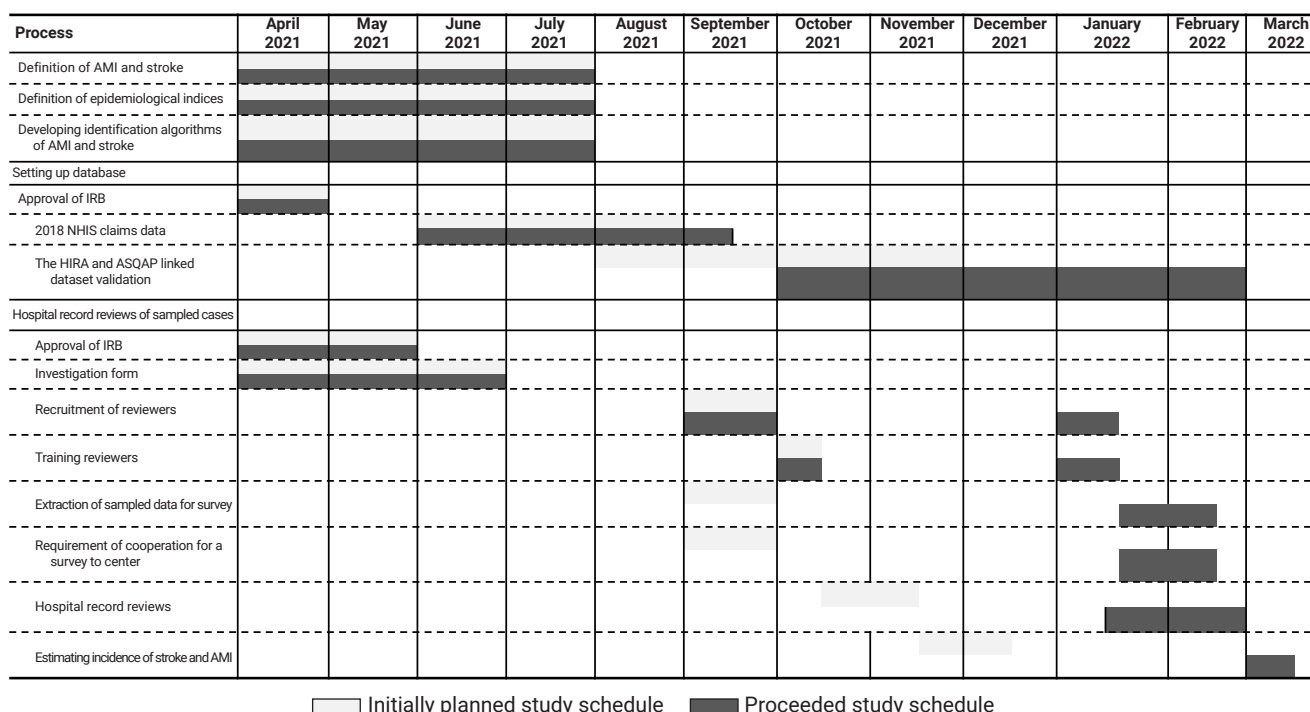


Figure 6. Study milestones.

AMI, acute myocardial infarction; IRB, Institutional Review Board; NHIS, National Health Insurance Service; HIRA, Health Insurance Review and Assessment Service; ASQAP, Acute Stroke Quality Assessment Program.

specific ICD codes, was used to detect hemorrhagic stroke cases. A preliminary survey conducted on 768 cases prior to the main hospital survey demonstrated that the PPV for identifying hemorrhagic stroke stood at an impressive 87.1%.

The AMI identification algorithm was tailored based on whether a hospital had the capacity to perform coronary angiography (CAG) (Figure 2B; Table S1). The categorization of being CAG-capable or not was based on whether the annual number of CAGs performed met or exceeded a threshold of 10. For CAG-capable hospitals, the algorithm included ICD codes, electrocardiogram (ECG) findings, cardiac enzyme levels (such as troponin I or troponin T), and CAG results. For CAG-incapable hospitals, the algorithm included ICD codes, ECG findings, and cardiac enzyme levels. A pilot study carried out before the main hospital survey, which involved a sample size of 757, found the PPV of the algorithm to be 73.3%.

Sampled Cases for the Hospital Survey

Due to the coronavirus disease 2019 (COVID-19)-related regulations and other logistical constraints, the planned hospital survey could not be conducted as initially intended. Instead of the originally planned 18 hospitals, only 14 hospitals participated in the survey. These consisted of 6 hospitals in the capital region (2 tertiary hospitals and 4 general hospitals) and 8 hospitals in the non-capital region (3 tertiary hospitals, 4 general hospitals, and 1 other hospital) (Table S2). The survey could not be carried out in the other hospital stratum of the capital region (Table S3).

There were also challenges in obtaining a sufficient number of algorithm-negative cases without ICD codes. This issue arose from communication gaps between the KDCA and the study teams, compounded by the tight timeline set for the study. As a result, we had to modify our approach to determining the sampling rates. We leaned on the NIHS-National Sample Cohort data regarding algorithm-negative cases without ICD codes. Further details regarding these adjustments can be found in Methods S1 and Figure S1.

In the hospital record review, we assessed 603 algorithm-positive cases and 483 algorithm-negative cases for stroke. Of these algorithm-negative cases, 54 cases had the ICD codes and 429 cases did not. For AMI, we evaluated 568 algorithm-positive cases and 546 algorithm-negative cases, with 60 of the latter having the ICD codes and 486 not having them (Table 1; Table S3). Through this review, we confirmed 578 cases (age, 72.3 ± 11.9 years; men, 51.9%) as stroke and 520 cases (age, 70.7 ± 10.9 years; men, 72.7%) as AMI (Tables 2 and 3). Of those confirmed as stroke, 88% involved ischemic stroke,

while 98.5% of the AMI cases had undergone CAG. The in-hospital mortality rate was 6.2% for stroke cases and 6.3% for AMI cases.

Performance of the Algorithms

The results from the hospital survey indicated that the algorithm we developed to identify stroke had a sensitivity of 94.3% and a specificity of 88.6%. For AMI detection, the algorithm exhibited a sensitivity of 97.9% and a specificity of 90.1%. Table 4 details the PPV, NPV1, and NPV2 for both the stroke and AMI identification algorithms in each stratum.

The stroke and AMI identification algorithms had high PPVs, indicating a high proportion of true positive cases identified by the algorithms. However, the NPV1 for algorithm-negative cases with the ICD codes was found to be low for stroke, indicating a higher proportion of false negative cases in this category. Additionally, the PPVs for stroke and AMI were found to be lower in tertiary hospitals located in the capital region than in other types of hospitals in the capital region and the non-capital hospitals (Table 4).

Incidence of AMI and Stroke

By applying the algorithms to the 2018 NHIS claims database, we derived estimates of 150,837 incident stroke cases and 40,529 incident AMI cases. Men accounted for 55.7% of stroke cases and 66.8% of AMI cases. It is worth noting that due to the small sample size for NPV1 (comprising 10% of the total algorithm-negative group), we used the overall NPV1 (38.9% for stroke and 81.7% for AMI) instead of detailing results by each stratum to estimate incident cases. In 2018, the crude incidence rates of stroke and AMI were estimated to be 294.9 and 79.2 cases per 100,000 person-years, respectively. After standardization for age and sex, the incidence rates were calculated as 180.2 cases per 100,000 person-years for stroke and 46.1 cases per 100,000 person-years for AMI. Notably, the age-standardized incidence rates of both stroke and AMI were higher in men than in women (Table 5). When applying the individual NPV1 value obtained in each stratum instead of the overall NPV1, we found that the number of incident cases and the crude incidence rate was 152,241 cases and 297.6 per 100,000 person-years for stroke and 42,801 cases and 83.7 per 100,000 person-years for AMI, respectively.

Excluding recurrent cases, the estimated number of incident cases was 131,347 for stroke and 39,720 for AMI. The crude incidence proportions were 256.0 cases per 100,000 people for stroke and 77.4 cases per 100,000 people for AMI. After age and sex standardization, the incidence proportions were determined to be 154.1 cases per 100,000 people for stroke and 44.4 cases per 100,000 people for AMI. Similar to the incidence rates, the age-standardized

Table 1. Results of hospital record reviews in stroke and AMI

Total centers	Stroke	Not a stroke	Total	AMI	Not an AMI	Total
Total cases	578	508	1,086	520	594	1,114
Algorithm-positive with the ICD codes	545	58	603	509	59	568
Algorithm-negative with the ICD codes	33	21	54	11	49	60
Algorithm-negative without the ICD codes	0	429	429	0	486	486
Cases in the capital region						
Tertiary hospitals	128	88	216	117	160	277
Algorithm-positive with the ICD codes	121	29	150	116	24	150
Algorithm-negative with the ICD codes	7	6	13	1	14	15
Algorithm-negative without the ICD codes	0	53	53	0	112	112
General hospitals	111	173	284	124	131	255
Algorithm-positive with the ICD codes	108	15	123	123	12	135
Algorithm-negative with the ICD codes	3	8	11	1	11	12
Algorithm-negative without the ICD codes	0	150	150	0	108	108
Cases in non-capital regions						
Tertiary hospitals	175	130	305	151	159	310
Algorithm-positive with the ICD codes	163	4	167	145	5	150
Algorithm-negative with the ICD codes	12	3	15	6	14	20
Algorithm-negative without the ICD codes	0	123	123	0	140	140
Non-capital regions						
General hospitals	144	97	241	116	132	248
Algorithm-positive with the ICD codes	135	8	143	114	7	121
Algorithm-negative with the ICD codes	9	4	13	2	10	12
Algorithm-negative without the ICD codes	0	85	85	0	115	115
Hospitals	20	20	40	12	12	24
Algorithm-positive with the ICD codes	18	2	20	11	1	12
Algorithm-negative with the ICD codes	2	0	2	1	0	1
Algorithm-negative without the ICD codes	0	18	18	0	11	11

AMI, acute myocardial infarction; ICD, International Classification of Diseases.

Table 2. Baseline characteristics of sampled cases of stroke for the hospital survey

Variable	Total (n=1,086)	Stroke (n=578)	Not a stroke (n=508)	P
Age (y)	64.9±19.0	72.3±11.9	56.5±21.9	<0.001
Male	560 (51.6)	300 (51.9)	260 (51.2)	0.81
Admission routes				<0.001
Direct visit	537 (49.4)	471 (81.5)	66 (13.0)	
During hospitalization	6 (0.6)	6 (1.0)	0 (0.0)	
Transfer	97 (8.9)	85 (14.7)	12 (2.4)	
Unknown	446 (41.1)	16 (2.8)	430 (84.6)	
Type of stroke				<0.001
No stroke	508 (46.8)	0 (0)	508 (100.0)	
Ischemic stroke	511 (47.0)	511 (88.4)	0 (0)	
Hemorrhagic stroke	67 (6.2)	67 (11.6)	0 (0)	
History of MI	15 (1.4)	15 (2.6)	0 (0)	0.39
History of stroke	168 (15.5)	130 (22.5)	38 (7.5)	<0.001
In-hospital mortality	38 (3.5)	36 (6.2)	2 (0.4)	<0.001

Data are presented as mean±standard deviation or n (%).

MI, myocardial infarction.

Table 3. Baseline characteristics of sampled cases of AMI for the hospital survey

Variable	Total (n = 1,114)	AMI (n = 520)	Not an AMI (n = 594)	p
Age (y)	61.3 ± 21.7	70.7 ± 10.9	53.0 ± 25.1	< 0.001
Male	691 (62.0)	378 (72.7)	313 (52.7)	< 0.001
Admission routes				< 0.001
Direct visit	484 (43.5)	408 (78.5)	76 (12.8)	
During hospitalization	15 (1.35)	13 (2.5)	2 (0.3)	
Transfer	107 (9.6)	99 (19.0)	8 (1.3)	
Unknown	508 (45.6)	0 (0.0)	508 (85.5)	
Chest pain	448 (40.2)	409 (78.7)	39 (6.6)	
Serum troponin I or T test	715 (64.2)	518 (99.6)	197 (33.2)	< 0.001
ECG	609 (54.7)	518 (99.6)	91 (15.3)	< 0.001
CAG	586 (52.6)	512 (98.5)	74 (12.5)	< 0.001
History of MI	79 (7.1)	47 (9.0)	32 (5.4)	0.001
History of angina	43 (3.9)	36 (6.9)	7 (1.2)	< 0.001
In-hospital mortality	42 (3.8)	33 (6.3)	9 (1.5)	< 0.001

Data are presented as mean ± standard deviation or n (%).

AMI, acute myocardial infarction; ECG, electrocardiogram; CAG, coronary angiography; MI, myocardial infarction.

Table 4. Validation of the identification algorithms for stroke and AMI after weighting

Region	Hospital	PPV (%)	NPV1 (%)	NPV2 (%)
Stroke				
Capital region	Tertiary hospital	80.7	53.3	100
	General hospital	87.8	69.4	100
Non-capital region	Tertiary hospital	97.6	20.0	100
	General hospital	94.4	31.2	100
	Other hospital	90.0	0	100
AMI				
Capital region	Tertiary hospital	77.3	95.0	100
	General hospital	91.1	91.7	100
Non-capital region	Tertiary hospital	96.7	80.0	100
	General hospital	94.2	83.3	100
	Other hospital	91.7	0	100

AMI, acute myocardial infarction; PPV, positive predictive value; NPV, negative predictive value.

incidence proportions of stroke and AMI were higher in men (Table 5).

Discussion

Our study showed that there were 150,837 incident stroke cases and 40,529 incident AMI cases across the Republic of Korea in 2018. When we excluded recurrent cases, the figures decreased to 131,347 for stroke and 39,270 for AMI. Compared to earlier studies in the Republic of Korea, which ranged from 73,501 to 130,025 [16,23,35], our estimated case number for stroke was much higher. Furthermore, our results exceeded those of the 2019 Global Burden of Disease study, which estimated 92,934 stroke cases in 2019 [2]. Likewise, the number of incident AMI cases in our study was also higher than those documented in previous studies in the Republic of Korea, which ranged from 15,893 to 25,531

[15,16,29], but was slightly lower than the figure reported in the first attempt to develop the national surveillance system (50,879 cases in 2004) [23]. Overall, our crude incidence rate and proportion exceeded those reported in previous Korean studies [15,16,23,35]. The marked rise in incident stroke and AMI cases compared to older studies might be attributed to the swift aging of the Korean population [36] coupled with an increased prevalence of traditional vascular risk factors stemming from the westernization of lifestyles [37]. While directly comparing our findings with these earlier studies can be complex due to variation in reference populations, disease definitions, and standardization methods, the upward trend in stroke and AMI case numbers we observed points to possible socioeconomic burdens linked to these health conditions.

The age- and sex-standardized incidence rates and proportions, based on the 2018 mid-year population of Korea,

Table 5. Crude and age- and sex-standardized incidence of stroke and AMI in 2018

Variable	Stroke	AMI
Total incident cases	150,837	40,529
Total patients	131,347	39,720
Incidence rate		
Total (cases/100,000 person-year) (95% CI)		
Crude incidence rate	294.9 (293.4–296.4)	79.2 (78.5–80.0)
Age, sex-standardized incidence rate ^{a)}	180.2 (178.3–182.2)	46.1 (45.1–47.0)
Age-standardized incidence rate ^{b)}	175.6 (174.6–176.5)	46.0 (45.5–46.4)
Male (cases/100,000 person-year) (95% CI)		
Crude incidence rate	329.1 (326.9–331.3)	106.0 (104.8–107.3)
Age-standardized incidence rate)	196.3 (194.9–197.8)	62.2 (61.4–63.0)
Female (cases/100,000 person-year) (95% CI)		
Crude incidence rate	260.8 (258.9–262.8)	52.6 (51.7–53.5)
Age-standardized incidence rate ^{a)}	164.0 (162.7–165.4)	29.9 (29.3–30.3)
Incidence proportion		
Total (cases/100,000 people) (95% CI)		
Crude incidence proportion	256.0 (254.6–257.4)	77.4 (76.7–78.2)
Age, sex-standardized incidence proportion ^{a)}	154.1 (152.3–155.9)	44.4 (43.5–45.3)
Age-standardized incidence proportion ^{b)}	149.1 (147.1–151.1)	44.2 (43.8–44.7)
Male (cases/100,000 people) (95% CI)		
Crude incidence proportion	283.0 (280.9–285.0)	103.7 (102.5–105.0)
Age-standardized incidence proportion ^{a)}	166.4 (165.1–167.7)	60.2 (59.4–60.9)
Female (cases/100,000 people)		
Crude incidence proportion	229.2 (227.3–231.0)	51.2 (50.4–52.1)
Age-standardized incidence proportion ^{a)}	141.7 (130.5–142.9)	28.6 (28.1–29.1)

AMI, myocardial infarction; CI, confidence interval.

^{a)}2005 Mid-year population in the Republic of Korea. ^{b)}2000 World Health Organization standard population.

were either comparable or slightly lower than those in the earlier Korean studies between 2004 and 2016 [15,16,23,35]. Our study's age-standardized incidence proportion of stroke (at 149.1 per 100,000 people using the 2000 WHO standard population) was similar to or slightly higher than the incidence of first-ever stroke observed in high-income countries [1,2,38–42], but lower than that in other Asian countries [43–45]. Regarding the age-standardized incidence proportion of AMI, our finding of 44.2 per 100,000 people was in line with the incidence of first-ever AMI in high-income Asian countries, but lower than the figures reported in Western countries [1,3,45–48].

We employed a 2-stage stratified sampling method to select cases for the hospital survey considering geographic location and hospital size in order to reflect the variability across different medical settings. This method is widely acknowledged for its effectiveness in estimating sensitivity, specificity, PPV, and NPV in national sample surveys while minimizing the standard error [30,31]. The hospital survey was conducted in 14 hospitals from 5 strata, involving a total of 2,200 cases (1,114 AMI and 1,086 stroke cases). Due to the restricted hospital access amid the COVID-19 pandemic and the constrained timeframe of our study,

the numbers of cases and hospitals for the hospital survey were inadequate for comprehensive algorithm validation. In particular, the proportion of algorithm-negative cases with the corresponding ICD codes was only 10%, which limited the evaluation. This small sample size resulted in the variability of NPV1 across strata (Table 4) and the adoption of the overall NPV1 instead of the stratum-specific NPV1.

The NPV for cases classified as algorithm-negative with the ICD codes demonstrated lower values in our hospital survey. Specifically, the NPV1 for stroke was notably lower than that for AMI. This discrepancy could be largely attributed to the limited sampling of the algorithm-negative cases with ICD codes for the hospital survey: only 54 cases for stroke and 60 cases for AMI were surveyed (Table 1). This small sample size inherently limited the precision of our evaluation. Moreover, the difference can be attributed to the distinct disease characteristics and the nuances in ICD coding for AMI and stroke. The stroke ICD codes, in particular, clearly distinguish between acute and chronic stages of the condition. An analysis solely based on the ICD codes of the 2018 NHIS claims data revealed an overestimation in the reported numbers of both stroke and AMI compared to our estimates for incident stroke and AMI cases (Table

S4). This overestimation seemed to be more apparent for stroke than AMI. Additionally, clinical practices for AMI, including diagnostic testing and treatment, are generally less complex than those for stroke. The complexity of stroke care might have contributed to the lower NPV1 for stroke in our study.

The PPVs in the capital region were lower than those in the non-capital region, and tertiary hospitals in the capital region exhibited the lowest PPVs among the 5 strata. This reduced PPV might be explained by the fact that, in these strata, a considerable number of patients are hospitalized long after the acute phase has passed [49]. The influence of hospital size and geographic location on the PPV and NPV underscores the need for a more comprehensive hospital survey with a larger sample size to procure PPVs and NPVs based on region, size, and other characteristics of our healthcare system.

This study has several limitations. First, we relied on the overall NPV1 rather than the stratum-specific NPV1 due to the limited sample size of algorithm-negative cases with the ICD codes per stratum, which ranged from 1 to 20 for stroke and 2 to 15 for AMI. Second, the use of stratum-specific weighted values, derived from a hospital survey with limited case numbers from 14 hospitals across 5 strata in 6 administrative regions, may impede the generalizability of our findings. This sample might not have adequately represented the diversity of the entire country. Third, our inability to include non-hospitalized incident cases, including pre-hospitalization fatalities, is another limitation. Fourth, the short study period and the small number of patients surveyed restrict our capacity to evaluate the algorithm's accuracy in determining the exact incidence rate. Fifth, we could not estimate the lifetime first-ever incident cases due to inadequate historical data on previous stroke and AMI from the claims data.

However, the strengths of our study are noteworthy. To our knowledge, this is the first study to validate a developed algorithm, including PPV, NPV, sensitivity, and specificity, through an extensive hospital survey using a 2-staged sampling strategy for national representativeness. Unlike previous research focused mainly on the quality of acute care in AMI and stroke cases, our study uniquely estimated incidence rates and proportions using PPV and NPV stratification. Although few studies have specifically evaluated the validity of identification algorithms, our study demonstrated superior sensitivity, specificity, and PPV than those relying predominantly on ICD codes [17–23].

Additionally, our study highlights the feasibility of establishing a national surveillance system using claims

data and identification algorithms for tracking the incidence of stroke and AMI. Such a system is invaluable for ongoing monitoring of these diseases and supporting nationwide epidemiological research. However, to implement effective 2-stage sampling for national hospital surveys and generate comprehensive national statistics, a unified platform for collaboration and streamlined data collection is imperative. Further studies with larger samples and a broader range of hospitals are essential to develop robust sampling strategies, ensuring accurate incidence estimates that reflect diverse healthcare settings.

Looking ahead, this study underscores the necessity for a national surveillance platform to minimize bias through a comprehensive hospital survey spanning the entire country based on claims data. It is important to continue adjusting and updating stroke and AMI identification algorithms based on extensive surveys with sample case numbers and study duration. Developing methods to estimate incident cases, including non-hospitalized ones and fatalities, is also crucial. Expanding the number and diversity of participating hospitals and allowing sufficient time for surveys will enhance the comprehensiveness and representativeness of our statistics. Moreover, the swift adaptation of the ICD-11 classification system is important for improving the accuracy of estimates [49]. Further research and the establishment of a well-organized platform for this surveillance system are essential steps forward.

Supplementary Material

Table S1. Key identifiers of AMI and stroke algorithms; **Table S2.** Planned number of cases for hospital survey; **Table S3.** Number of cases of completed hospital survey; **Table S4.** Case numbers of AMI and stroke solely based on the ICD codes obtained from the 2018 NHIS claims data; **Figure S1.** Modification of the calculation of sampling rates of each stratum for algorithm-negative cases without the International Classification of Diseases (ICD) codes; **Methods S1.** Supplementary method for calculating the sampling rate. Supplementary data are available at <https://doi.org/10.24171/j.phrp.2023.0248>.

Notes

Ethics Approval

The Institutional Review Board (IRB) of Seoul National University (IRB No: E-2104-135-1213, E-2109-031-1252, and H-2106-064-1225) approved this study. Additionally, the requirement for informed consent was waived by the IRB because of the retrospective nature of this study.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Funding

This work was funded by the Korean Disease Control and Prevention Agency.

Availability of Data

All data analyzed during this study are included in this published article. While the datasets are not publicly available, inquiries regarding access to the study data can be directed to the corresponding author.

Authors' Contributions

Conceptualization: HJK, JoL, HJB; Data curation: SEK, JP, JES, JiL; Formal analysis: TJK, HSL, SEK, JP, JES, JiL; Funding acquisition: HJB; Investigation: TJK, HSL, SEK, JYK, JP, JES, JiL; Methodology: TJK, HSL, SEK, JYK, JiL, JHC, HCK, DHS, HYL, BJK, WKS, JMP, SJL, KHJ; Project administration: JHH, JP, JoL, KH, HJK, HJB; Software: SEK, JP, JES, JiL; Supervision: HJK, SUK, YCH, HSK, JuL, HJB; Validation: TJK, HSL, SEK, JYK, JP, JES, JiL; Visualization: TJK, SEK, JP; Writing—original draft: TJK, HJB; Writing—review & editing: all authors. All authors read and approved the final manuscript.

Additional Contributions

We wish to extend our sincere gratitude to the National Health Insurance Service and the Health Insurance Review and Assessment Service for their valuable contribution in providing the data for our research.

References

- Dai H, Much AA, Maor E, et al. Global, regional, and national burden of ischaemic heart disease and its attributable risk factors, 1990-2017: results from the Global Burden of Disease Study 2017. *Eur Heart J Qual Care Clin Outcomes* 2022;8:50–60.
- GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol* 2021;20:795–820.
- Virani SS, Alonso A, Aparicio HJ, et al. Heart disease and stroke statistics: 2021 update: a report from the American Heart Association. *Circulation* 2021;143:e254–743.
- Feigin V, Norrving B, Sudlow CL, et al. Updated criteria for population-based stroke and transient ischemic attack incidence studies for the 21st century. *Stroke* 2018;49:2248–55.
- Krishnamurthi RV, Feigin VL, Forouzanfar MH, et al. Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet Glob Health* 2013;1:e259–81.
- Feigin VL, Roth GA, Naghavi M, et al. Global burden of stroke and risk factors in 188 countries, during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet Neurol* 2016;15:913–24.
- Broderick J, Brott T, Kothari R, et al. The Greater Cincinnati/Northern Kentucky Stroke Study: preliminary first-ever and total incidence rates of stroke among blacks. *Stroke* 1998;29:415–21.
- Zhu Z, Bundy JD, Mills KT, et al. Secular trends in cardiovascular health in US adults (from NHANES 2007 to 2018). *Am J Cardiol* 2021; 159:121–8.
- 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.
- Kwon S. Thirty years of national health insurance in South Korea: lessons for achieving universal health care coverage. *Health Policy Plan* 2009;24:63–71.
- Park JS, Lee CH. Clinical study using healthcare claims database. *J Rheum Dis* 2021;28:119–25.
- Kim TJ, Lee JS, Kim JW, et al. Building linked big data for stroke in Korea: linkage of stroke registry and National Health Insurance claims data. *J Korean Med Sci* 2018;33:e343.
- Kim JY, Lee KJ, Kang J, et al. Development of stroke identification algorithm for claims data using the multicenter stroke registry database. *PLoS One* 2020;15:e0228997.
- Lee HH, Cho SM, Lee H, et al. Korea Heart Disease Fact Sheet 2020: analysis of nationwide data. *Korean Circ J* 2021;51:495–503.
- Kim RB, Kim HS, Kang DR, et al. The trend in incidence and case-fatality of hospitalized acute myocardial infarction patients in Korea, 2007 to 2016. *J Korean Med Sci* 2019;34:e322.
- Kim RB, Kim BG, Kim YM, et al. Trends in the incidence of hospitalized acute myocardial infarction and stroke in Korea, 2006-2010. *J Korean Med Sci* 2013;28:16–24.
- Goldstein LB. Accuracy of ICD-9-CM coding for the identification of patients with acute ischemic stroke: effect of modifier codes. *Stroke* 1998;29:1602–4.
- Benesch C, Witter DM, Wilder AL, et al. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. *Neurology* 1997;49:660–4.
- May DS, Kittner SJ. Use of Medicare claims data to estimate national trends in stroke incidence, 1985-1991. *Stroke* 1994;25:2343–7.
- Hsieh CY, Chen CH, Li CY, et al. Validating the diagnosis of acute ischemic stroke in a National Health Insurance claims database. *J Formos Med Assoc* 2015;114:254–9.
- Shima D, Ii Y, Higa S, et al. Validation of novel identification algorithms for major adverse cardiovascular events in a Japanese claims database. *J Clin Hypertens (Greenwich)* 2021;23:646–55.
- Kumamaru H, Judd SE, Curtis JR, et al. Validity of claims-based stroke algorithms in contemporary Medicare data: reasons for geographic and racial differences in stroke (REGARDS) study linked with medicare claims. *Circ Cardiovasc Qual Outcomes* 2014;7:611–9.
- Health Insurance Review and Assessment Service; Korea Disease Control and Prevention Agency (KDCA). Construction of national surveillance system for cardiovascular & cerebrovascular diseases [Internet]. KDCA; 2006 [cited 2023 Sep 1]. Available from: http://library.nih.go.kr/ncmiklib/archive/rom/reportView.do?upd_yn=Y&rep_id=RP00002977MNU1154-MNU0004-MNU1889&fid=28&q_type=&q_value=&cid=62424&pageNum=1. Korean.
- Porta MS. A dictionary of epidemiology. 5th ed. Oxford University Press; 2008.
- Modig K, Berglund A, Talback M, et al. Estimating incidence and prevalence from population registers: example from myocardial

- infarction. *Scand J Public Health* 2017;45:5–13.
26. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, et al. Myocardial infarction and coronary deaths in the World Health Organization MONICA Project: registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation* 1994;90:583–612.
27. Hopkins RS, Jajosky RA, Hall PA, et al. Summary of notifiable diseases: United States, 2003. *MMWR Morb Mortal Wkly Rep* 2005;52:1–85.
28. Aho K, Harmsen P, Hatano S, et al. Cerebrovascular disease in the community: results of a WHO collaborative study. *Bull World Health Organ* 1980;58:113–30.
29. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). *J Am Coll Cardiol* 2018;72:2231–64.
30. Sedgwick P. Stratified cluster sampling. *BMJ* 2013;347:f7016.
31. McNamee R. Optimal designs of two-stage studies for estimation of sensitivity, specificity and positive predictive value. *Stat Med* 2002;21:3609–25.
32. Park HK, Kim SE, Cho YJ, et al. Quality of acute stroke care in Korea (2008-2014): retrospective analysis of the nationwide and nonselective data for quality of acute stroke care. *Eur Stroke J* 2019;4:337–46.
33. Lai D, Chang KC, Rahbar MH, et al. Optimal allocation of sample sizes to multicenter clinical trials. *J Biopharm Stat* 2013;23:818–28.
34. Casella G, Berger RL. *Statistical inference*. 2nd ed. Thomson Learning; 2002.
35. Ministry of Health and Welfare. Estimating the incidence of acute myocardial infarction and stroke based on the national health insurance claims data in Korea (TRKO201700004461). Ministry of Health and Welfare; 2017. Korean.
36. Kim KW, Kim OS. Super aging in South Korea unstoppable but mitigatable: a sub-national scale population projection for best policy planning. *Spat Demogr* 2020;8:155–73.
37. Jo G, Oh H, Singh GM, et al. Impact of dietary risk factors on cardiometabolic and cancer mortality burden among Korean adults: results from nationally representative repeated cross-sectional surveys 1998-2016. *Nutr Res Pract* 2020;14:384–400.
38. Li L, Scott CA, Rothwell PM, et al. Trends in stroke incidence in high-income countries in the 21st century: population-based study and systematic review. *Stroke* 2020;51:1372–80.
39. Medin J, Nordlund A, Ekberg K, et al. Increasing stroke incidence in Sweden between 1989 and 2000 among persons aged 30 to 65 years: evidence from the Swedish Hospital Discharge Register. *Stroke* 2004;35:1047–51.
40. Hammar N, Alfredsson L, Rosen M, et al. A national record linkage to study acute myocardial infarction incidence and case fatality in Sweden. *Int J Epidemiol* 2001;30 Suppl 1:S30–4.
41. Yafasova A, Fosbol EL, Christiansen MN, et al. Time trends in incidence, comorbidity, and mortality of ischemic stroke in Denmark (1996-2016). *Neurology* 2020;95:e2343–53.
42. Meretoja A, Roine RO, Kaste M, et al. Stroke monitoring on a national level: PERFECT Stroke, a comprehensive, registry-linkage stroke database in Finland. *Stroke* 2010;41:2239–46.
43. Venketasubramanian N, Yoon BW, Pandian J, et al. Stroke epidemiology in South, East, and South-East Asia: a review. *J Stroke* 2017;19:286–94.
44. Tan CS, Muller-Riemenschneider F, Ng SH, et al. Trends in stroke incidence and 28-day case fatality in a nationwide stroke registry of a multiethnic Asian population. *Stroke* 2015;46:2728–34.
45. Wang Y, Zhou L, Guo J, et al. Secular trends of stroke incidence and mortality in China, 1990 to 2016: the Global Burden of Disease Study 2016. *J Stroke Cerebrovasc Dis* 2020;29:104959.
46. Amini M, Zayeri F, Salehi M. Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from Global Burden of Disease Study 2017. *BMC Public Health* 2021;21:401.
47. Roth GA, Mensah GA, Johnson CO, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 Study. *J Am Coll Cardiol* 2020;76:2982–3021.
48. Camacho X, Nedkoff L, Wright FL, et al. Relative contribution of trends in myocardial infarction event rates and case fatality to declines in mortality: an international comparative study of 1.95 million events in 80.4 million people in four countries. *Lancet Public Health* 2022;7:e229–39.
49. World Health Organization (WHO). International classification of diseases 11th revision (ICD-11), 2018 [Internet]. WHO; 2018 [cited 2018 Jun 18]. Available from: <https://icd.who.int/browse11>.