



OPEN

# Lifetime risks and health impacts of hemorrhagic and ischemic stroke in South Korea

Soyeon Cheon<sup>1</sup>, Hyangsook Lee<sup>2,3</sup>, Jiyeon Won<sup>2,3</sup>, Bo-Hyoung Jang<sup>4</sup> & Jung-Der Wang<sup>1</sup>✉

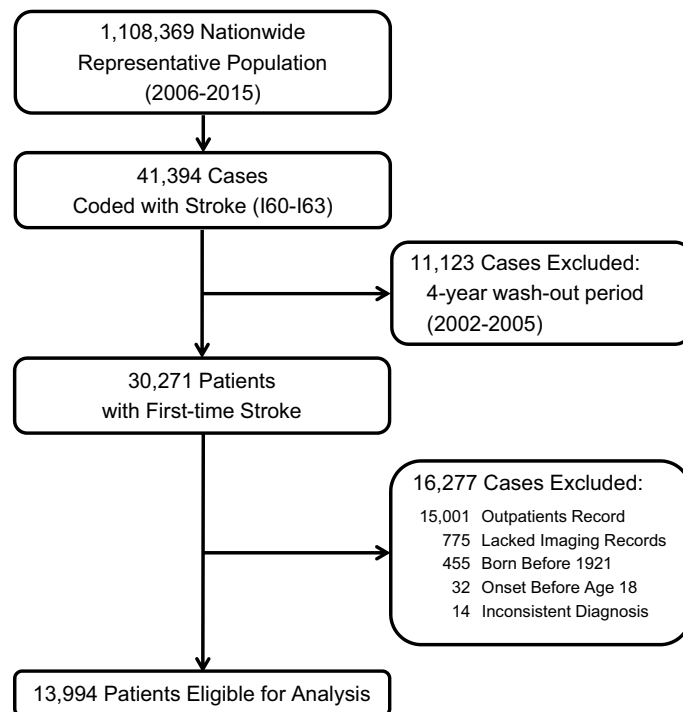
This study is aimed toward estimating the lifetime risks, life expectancy, expected years of life lost (EYLL), and lifetime costs related to different subtypes of stroke in South Korea. We included 13,994 patients diagnosed with stroke (ICD-10, I60-I63) in the National Health Insurance Service-National Sample Cohort of Korea between 2006 and 2015. Lifetime risks were calculated using the cumulative incidence rate for patients aged 18–84. Lifetime survival data were obtained through the Kaplan–Meier method and extrapolated with a rolling-over extrapolation algorithm. The lifetime costs were estimated by multiplying the average monthly expenditures with the survival probabilities and adding the values over lifetime. The lifetime risks of stroke in Korea have been decreasing consistently over the last decade with the exception of subarachnoid hemorrhage in females, which appears to have slightly increased. The EYLL is higher in hemorrhagic stroke than in ischemic stroke (6–9.7 vs. 4.7). Expected lifetime costs reimbursed by the NHIS would amount to about \$71,406 accompanied with \$14,921 copayment from the patients for hemorrhagic stroke, and \$50,551 and \$11,666, respectively, for ischemic stroke. Further studies are warranted to combine survival with quality of life and functional disability to obtain a more detailed outcome assessment of the potential impact of the prevention of stroke.

Being the second most common cause of deaths and the third most common cause of disability worldwide, strokes result not only in mortality but also in morbidities affecting the lives of survivors and their family members<sup>1</sup>. Among the top 30 causes of years of life lost, only seven of them also appear in the top 30 causes of years lived with disability, including stroke<sup>2</sup>. With advances in treatment, as well as the control of hypertension for prevention of stroke during the last couple of decades, we are witnessing a decline in stroke incidence and mortality<sup>3</sup>. In a recently updated report of stroke statistics in South Korea, the age- and sex-standardised incidence rates of first-ever stroke showed a gradual decrease from 105.8 per 100,000 person-years in 2007 to 92.2 in 2013<sup>4</sup>. However, an ageing population as well as population growth itself still continues to increase the absolute burden of stroke<sup>5</sup>, which is also the case in Korea. Stroke has recently dropped to the third most common cause of deaths after cancer (all types combined) and heart disease<sup>6</sup>, but the prevalence and medical expenses related to stroke have been continuously growing over the last decade<sup>7</sup>.

An increasing burden of stroke is inevitable unless there is a significant drop in the incidence rate, and therefore, along with improving quality of care, policy makers must emphasise effective prevention of stroke. Especially in countries with an ageing and/or aged population, there has been a continuous emphasis placed on preventive medicine<sup>1,8</sup>. Nonetheless, the difficulties in quantifying the anticipated savings of life years and/or costs due to prevention have been an obstacle for preventive medicine in terms of receiving sufficient resources<sup>9–11</sup>. In addition to determining the incidence rate as the first step to show the impact of prevention, it is also necessary to quantify the consequences (or pay-off in economic terms) of stroke in units that are directly comparable to other clinical services in order for policy makers to allocate resources fairly.

Although there are a number of studies that describe the various outcomes of stroke in Korea<sup>12–15</sup>, to our knowledge, there is no evidence of the lifetime health benefits and costs savings resulting from successful prevention of stroke. A newly developed method<sup>16–18</sup> for extrapolating lifetime outcomes beyond a given observation

<sup>1</sup>Department of Public Health, National Cheng Kung University College of Medicine, No. 1, University Road, Tainan, Taiwan. <sup>2</sup>Acupuncture and Meridian Science Research Center, College of Korean Medicine, Kyung Hee University, Seoul, Korea. <sup>3</sup>Department of Science in Korean Medicine, Graduate School, Kyung Hee University, Seoul, Korea. <sup>4</sup>Department of Preventive Medicine, College of Korean Medicine, Kyung Hee University, Seoul, Korea. ✉email: jdwang121@gmail.com



**Figure 1.** Flow diagram for establishing the index cohort.

period, which has been approved by international scholars<sup>19</sup>, now makes it possible to obtain an accurate estimation of life expectancy (LE), expected years of life lost (EYLL), and lifetime healthcare expenditures for various diseases. By applying this novel method to real-world Korean data, we aimed to estimate: (1) the lifetime risks, (2) LE and EYLL, and (3) direct medical costs over lifetime for insurers and patients by utilizing a decade of national health insurance reimbursement data from patients diagnosed with hemorrhagic and ischemic stroke in South Korea.

## Results

Reimbursement records for 41,394 patients coded with hemorrhagic or ischemic stroke during the period between 2006 and 2015 were found, and among them, 30,271 patients had a first-time diagnosis of stroke (Fig. 1). To assure the validity of diagnosis, we excluded 15,776 cases without any hospitalization records and/or imaging evidence of either computed tomography (CT) or magnetic resonance imaging (MRI). Furthermore, 487 patients who did not have accurate age information or who were younger than 18 years old, and 14 patients without consistent diagnosis on the index date were excluded. A total of 13,994 patients were eligible for analysis out of the 1 million sample population cohort. Among the patients who were excluded due to a lack of hospitalization records (15,001), 5,566 patients had made at least three visits to an outpatient clinic.

Demographic and clinical characteristics at diagnosis for the included and excluded patients are summarised according to different subtypes of stroke in Table 1 and Supplementary Table S4. Approximately 73.5% of the patients had a cerebral infarction (CI), followed by 13.4% with intracerebral hemorrhage (ICH), and 8.4% with subarachnoid hemorrhage (SAH), which corroborates with a recent report<sup>20</sup>, indicating the validity of the included cohort. A post-hoc analysis showed that there were statistically significant differences among the age distributions of patients in the different stroke subtypes (Kruskal–Wallis test,  $P < 0.0001$ ), where patients with SAH were the youngest. About 10% of the patients in the cohort were beneficiaries of medical aid services in Korea, which is higher than the average 3.4% of beneficiaries in the total population reported between 2006 and 2015<sup>21</sup>. The most common comorbidity among the stroke patients was hypertension, followed by hyperlipidemia and diabetes. When compared with the included patients, the excluded patients generally showed lower prevalence rates of pre-existing comorbidities, which is consistent with our hypothesis that these are, at most, relatively mild cases of stroke<sup>22</sup>. Patients with ischemic stroke scored the lowest on the stroke severity index (SSI), indicating less severe clinical conditions during their first hospital stay due to stroke compared with those with hemorrhagic stroke (median 4.1 vs. 12.1, respectively).

**Lifetime risks of stroke.** The lifetime risks of hemorrhagic and ischemic strokes, calculated using the cumulative incidence rate (CIR), in males and females are summarised in Fig. 2 and Supplementary Table S5. In general, the lifetime risk of stroke for males was higher than females, except for SAH, during the last decade. The estimated lifetime risk of ischemic stroke ( $CIR_{18-84}$ ) seemed to consistently decrease over the 10-year observation period, from 19.5 to 15% and from 16.1 to 9.9%, for males and females, respectively. When adjusted with prevalence rates, the estimated results were slightly higher than the original estimates, due to the smaller size

	SAH N = 1,182	ICH N = 2,022	Other NIH N = 508	CI N = 10,282	Overall N = 13,994	Excluded <sup>a</sup> N = 5,566
<b>Male</b>	458 (38.8)	1,152 (57.0)	344 (67.7)	5,584 (54.3)	7,538 (53.9)	2,492 (44.8)
<b>Age, years (mean)</b>	56.3	61.6	64.9	67.7	65.7	66.1
18–44	229 (19.4)	221 (10.9)	53 (10.4)	431 (4.2)	934 (6.7)	244 (4.4)
45–54	343 (29.0)	446 (22.1)	75 (14.8)	1,255 (12.2)	2,119 (15.1)	668 (12.0)
55–64	285 (24.1)	474 (23.4)	81 (15.9)	2,028 (19.7)	2,868 (20.5)	1,281 (23.0)
65–74	190 (16.1)	453 (22.4)	150 (29.5)	3,151 (30.7)	3,944 (28.2)	1,976 (35.5)
75–84	111 (9.4)	362 (17.9)	121 (23.8)	2,863 (27.8)	3,457 (24.7)	1,257 (22.6)
85 +	24 (2.0)	66 (3.3)	28 (5.5)	554 (5.4)	672 (4.8)	140 (2.5)
<b>Residence area</b>						
Seoul metropolitan city	215 (18.2)	356 (17.6)	113 (22.2)	1,562 (15.2)	2,246 (16.1)	1,115 (20.0)
Other metropolitan cities	325 (27.5)	515 (25.5)	95 (18.7)	2,428 (23.6)	3,363 (24.0)	1,191 (21.4)
Non-metropolitan area	642 (54.3)	1,151 (56.9)	300 (59.1)	6,292 (61.2)	8,385 (59.9)	3,260 (58.6)
<b>Insurance type</b>						
Self-employed insured	441 (37.3)	777 (38.4)	166 (32.7)	3,407 (33.1)	4,791 (34.2)	1,767 (31.8)
Employed insured	666 (56.4)	1,073 (53.1)	284 (55.9)	5,736 (55.8)	7,759 (55.5)	3,114 (56.0)
Medical aid beneficiary	75 (6.4)	172 (8.5)	58 (11.4)	1,139 (11.1)	1,444 (10.3)	685 (12.3)
<b>Comorbidities</b>						
Atrial fibrillation	41 (3.5)	90 (4.5)	40 (7.9)	1,130 (11.0)	1,301 (9.3)	127 (2.3)
CHF	36 (3.1)	76 (3.8)	29 (5.7)	697 (6.8)	838 (6.0)	125 (2.3)
CKD	11 (0.9)	75 (3.7)	16 (3.2)	232 (2.3)	334 (2.4)	65 (1.2)
COPD	75 (6.4)	167 (8.3)	64 (12.6)	1,076 (10.5)	1,382 (9.9)	303 (5.4)
Diabetes	255 (21.6)	614 (30.4)	154 (30.3)	4,025 (39.2)	5,048 (36.1)	909 (16.3)
Hyperlipidemia	251 (21.2)	563 (27.8)	149 (29.3)	5,774 (56.2)	6,737 (48.1)	936 (16.8)
Hypertension	595 (50.3)	1,419 (70.2)	298 (58.7)	6,669 (64.9)	8,981 (64.2)	1509 (27.1)
IHD	142 (12.0)	279 (13.8)	92 (18.1)	1,986 (19.3)	2,499 (17.9)	563 (10.1)
TIA	167 (14.1)	61 (3.0)	33 (6.5)	679 (6.6)	940 (6.7)	286 (5.1)
<b>SSI (median)</b>	15.6	12.1	9.7	4.1	5.8	NA

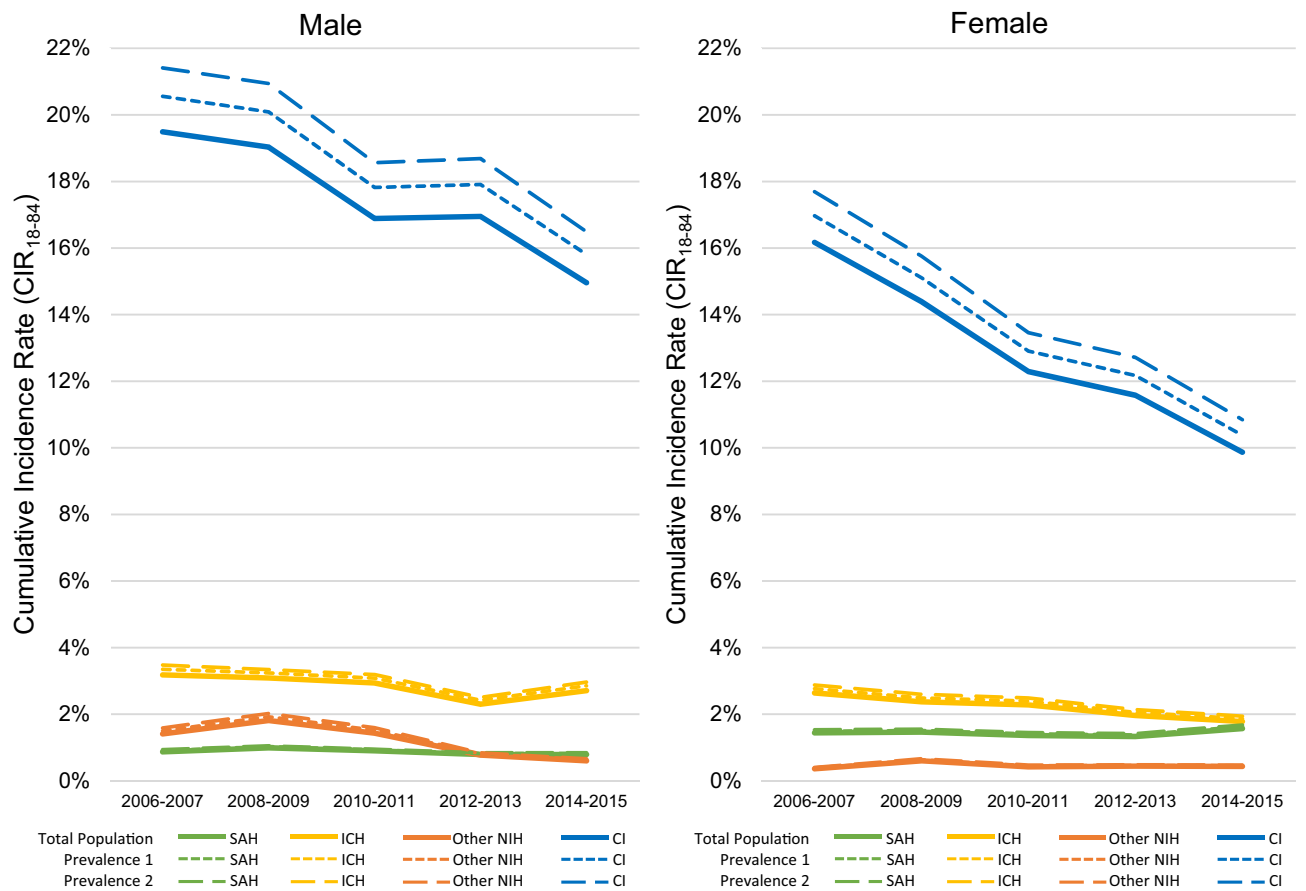
**Table 1.** Baseline demographic and clinical characteristics of stroke patients from the Korean NHIS-National Sample Cohort. Values are numbers (percentage) unless stated otherwise. *CHF* congestive heart failure, *CI* cerebral infarction, *CKD* chronic kidney disease, *COPD* chronic obstructive pulmonary disease, *ICH* intracerebral hemorrhage, *IHD* ischemic heart disease (including myocardial infarction), *NA* not applicable, *NHIS* National Health Insurance Service, *NIH* non-traumatic intracranial hemorrhage, *SAH* subarachnoid hemorrhage, *SSI* stroke severity index (ranging from 4.1 to 27.11; a higher score indicates more severe status), *TIA* transient ischemic attack. <sup>a</sup>This number is patients who have visited an outpatient clinic more than three times with a stroke diagnosis code.

of the denominator. The lifetime risk of hemorrhagic stroke showed a negligible but decreasing trend over the years, except for SAH in females.

**Validation of the extrapolation method.** Mean survival for the 10-year period was estimated using the Kaplan–Meier method and the semi-parametric extrapolation method for comparison, for which the results are summarised in Supplementary Table S6. The extrapolated estimates based on the first 5 years of follow up were close to those obtained using the Kaplan–Meier method, showing relative biases of less than 1.4% in all subtypes of stroke, thus supporting the validity of this method.

**Estimation of life expectancy (LE) and expected years of life lost (EYLL).** The average life span of the age-, sex-, and calendar year-matched reference population was similar among the different types of hemorrhagic and ischemic stroke (83.2–85.1 years), but the estimated LE after diagnosis and EYLL for the stroke cohort varied among the different subtypes (Table 2, Supplementary Fig. S1). Patients with SAH were estimated to have the highest LE (17.9 years) yet also the highest EYLL (9.7 years), which was followed by patients with ICH, with 7.8 years of EYLL.

**Lifetime healthcare expenditures.** The lifetime healthcare expenditures for hemorrhagic and ischemic strokes for the Korean National Health Insurance Service (NHIS) and the copayment expenses for patients are summarised in Table 2 and Supplementary Table S7. Based in the calculations, ICH resulted in the largest financial burden to the NHIS, with an estimated \$71,406, which was followed by SAH (\$58,741), and CI (\$50,551) for each incident stroke case when adjusted with a 3% annual discount rate. Copayment expenses



**Figure 2.** Lifetime probability of developing different subtypes of stroke.  $CIR_{18-84}$  of different subtypes of stroke from 2006 to 2015. The three line patterns represent different CIR estimations with slightly different denominators: the solid line indicates the number of total population from vital statistics in each strata; the dotted and broken lines indicate estimations after subtracting the number of prevalent cases of stroke from the denominator based on two different prevalence analyses from Korea (prevalence 1 and 2)<sup>4</sup>. *CI* cerebral infarction, *ICH* intracerebral hemorrhage, *NIH* non-traumatic intracranial hemorrhage, *SAH* subarachnoid hemorrhage.

Stroke type	N	Age at Dx	LE after Dx	EYLL	Expenditures covered by NHIS	Copayment expenses
SAH	1,182	56.31 (0.40)	17.90 (1.76)	9.66 (1.76)	58,741 (4,031)	12,290 (1,052)
ICH	2022	61.60 (0.31)	14.44 (0.96)	7.80 (1.00)	71,406 (4,328)	14,921 (867)
Other NIH	508	64.94 (0.65)	13.30 (2.05)	6.01 (2.19)	49,911 (5,526)	9,684 (1,037)
CI	10,282	67.65 (0.12)	12.75 (0.50)	4.68 (0.49)	50,551 (1,716)	11,666 (344)

**Table 2.** Life expectancy (LE), expected years of life lost (EYLL), and lifetime direct medical costs per case by different subtypes of stroke in Korea. US\$1 = 1,131.2 KRW in 2015. Values are mean (SE) in years or US\$. The medical costs are adjusted with 3% annual discount rates here. *CI* cerebral infarction, *Dx* diagnosis, *ICH* intracerebral hemorrhage, *N* number, *NHIS* National Health Insurance Service, *NIH* non-traumatic intracranial hemorrhage, *SAH* subarachnoid hemorrhage, *SE* standard error.

comprised approximately 16.2–17.3% of the total costs for hemorrhagic stroke, whereas the proportion was 18.8% for ischemic stroke.

### Discussion

The present study is the first to quantify both the likelihood (Fig. 2) and consequences (Table 2) of hemorrhagic and ischemic strokes in Korea, which will provide stakeholders with an opportunity to compare the potential impacts of prevention versus other healthcare services. We found that the lifetime risks of ischemic stroke ( $CIR_{18-84}$ ) have been decreasing for both genders over the last decade, while those for different subtypes of hemorrhagic stroke did not show such an obvious trend, especially among males with ICH and females with SAH (Fig. 2). Although the lifetime risk is much higher in ischemic stroke than in the different subtypes

of hemorrhagic strokes, the EYLLs and lifetime health expenditures are higher in the latter events. Namely, the EYLL (or, loss of life expectancy) for patients with hemorrhagic strokes would be 6–9.7 years with lifetime healthcare expenditures of \$59,595–\$86,327, while those of ischemic stroke would be 4.7 years and \$62,217 per patient (Table 2).

Besides making any more inferences, we must first justify the validity of these estimations. First, our results of lifetime risks or  $CIR_{18-84}$  corroborate those of another study of stroke in Korea<sup>3</sup>. Since the lifetime risk of a disease has an intuitive interpretation of personal risk under probability terms and can be directly compared across different time periods<sup>23,24</sup> and countries<sup>25–29</sup>, this information could quickly catch the attention of both lay people and policy makers. Second, since accurate quantifications of LE and lifetime healthcare expenditures are crucial for a fair comparison of the societal impacts between prevention and all other direct clinical services, we must also validate our extrapolation method used to determine lifetime survival. In addition to the theoretical justifications<sup>18,19</sup>, we performed empirical validations by using the first five years of data to extrapolate to the end of the 10th year and compared them with actual Kaplan–Meier’s estimates made for the 10 years of follow-up. Because the relative biases were all below 1.4% (Supplementary Table S6), this assured the accuracy of the estimated lifetime survival function based on the 10 years of follow-up from 2006 to 2015, which was much closer to the LE of usual strokes (Table 2). Third, we estimated EYLL as the indicator of potential savings from prevention of loss of LE. Because EYLL is the difference between the LE of the stroke cohort and that of the age-, sex-, and calendar year-matched referents, it quantifies the lifetime health benefits brought about by successful prevention of a case of stroke. In addition, because it has already been adjusted for differences in age and sex<sup>30,31</sup>, it inherently controls lead time bias when early diagnosis is applied as a strategy for prevention. Matching for the same calendar year would adjust for progressive improvements in medical technology over time. Therefore, these results are presented in units that make prevention directly comparable for value with all other direct clinical services such as diagnosis, treatment, rehabilitation, alternative medicines, etc., and/or other health conditions, and thus could aid health policy makers in their decision-making processes related to more equitable and efficient allocation of scarce resources.

Exploration of trends of lifetime risks of stroke subtypes raises issues of prevention strategies in Korea: the lifetime risk of ischemic stroke ( $CIR_{18-84}$ ) dropped to about 1 in 6 males and 1 in 10 females from 2014 to 2015. The lifetime risks for hemorrhagic stroke also showed a small yet decreasing trend (Fig. 2). While other types of stroke showed higher lifetime risk in men than in women, consistently over the last decade, this appeared to be reversed in the case of SAH (e.g., in the time period between 2014 and 2015, for males, 1 in 127 and for women, 1 in 64). Other studies have reported similar lifetime risk of stroke between men and women<sup>3,26</sup>, and the same trend of a much greater lifetime risk of SAH in women was observed in a study conducted in Japan<sup>26</sup>. Previous studies have reported that men have a higher stroke incidence than women in general, but in the case of SAH, the incidence is higher in women than in men<sup>32,33</sup>. The mechanism behind this phenomenon is still unclear, but some studies propose hormonal factors as a possible explanation<sup>33–35</sup>. The estimates did not differ to any great degree when the population at risk for calculation was adjusted by two different reported prevalence rates of stroke in Korea<sup>4</sup>. In these prevalence rates, new cases from the same year were also included, so the estimates from this sensitivity analysis would lead to a minor over-estimation, or the upper bound of the CIR.

Our estimation of lifetime expenditures is based on incident cases, and as such is more accurate compared with previous studies based on prevalent cases<sup>13,36–38</sup>. While such studies reported a comprehensive approach, including both direct and indirect costs, many components of the costs were calculated by simply multiplying specific pre-determined amounts by the number of stroke patients. On the other hand, the strength of the current study lies in the fact that the analysis is from real-world data that incorporates all logged expenses in the NHIS reimbursement dataset after the diagnosis of stroke. Another recent study has estimated the 5-year cumulative costs of 3,393 patients of ischemic stroke in Korea<sup>39</sup>. Although they also analyzed the NHIS dataset, unlike our national sample cohort, their customized dataset allowed their estimates to include costs of long term care. However, their 5-year cumulative costs were summed up throughout the 5 years by taking the average daily costs for those followed for 5 years and were not weighted for survival rates at each time point. In contrast, we applied Kaplan–Meier method to account for survival rate and censored patients at each time points, which would be the average lifetime medical costs for a stroke patient. Furthermore, the rolling-over extrapolation algorithm used in this study has been demonstrated to show accurate, robust results regardless of whether the index cohort’s excess hazard is constant, decreasing, or increasing after censoring observations<sup>18</sup>. Therefore, while previous studies have shown a rising burden of stroke year after year, the lifetime health expenditures of different subtypes of stroke estimated in the present study provide additional evidence on how much financial burden per case could be avoided for the Korean NHIS and individuals thanks to the prevention of stroke.

There are several limitations that must be acknowledged in this study. First, while the NHIS dataset has many advantages, such as being real-world data representative of Korea, large in size, with a 10-year follow up, and relatively free of recall bias, it involves input of data (e.g., diagnosis) from various levels of hospitals and clinics, so the validity of the diagnoses of different subtypes of stroke may thus be in question. However, this concern could be partially dealt with by having clear inclusion and exclusion criteria. In this study, we included stroke patients who were hospitalised with diagnosis codes I60–I63 (ICD-10) accompanied with imaging records such as CTs and MRIs, to enhance the validity of the diagnosis. While this step was crucial since the inclusion of only cases with a valid diagnosis was a prerequisite for the entire study, it may have resulted in excluding some mild stroke cases that did not require hospitalization. The potential impact of this inclusion criteria thus could have resulted in underestimation of the results. Second, the estimates were based on currently available data from 2006 to 2015, but the true outcomes for these patients in the future may reflect improvements in stroke care. This means that the lifetime survival results from our study could easily have been underestimated, while the EYLL could have been over-estimated. Therefore, the results should be interpreted carefully with consideration of the possible direction of such effects. Another limitation of this study is that we only analyzed the lifetime cost of

stroke in terms of direct medical costs that are covered by the NHIS. In Korea, however, NHIS reimburses over 80% of the care provided to patients with stroke in hospitals, meaning that our study did include most of the expenses. Therefore, the missing portion would only lead to an underestimation of the total direct medical costs<sup>13</sup>. Also, many studies on the cost of illness have laid out different types of cost, including direct non-medical costs as well as indirect costs such as productivity loss. Although we did not include other compositions of costs here, we could refer to previous reports to get an overall idea of the total costs. For example, a review on the indirect costs of stroke have reported the median proportion of indirect costs to total costs to be 32%<sup>40</sup>; thus, with our results for direct costs, one may roughly estimate the total lifetime costs. For a more in-depth understanding, further studies on different types of costs, such as indirect costs from productivity loss, are warranted.

Future studies may include further stratification of the cohort, such as different levels of stroke severity, different types of comorbidities, or different treatment modalities. Especially in Korea, where many people seek traditional Korean medicine for stroke care<sup>41</sup>, it would be worthwhile to establish a cohort of patients who utilised both conventional and traditional Korean medicine to see the outcome or consequences of integrated care. Also, other outcomes reflecting important values in healthcare, such as quality of life or years living with functional disability could be integrated with the results from the current study to provide further evidence on lifetime outcomes of stroke in Korea.

## Conclusion

For prevention technology to be fairly compared with other healthcare services in order to determine resource allocations, strong evidence that comprises both likelihood and consequences of events is necessary. While the lifetime incidence of stroke appears to have decreased in the last decade, except for SAH in females, the health and financial impact of stroke across the lifetime horizon is still significant, and this burden could be alleviated by successful prevention. The results of this study, together with future works on other important outcomes, such as quality of life and functional disability, may provide the best available real-world evidence for economic evaluation of outcomes of all healthcare services, including prevention of stroke in Korea.

## Methods

**Data source.** We used the Korean National Health Insurance Service-National Sample Cohort (NHIS-NSC) from January 2002 to December 2015 to construct our study cohort of adult patients suffering a first-time stroke in South Korea. Access to the database can be obtained by submitting an application to the Korean National Health Insurance Service (NHIS) with ethics approval from the researcher's institutional review board and a study proposal. The NHIS-NSC is a representative population-based cohort established in 2002. It includes various information, such as participants' insurance eligibility, diagnostic work-ups and treatments in both clinics and hospitals, and claimed reimbursement data<sup>42</sup>. More details on how the dataset was constructed and maintained can be found in other publications<sup>42,43</sup>. This study was approved by the institutional review board of Kyung Hee University (KHSIRB-18-026), and the board granted a waiver to conduct the study without written informed consent due to its retrospective study design using an anonymized dataset. All methods were performed in accordance with the relevant guidelines and regulations.

**Study population.** Adult patients aged  $\geq 18$  with primary diagnosis codes indicating either hemorrhagic or ischemic stroke from 2006 to 2015 were included in this study. We also searched through records between 2002 and 2005 to look for any prevalent cases or sequelae of stroke. Any individual with such a record was excluded to ensure that all included cases were first-time stroke. Following the diagnostic codes of the ICD-10 (International classification of diseases, 10th revision), codes I60–I62 were used to identify hemorrhagic stroke; I63 was used for ischemic stroke, and I60–I69 was applied to any type of stroke or sequelae of stroke. Because claim data could include exaggerated diagnoses of more severe diseases to avoid rejection, we adopted a common strategy of including only hospitalized cases with stroke to assure the validity of the diagnostic codes<sup>44,45</sup>. In the Korean dataset specifically, the Health Insurance Review and Assessment Service (HIRA) of Korea reported that diagnostic codes on actual hospital inpatient records and the claims dataset were matched 82% overall (94.9% for stroke), but the rate dropped to only 44.4% in outpatient records<sup>46</sup>. Based on this, we are concerned that more than half of those excluded may not have been actual stroke cases, or, the severity of stroke was so mild that they did not require hospitalization, such as a transient ischemic attack. Thus, to assure the validity of the stroke diagnosis in the cohort, patients without hospital admission records were excluded. Among those excluded by this criteria, we identified patients who visited an outpatient clinic at least three times for a comparison with included patients. Patients without diagnostic brain imaging records, i.e., computed tomography (CT) or magnetic resonance imaging (MRI), accurate age information, i.e., patients born before 1921, or without consistent diagnosis on the index date were also excluded.

From the established cohort, we extracted various information, including demographic data, area of residence, insurance type (e.g., self-employed, employed, or medical aid beneficiary), survival status, dates and payments for healthcare utilization, and diagnoses. As baseline clinical characteristics, information on nine comorbidities (atrial fibrillation, congestive heart failure, chronic kidney disease, chronic obstructive pulmonary disease, diabetes, hyperlipidemia, hypertension, ischemic heart disease and transient ischemic attack) that were present before the occurrence of stroke was collected. In addition, we calculated the stroke severity index (SSI)<sup>47</sup> using information on healthcare utilization during the patients' first hospital stay. SSI scores range from 4.1 to 27.11, where a higher score means a more severe status. The codes used to define imaging records, comorbidities and predictors of SSI are summarised in Supplementary Tables S1, S2 and S3.

**Lifetime risk of stroke estimated by CIR.** From the study population, we identified new age- and sex-specific stroke cases from 2006 to 2015. Then, the number of new cases for each year was divided by the sampling fraction of the corresponding years. This numerator was used to calculate the age- and sex-specific incidence rates for every two years, with the corresponding population at risk from census records<sup>48</sup> as the denominators. Afterwards, the cumulative incidence rate (CIR) for ages 18–84 ( $CIR_{18-84}$ ) was calculated using the formula below<sup>49</sup>:

$$CIR_{18-84} = 1 - \exp\left[-\sum_i (IR_i)(\omega_i)\right], \text{ where } i = 18 - 24, 25 - 29, 30 - 34, \dots, 80 - 84$$

In the formula, the incidence rate for the  $i$ th age group is represented by  $IR_i$ , and  $\omega_i$  is the width of the  $i$ th age group. While the CIR formula provides each year of age an equal weight during calculation, one may be concerned that the relatively large variation, resulting from small sample sizes of denominators and numerators for people living longer than life expectancy, would be over-represented in the final estimation. Therefore, we decided to calculate the CIR of stroke patients for ages 18–84 only.

As a sensitivity analysis, we also estimated the  $CIR_{18-84}$  after subtracting the number of prevalent cases from the denominators in each strata based on two different analysis results from Korea<sup>4</sup>.

**Survival analysis and extrapolation.** We applied a semi-parametric method that could be used to quantify the lifetime survival function of the index cohort using that of the age-, sex-, and calendar year-matched reference population. This survival function of the reference population was generated from the complete life tables of South Korea<sup>50</sup> using Monte-Carlo methods<sup>17,18,51</sup>. In brief, the lifetime survival function of the matched referents would have been the LE of the patients if the patients had not developed stroke. For the index cohort, we first used the Kaplan–Meier method to estimate survival from the onset of stroke to the end of follow-up. The survival of the index cohort was extrapolated beyond the end of follow-up to lifetime by establishing a logit transformation of the survival ratio between the index cohort and referents, followed by a rolling-over extrapolation algorithm<sup>18</sup>. Based on the estimated survival function, the LE and EYLL were estimated for four subtypes of stroke, including subarachnoid hemorrhage (SAH), intracerebral hemorrhage (ICH), other non-traumatic intracranial hemorrhage (NIH), and cerebral infarction (CI). In addition, we validated the extrapolation method by adopting the data from 2006 to 2010 and extrapolating to the end of 2015 and then compared them with the actual Kaplan–Meier’s estimate of each subcohort. The relative bias was the difference between the extrapolated estimate and the 10-year follow-up estimate divided by the 10-year follow-up estimate. To obtain the Kaplan–Meier’s estimate, we used Proc Lifetest restricted to the largest observation on the SAS Enterprise Guide (ver. 7.13). The rest of the analyses were conducted in R 3.4.1 using the iSQoL2 (ver. 4.1) package<sup>52</sup>.

**Lifetime healthcare expenditures.** Based on the healthcare service utilization fees claimed in our dataset, the lifetime healthcare expenditures covered by NHIS and copayment expenses for patients were calculated in this study. This included direct medical costs incurred for all types of hospitalization, clinical visits, and pharmacies registered in the Korean NHIS from the diagnosis of stroke until being censored or declared deceased. First, we calculated the observed monthly average cost for every case after diagnosis. The cost beyond follow-up was extrapolated using a rolling-over extrapolation algorithm, assuming that medical expenditures typically rise near the end of life. Then, the product of the monthly survival rates and monthly mean costs were summed up to estimate the lifetime cost<sup>18</sup>. Lastly, we adjusted the extrapolated costs with a 3% annual discount rate, and added 0% and 5% rates for the sensitivity analysis. More detailed methods and mathematical proofs have been described in previous reports<sup>17,18</sup>. This calculation was also performed using the iSQoL2 package. The original values coded with Korean Won (KRW) were all converted to US\$ based on the average exchange rate in 2015 (1 US\$ for 1,131.2 KRW)<sup>53</sup>.

## Data availability

The datasets generated and analyzed during the current study are not publicly available due to restrictions by the Korean NHIS. However, interested parties may submit a separate application to the NHIS for access. The NHIS accepts applications via their website (<https://nhiss.nhis.or.kr>), and ethics approval from the researcher’s institutional review board and a study proposal are required<sup>43</sup>. The iSQoL2 package for Windows is freely available at <http://sites.stat.sinica.edu.tw/isqol/> and provided with an English manual.

Received: 28 January 2020; Accepted: 13 August 2020

Published online: 03 September 2020

## References

1. Feigin, V. L., Norrving, B. & Mensah, G. A. Global burden of stroke. *Circ. Res.* **120**, 439–448 (2017).
2. Vos, T. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1211–1259 (2017).
3. Feigin, V. L. *et al.* Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016. *N. Engl. J. Med.* **379**, 2429–2437 (2018).
4. Kim, J. Y. *et al.* *Stroke Statistics in Korea 2018: A Report from the Epidemiology Research Council of the Korean Stroke Society.* (2019).
5. Feigin, V. L. *et al.* Update on the global burden of ischemic and hemorrhagic stroke in 1990–2013: The GBD 2013 study. *Neuroepidemiology* **45**, 161–176 (2015).
6. Kim, J. S. Stroke becomes the 3rd important cause of death in Korea; is it a time to toast?. *J. Stroke* **16**, 55 (2014).
7. NHIS & HIRA. *2017 National Health Insurance Statistical Yearbook* (Wonju, Korea, 2018).

8. FaniMarvasti, F. & Stafford, R. S. From sick care to health care—reengineering prevention into the US system. *N. Engl. J. Med.* **367**, 889–891 (2012).
9. Bloom, D. E. *et al.* Macroeconomic implications of population ageing and selected policy responses. *Lancet* **385**, 649–657 (2015).
10. Ovbigele, B. *et al.* Forecasting the future of stroke in the united states. *Stroke* **44**, 2361–2375 (2013).
11. Wang, F., Wang, J.-D. & Huang, Y.-X. Health expenditures spent for prevention, economic performance, and social welfare. *Health Econ. Rev.* **6**, 45 (2016).
12. Hong, K.-S. *et al.* Stroke statistics in Korea: Part I. Epidemiology and risk factors: A report from the Korean stroke society and clinical research center for stroke. *J. Stroke* **15**, 2 (2013).
13. Cha, Y.-J. The economic burden of stroke based on South Korea's National Health Insurance Claims Database. *Int. J. Health Policy Manag.* **7**, 904–909 (2018).
14. Kwon, S. *et al.* Health-related quality of life and related factors in stroke survivors: Data from Korea National Health and Nutrition Examination Survey (KNHANES) 2008 to 2014. *PLoS One* **13**, e0195713 (2018).
15. Hong, K.-S. *et al.* Burden of ischemic stroke in Korea: Analysis of disability-adjusted life years lost. *J. Clin. Neurol.* **7**, 77–84 (2011).
16. Hwang, J. S., Tsauo, J. & Wang, J. Estimation of expected quality adjusted survival by cross-sectional survey. *Stat. Med.* **15**, 93–102 (1996).
17. Hwang, J. S. & Wang, J. D. Monte Carlo estimation of extrapolation of quality-adjusted survival for follow-up studies. *Stat. Med.* **18**, 1627–1640 (1999).
18. Hwang, J.-S., Hu, T.-H., Lee, L.J.-H. & Wang, J.-D. Estimating lifetime medical costs from censored claims data. *Health Econ.* **26**, e332–e344 (2017).
19. Jackson, C. *et al.* Extrapolating survival from randomized trials using external data: A review of methods. *Med. Decis. Mak.* **37**, 377–390 (2017).
20. Kim, J. Y. *et al.* Executive summary of stroke statistics in Korea 2018: A report from the Epidemiology Research Council of the Korean Stroke Society. *J. Stroke* **21**, 42–59 (2019).
21. Number of Medical Aid Beneficiaries (1997–2017). <https://www.index.go.kr/main.do> (2018).
22. Schmidt, M., Jacobsen, J. B., Johnsen, S. P., Botker, H. E. & Sørensen, H. T. Eighteen-year trends in stroke mortality and the prognostic influence of comorbidity. *Neurology* **82**, 340–350 (2014).
23. Seshadri, S. & Wolf, P. A. Lifetime risk of stroke and dementia: Current concepts, and estimates from the Framingham Study. *Lancet Neurol.* **6**, 1106–1114 (2007).
24. Carandang, R. *et al.* Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. *JAMA* **296**, 2939–2946 (2006).
25. Seshadri, S. *et al.* The lifetime risk of stroke: Estimates from the Framingham Study. *Stroke* **37**, 345–350 (2006).
26. Turin Tanvir, C. *et al.* Lifetime Risk of Stroke in Japan. *Stroke* **41**, 1552–1554 (2010).
27. Berry, J. D. *et al.* Lifetime risks of cardiovascular disease. *N. Engl. J. Med.* **366**, 321–329 (2012).
28. Wang, Y. *et al.* Lifetime risk of stroke in young-aged and middle-aged Chinese population: The Chinese Multi-Provincial Cohort Study. *J. Hypertens.* **34**, 2434–2440 (2016).
29. Licher, S. *et al.* Lifetime risk and multimorbidity of non-communicable diseases and disease-free life expectancy in the general population: A population-based cohort study. *PLoS Med.* **16**, e1002741–e1002741 (2019).
30. Chu, P.-C., Wang, J.-D., Hwang, J.-S. & Chang, Y.-Y. Estimation of life expectancy and the expected years of life lost in patients with major cancers: Extrapolation of survival curves under high-censored rates. *Value Health* **11**, 1102–1109 (2008).
31. Yang, S.-C. *et al.* Cost-effectiveness of implementing computed tomography screening for lung cancer in Taiwan. *Lung Cancer* **108**, 183–191 (2017).
32. Barker-Collo, S. *et al.* Sex differences in stroke incidence, prevalence, mortality and disability-adjusted life years: Results from the global burden of disease study 2013. *Neuroepidemiology* **45**, 203–214 (2015).
33. Appelros, P., Stegmayr, B. & Terént, A. Sex differences in stroke epidemiology. *Stroke* **40**, 1082–1090 (2009).
34. de Rooij, N. K., Linn, F. H. H., van der Plas, J. A., Algra, A. & Rinkel, G. J. E. Incidence of subarachnoid haemorrhage: A systematic review with emphasis on region, age, gender and time trends. *J. Neurol. Neurosurg. Psychiatry* **78**, 1365–1372 (2007).
35. Turan, N. *et al.* Sex differences in the formation of intracranial aneurysms and incidence and outcome of subarachnoid hemorrhage: Review of experimental and human studies. *Transl. Stroke. Res.* **7**, 12–19 (2016).
36. Lim, S. J. *et al.* Socioeconomic costs of stroke in Korea: Estimated from the Korea national health insurance claims database. *J. Prev. Med. Public Health* **42**, 251–260 (2009).
37. Kim, H.-J. *et al.* The economic burden of stroke in 2010 in Korea. *J. Korean Med. Assoc.* **55**, 1226–1236 (2012).
38. Kang, H.-Y., Lim, S.-J., Suh, H. S. & Liew, D. Estimating the lifetime economic burden of stroke according to the age of onset in South Korea: A cost of illness study. *BMC Public Health* **11**, 646 (2011).
39. Kim, S.-E. *et al.* Three-month modified Rankin Scale as a determinant of 5-year cumulative costs after ischemic stroke: An analysis of 11,136 patients in Korea. *Neurology* **94**, e978–e991 (2020).
40. Joo, H., George, M. G., Fang, J. & Wang, G. A literature review of indirect costs associated with stroke. *J. Stroke Cerebrovasc. Dis.* **23**, 1753–1763 (2014).
41. Park, S.-U. *et al.* Integrative treatment modalities for stroke victims in Korea. *Complement. Ther. Clin. Pract.* **20**, 37–41 (2014).
42. Lee, J., Lee, J. S., Park, S.-H., Shin, S. A. & Kim, K. Cohort profile: The national health insurance service-national sample cohort (NHIS-NSC), South Korea. *Int. J. Epidemiol.* **46**, e15–e15 (2017).
43. Cheol Seong, S. *et al.* Data resource profile: The National Health Information Database of the National Health Insurance Service in South Korea. *Int. J. Epidemiol.* **46**, 799–800 (2017).
44. Lee, Y. S., Kwon, S., Chae, Y., Jang, B. H. & Ko, S. G. A retrospective cohort study on the outcomes of ischemic stroke patients with adjuvant Korean Medicine treatment. *Sci. Rep.* **8**, 1656 (2018).
45. Seo, S. R. *et al.* The incidence of stroke by socioeconomic status, age, sex, and stroke subtype: A nationwide study in Korea. *J. Prev. Med. Public Health* **47**, 104–112 (2014).
46. HIRA. *Assessment on Matching Rate of Disease Codes between NHIS Claims Data and Hospital Records and Proposal for Enhancement.* (2018).
47. Sung, S.-F. *et al.* Developing a stroke severity index based on administrative data was feasible using data mining techniques. *J. Clin. Epidemiol.* **68**, 1292–1300 (2015).
48. *Census Population by administrative district/sex/age.* <https://kosis.kr/> (2018).
49. Miettinen, O. Estimability and estimation in case-referent studies. *Am. J. Epidemiol.* **103**, 226–235 (1976).
50. *Complete Life Tables (1970–2016).* <https://kosis.kr/eng/> (2018).
51. Fang, C. T. *et al.* Life expectancy of patients with newly-diagnosed HIV infection in the era of highly active antiretroviral therapy. *QJM* **100**, 97–105 (2007).
52. *iSQoL2 package for windows.* <https://sites.stat.sinica.edu.tw/isqol/v4.1> (2018).
53. *Official exchange rate (LCU per US\$, period average).* <https://data.worldbank.org/indicator/PA.NUS.FCRF?locations=KR&view=chart> (2019).



### Author contributions

This article was conceived and designed by S.C. and J.D.W. S.C. and J.W. constructed the index cohort from the NHIS-NSC data. S.C. analyzed the data and wrote the manuscript. J.D.W. reviewed the data analysis. H.L., B.H.J., and J.D.W. reviewed the interpretations. All authors approved the final version of the manuscript for publication.

### Funding

This research was supported by the Ministry of Science and Technology of Taiwan (MOST 107-2627-M-006-007) and by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2014R1A1A2055507 and NRF-2017R1A2B4006407). This study used NHIS-NSC data (NHIS-2018-2-210) from the National Health Insurance Service (NHIS). The funders did not have any direct involvement in the design of the study, collection, analysis, and interpretation of data or in writing the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-71439-3>.

**Correspondence** and requests for materials should be addressed to J.-D.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020