

Original Article



Traditional and Genetic Risk Score and Stroke Risk Prediction in Korea

Keum Ji Jung , PhD¹, Semi Hwang , BA¹, Sunmi Lee , PhD²,
Hyeon Chang Kim , MD, PhD³, and Sun Ha Jee , PhD, MPH¹

¹Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea

²Health Insurance Policy Research Institute, National Health Insurance Service, Wonju, Korea

³Department of Preventive Medicine and Public Health, Yonsei University College of Medicine, Seoul, Korea



Received: Jan 30, 2018

Revised: Mar 14, 2018

Accepted: Mar 28, 2018

Correspondence to

Sun Ha Jee, PhD, MPH

Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, 50-1, Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.
E-mail: jsunha@yuhs.ac

Copyright © 2018. The Korean Society of Cardiology

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Keum Ji Jung
<https://orcid.org/0000-0003-4993-0666>
Semi Hwang
<https://orcid.org/0000-0002-7870-9963>
Sunmi Lee
<https://orcid.org/0000-0003-3051-2798>
Hyeon Chang Kim
<https://orcid.org/0000-0002-8210-756X>
Sun Ha Jee
<https://orcid.org/0000-0001-9519-3068>

Funding

This study was funded by a grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (HI13C0715, HI14C2686).

ABSTRACT

Background and Objectives: Whether using both traditional risk factors and genetic variants for stroke as opposed to using either of the 2 alone improves the prediction of stroke risk remains unclear. The purpose of this study was to compare the predictability of stroke risk between models using traditional risk score (TRS) and genetic risk score (GRS).

Methods: We used a case-cohort study from the Korean Cancer Prevention Study-II (KCPS-II) Biobank (n=156,701). We genotyped 72 single nucleotide polymorphisms (SNPs) identified in genome-wide association study (GWAS) on the KCPS-II sub-cohort members and stroke cases. We calculated GRS by summing the number of risk alleles. Prediction models with or without GRS were evaluated in terms of the area under the receiver operating characteristic curve (AUROC).

Results: Sixteen out of 72 SNPs identified in GWAS showed significant associations with stroke, with an odds ratio greater than 2.0. For participants aged <40 years, AUROCs for incident stroke were 0.58, 0.65, and 0.67 in models using modifiable TRS only, GRS only, and TRS plus GRS, respectively, showing that GRS only model had better prediction than TRS only. For participants aged ≥40 years, however, TRS only model had better prediction than GRS only model. Favorable levels of traditional risk were associated with significantly lower stroke risks within each genetic risk category.

Conclusions: TRS and GRS were both independently associated with stroke risk. Using genetic variants in addition to traditional risk factors may be the most accurate way of predicting stroke risk, particularly in relatively younger individuals.

Keywords: Risk factors; Epidemiologic methods; Genetics; Stroke

INTRODUCTION

Stroke is the third leading cause of death and disability in Korea as well as worldwide.¹⁾ Approximately 9% of total deaths are attributable to stroke-related causes, and the crude death rate of Koreans due to stroke was 48.2 per 100,000 persons in 2014. Moreover, the annual medical cost for stroke accounts for more than 10% of the total medical expenditures in Korea.^{1,2)}

Estimating absolute risk of stroke typically relies on prediction models developed using data from prospective cohort studies. Numerous previous studies have identified a

Conflict of Interest

The authors have no financial conflicts of interest.

Author Contributions

Conceptualization: Jung KJ, Jee SH; Data curation: Jung KJ, Lee S, Jee SH; Formal analysis: Jung KJ, Hwang S, Lee S, Jee SH; Funding acquisition: Jee SH; Investigation: Jung KJ, Jee SH; Methodology: Jung KJ, Hwang S, Lee S, Kim HC, Jee SH; Supervision: Kim HC, Jee SH; Validation: Jung KJ, Jee SH; Writing - original draft: Jung KJ, Kim HC, Jee SH; Writing - review & editing: Kim HC.

variety of traditional risk factors for stroke using standard ways of establishing prediction models including multivariate functions.³⁻⁵⁾ More recent studies have evaluated whether incorporating nontraditional risk factors (e.g., genetic variants) into models with traditional risk factors improves the accuracy of predicting stroke risk, but, it has shown only limited improvement in the overall prediction accuracy.⁶⁻⁸⁾

The rapid growth in genome-wide association study (GWAS) has resulted in the discovery of common genetic variants associated with stroke. This has enabled researchers to develop stroke risk prediction models using genetic variants identified in GWAS of stroke,⁹⁾¹⁰⁾ but the predictive validity of the models appears inconclusive. It may be due, in large part, to the possibility that genetic variants identified in GWAS of stroke may reflect the effect of modifiable traditional risk factors (e.g., hypertension, diabetes, dyslipidemia, and smoking), as well as the direct genetic effects on stroke.¹¹⁾ Furthermore, most of the identified genetic variants not only have small effects on stroke risk, but also are associated with modifiable traditional risk factors for stroke.¹²⁾ An alternative approach to improve predictability is to use genetic variants that are known to have large effect sizes and are unrelated to established risk factors for stroke.¹³⁾

Another issue with standard prediction modeling approaches is measurement errors associated with the quantification of traditional risk factors. If measurement errors for traditional risk factors exist, the predictability of prediction models using such factors tends to be reduced. A genetic risk score (GRS) is an approach that can potentially overcome the issue with large measurement errors.¹⁴⁾¹⁵⁾ Nevertheless, there is few research using these approaches altogether in developing and evaluating prediction models for stroke risk.

The aim of this study was, therefore, to develop prediction models for stroke risk using genetic variants that are known to have relatively large effect sizes (e.g., odds ratios [ORs] greater than 2.0) and unrelated to established risk factors for stroke. We also compared the predictability of stroke risk between models using traditional risk factors, GRS, and both traditional factors and GRS.

METHODS

Study subjects

The Korean Cancer Prevention Study-II (KCPS-II) Biobank was initiated in April 2004. It was supported by the Seoul City Government in December 2005 as a project of the Korean Metabolic Syndrome Research Initiatives study. A full description of KCPS-II has been previously published.¹⁵⁾¹⁶⁾ We have acquired written consent forms and blood samples from 156,701 participants. Among the total 156,701 participants, 1,047 participants who reported having a prevalent stroke were excluded. In addition, 12,107 participants who had missing values on systolic blood pressure, fasting blood glucose, total cholesterol, smoking status, and exercise were excluded. During the follow-up through December 2011, identified 823 (out of 144,594 participants) were recorded as incident cases of stroke. These data were used to develop both GWAS and prediction models.

For the case-cohort study, we selected a sub-cohort as a 3.4% random sample of all participants. Twenty-seven out of 4,856 randomly-selected participants were found to be diagnosed with stroke from our sub-cohort study, while 796 stroke cases were verified outside the sub-cohort.

In short, we included a total of 5,656 additional participants (4,833 plus 823 participants) in our case-cohort study design. The Institutional Review Board of Yonsei University reviewed and approved this study (approval No. 4-2011-0277). All participants have submitted an informed consent form at baseline survey.

Traditional risk factors

To develop prediction models using traditional risk factors, we first fitted Cox proportional hazards regression models to a basic set of classical risk factors: age, sex, hypertension, diabetes, dyslipidemia, and smoking status.

DNA extraction and single nucleotide polymorphism (SNP) analysis

The blood samples for DNA extraction used in the current study were obtained from the KCPS-II Biobank at baseline. We conducted genotyping in 2016, using a randomly selected sub-cohort of KCPS-II Biobank. Genotype data were produced using the Korean Chip (K-CHIP) obtained from the K-CHIP consortium. K-CHIP was designed by Center for Genome Science, Korea National Institute of Health, Korea (4845-301, 3000-3031). Seventy-four SNP were identified in GWAS of stroke in this study (**Supplementary Table 1**). Among the seventy-four SNPs, there were sixteen final SNPs, and cluster QC was successful (**Supplementary Figure 2**).

Outcome variables

Stroke events, comprising the occurrence of death due to stroke or the first occurrence of nonfatal stroke, were recorded.³⁾ We identified the events from insurance claims reported to the National Health Insurance System (NHIS). Since the NHIS is a national organization, follow-up was expected to be 100% completed. We ascertained nonfatal or stroke events, defined according to the International Classification of Diseases 10th revision (ICD-10, I60–I69) revision, from health insurance claims data from the NHIS. Cases of fatal stroke (ICD-10, I60–I69) were ascertained from the causes of death listed on death certificates.

We conducted a validation study in collaboration with the NHIS (formally Korean Medical Insurance Corporation). For participants who provided written permission for the use of their personal informations, 626 stroke events between 1993 and 1997 were confirmed with individual hospital medical records, and 83% of the stroke cases were validated.¹⁷⁾

Statistical analysis

To evaluate the general characteristics of the study participants, we calculated means and standard deviations (SDs), and presented proportions of hypertension, diabetes, dyslipidemia, and cigarette smoking status. We calculated GRS by summing the number of risk alleles over all SNPs. We then categorized the GRS into tertiles. Stroke risk associated with any one genotype was estimated as an OR and 95% confidence interval (CI), and was computed using logistic regression under an additive genetic model. We also used the area under the receiver operating characteristic curve (AUROC) analysis. Finally, Cox proportional hazards models were used to estimate the effect of GRS and traditional risk score (TRS) on stroke risk.

RESULTS

Table 1 describes the characteristics of the sub-cohort controls and stroke cases at their first health examination at baseline. Stroke cases were older and had higher systolic blood pressure, fasting glucose, and total cholesterol levels than the sub-cohort controls. Also, stroke cases

Table 1. General characteristics of study participants, in case-cohort design

Variables	Sub-cohort	Total stroke	p value
Participants number	5,263	823 (625)*	
Age (years)	43.1 (9.3)	52.4 (11.5)	<0.001
Systolic blood pressure (mmHg)	118.3 (14.2)	125.4 (15.9)	<0.001
Fasting blood glucose (mg/dL)	90.7 (18.4)	98.2 (26.1)	<0.001
Total cholesterol (mg/dL)	189.9 (33.3)	197.3 (35.3)	<0.010
LDL cholesterol (mg/dL)	113.8 (31.3)	117.6 (33.1)	<0.001
HDL cholesterol (mg/dL)	50.9 (10.0)	50.1 (11.5)	0.068
Triglyceride (mg/dL)	138.9 (89.8)	157.9 (98.8)	<0.001
Smoking status			0.761
Never	48.6	48.0	
Former	20.1	21.2	
Current	31.3	30.8	
Sex (female)	33.4	32.3	0.672
Hypertension	17.9	42.7	<0.001
Diabetes	4.9	14.8	<0.001
High cholesterol	34.2	44.3	<0.001

Data shown are mean (SD) or percentage (%).

HDL = high-density lipoprotein; LDL = low-density lipoprotein; SD = standard deviation.

*The number in parenthesis is incident.

Table 2. Allelic ORs for stroke in study participants

SNPs*	Chromosome	Gene	RA	RAF	HR (95% CI)
(24) rs17002646	22	<i>ME11</i>	C	1.0	2.2 (1.5–3.3)
(30) rs56680016	3	<i>CADPS</i> [†]	C	1.4	2.5 (1.8–3.6)
(34) rs75053900	10	<i>FAM208B</i>	C	98.0	2.6 (1.8–3.8)
(35) rs77412933	16	<i>LOC107984901</i>	A	1.4	2.5 (1.8–3.6)
(37) rs79159085	7	<i>LOC221946</i>	C	1.8	2.4 (1.7–3.3)
(38) rs79789141	1	-	C	98.0	2.6 (1.8–3.8)
(44) rs117868687	8	-	G	1.6	2.0 (1.2–2.5)
(48) rs139580491	15	-	C	97.0	2.1 (1.6–2.8)
(52) rs141824980	15	<i>GNB5</i> [†]	C	1.3	2.9 (2.0–4.1)
(55) rs143387922	9	-	C	98.0	2.1 (1.5–3.0)
(57) rs144579871	3	<i>LOC105374257</i>	A	98.0	2.0 (1.3–3.0)
(63) rs149912751	2	-	A	1.1	2.2 (1.5–3.4)
(67) rs191810437	4	<i>GPAT3</i> [‡]	A	1.2	3.1 (2.1–4.4)
(70) rs141886475	17	-	C	97.0	2.6 (1.3–5.2)
(71) rs184999606	13	<i>RCBTB2</i>	G	98.0	2.9 (1.4–6.0)
(72) rs188932107	12	<i>SOX</i> [‡]	G	98.0	2.8 (1.4–5.6)

CI = confidence interval; HR = hazard ratio; OR = odds ratio; RA = risk allele; RAF = risk allele frequency; SNP = single nucleotide polymorphism.

*Parentheses represent the SNP number in **Supplementary Table 1**; [†]These 2 genes are related to brain function; [‡]These 2 genes are related to dyslipidemia.

were more likely to have hypertension, diabetes, and dyslipidemia at baseline. Smoking status and exercise levels did not differ between the sub-cohort controls and stroke cases.

Table 2 shows 16 SNPs among 72 SNPs identified from GWAS with an OR of 2 or greater. Among the 16 SNPs used in the GRS calculation, 4 SNPs were associated with the brain function (rs56680016 in *CADPS* gene, rs141824980 in *GNB5* gene) and with dyslipidemia (rs191810437 in *GPAT3* gene, rs188932107 in *SOX5* gene). Four SNPs (rs12115631, rs142764455, rs147074671, and rs147154581) out of the 72 SNPs had stroke-related risk genes (*PTPRD*, *DLG2*, *TPK1*, and *GCLC*) (data not shown).

Table 3 shows prediction models for specific types of stroke using traditional risk scoring and count or weighted genetic-risk scoring. The AUROC value was highest in ischemic stroke (0.83) than hemorrhagic stroke (0.68). However, there were no differences between model 1 with count GRS, and model 2 with weighed GRS.

Table 3. Stroke type specific prediction models using TRS and count or weighted GRS, multiple logistic model

Variables	Total stroke (823/5,266)		Ischemic stroke (356/5,730)		Hemorrhagic stroke (216/5,870)	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Age (years)	0.08 (<0.01)	0.08 (<0.01)	0.09 (<0.01)	0.09 (<0.01)	0.02 (<0.01)	0.02 (<0.01)
Sex	0.14 (0.23)	0.13 (0.11)	-0.01 (0.97)	-0.01 (0.93)	-0.27 (0.18)	-0.28 (0.17)
Hypertension	0.69 (<0.01)	0.69 (<0.01)	0.71 (<0.01)	0.71 (<0.01)	0.50 (<0.01)	0.50 (<0.01)
Diabetes	0.46 (<0.01)	0.46 (<0.01)	0.67 (<0.01)	0.67 (<0.01)	0.25 (0.28)	0.25 (0.28)
Dyslipidemia	0.19 (0.02)	0.19 (0.02)	0.15 (0.20)	0.15 (0.21)	0.15 (0.29)	0.15 (0.31)
Ex-smokers	-0.01 (0.91)	-0.01 (0.93)	0.22 (0.21)	0.22 (0.22)	0.01 (0.94)	0.01 (0.95)
Current smokers	0.28 (0.02)	0.28 (0.02)	0.46 (0.01)	0.45 (0.01)	0.27 (0.15)	0.27 (0.16)
Count GRS	0.80 (<0.01)		0.74 (<0.01)		0.50 (<0.01)	
Weighted GRS		1.04 (<0.01)		0.98 (<0.01)		0.65 (<0.01)
AIC	4,016.79	4,015.44	2,174.24	2,172.52	2,172.52	1,804.18
AUROC	0.79	0.79	0.83	0.83	0.68	0.68

Data shown are β (p).

AIC = akaike information criterion; AUROC = area under the receiver operating characteristic curve; GRS = genetic risk score; TRS = traditional risk score.

Table 4. Total stroke prediction models, case-cohort design using Cox proportional hazard model

Variables	Age <40 years at baseline (104 cases/2,177 controls)			Age ≥40 years at baseline (521 cases/2,656 controls)			Total (625 cases/4,833 controls)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
TRS									
T1	1.0		1.0	1.0		1.0	1.0		1.0
T2	2.0 (1.3–3.0)		1.9 (1.3–2.9)	1.4 (0.8–2.8)		1.5 (0.8–2.9)	2.0 (1.5–2.7)		2.0 (1.5–2.7)
T3	6.6 (3.0–15)		8.0 (3.6–18)	6.0 (3.3–11)		6.4 (3.5–12)	8.4 (6.5–11)		8.5 (6.6–11)
GRS									
G1		1.0	1.0		1.0	1.0		1.0	1.0
G2		2.6 (0.7–12)	2.8 (0.7–11)		1.9 (1.2–3.2)	2.1 (1.3–3.5)		2.0 (1.3–3.4)	2.2 (1.4–3.5)
G3		9.6 (2.3–39)	9.5 (2.3–39)		4.9 (2.9–8.1)	5.4 (3.3–9.0)		5.5 (3.4–8.9)	5.9 (3.7–9.5)
AIC	22,121.3	22,103.8	22,085.2	111,613.7	111,713.9	111,487.3	134,304.8	134,601.8	134,132.9
DF	2	2	4	2	2	4	2	2	4
AUROC	0.58	0.65	0.67	0.72	0.62	0.76	0.75	0.63	0.78
Δ	Reference	0.07	0.09	Reference	-0.10	0.04	Reference	-0.12	0.03

TRS included age, sex, hypertension, diabetes, dyslipidemia, and smoking status; TRS was classified into tertile (T1, T2, T3). GRS included 16 SNPs using weighted method; GRS was classified into tertile (G1, G2, G3). AUROC was estimated using logistic model. Data shown are HR (95% CI).

AIC = akaike information criterion; AUROC = area under the receiver operating characteristic curve; CI = confidence interval; DF = degree of freedom; GRS = genetic risk score; HR = hazard ratio; SNP = single nucleotide polymorphism; TRS = traditional risk score.

Table 4 shows whether modifiable risk factors and GRS improve the prediction of stroke risk beyond age and sex, as determined by AUROC. We evaluated data according to different age groups. For participants aged <40 year at baseline, AUROCs (95% CI) for modifiable traditional risk factors (model 1), GRS only (model 2), and model 1 plus GRS (model 3) were 0.58 (model 1), 0.65 (model 2), and 0.67 (model 3) for incident stroke, respectively. Compared to model 1, the AUROC improved by 7% in model 2, and 9% in model 3. However, the predictability decreased in model 2 among participants aged 40 years and older.

Figure 1 shows AUROCs for incident stroke according to the different age groups. The gene-environment risk score model showed significant improvements in predictability by about 9% compared with the TRS model for the young-stroke risk category, while negligible differences were observed for old-stroke and even all stroke models. Favorable traditional risk was significantly associated with lower stroke risk within each genetic risk category (**Figure 2**).

Supplementary Figure 1 and **Supplementary Table 2** showed age-specific comparisons of HRs for GRS-stroke associations between logistic and Cox proportional hazard models. The association was highest among participants aged 50 to 59 years, but lowest among participants aged 60 years and older.

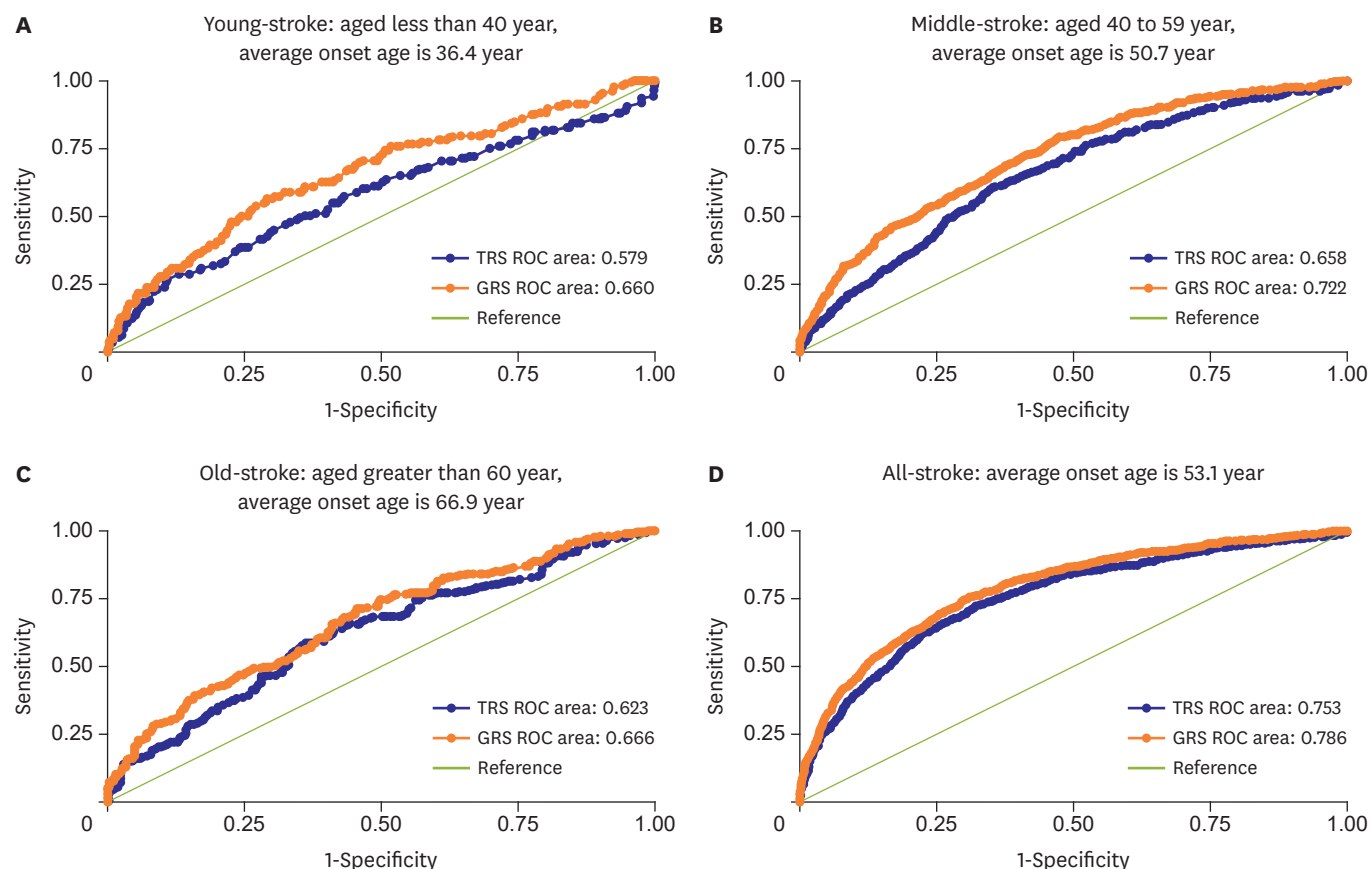


Figure 1. AUROC for incident stroke according to age groups. AUROC = area under the receiver operating characteristic curve; ROC = receiver operating characteristic.

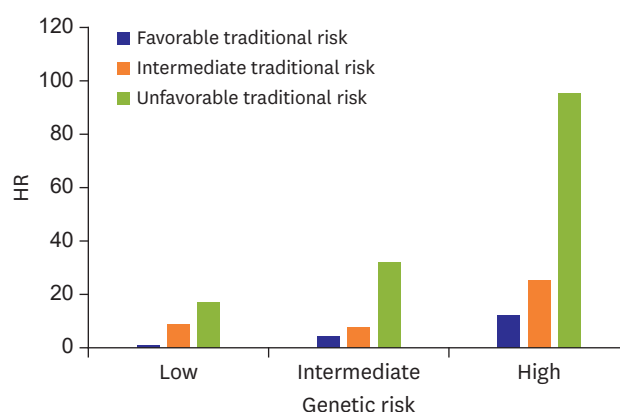


Figure 2. A favorable traditional risk for stroke according to genetic risk category.

DISCUSSION

This study aimed to determine whether the putative stroke risk alleles identified in GWAS exhibit higher predictive validity than traditional modifiable risk factors for stroke in Koreans. We found 16 of the 72 SNPs associated with stroke risk to have relatively big effect sizes (with ORs greater than 2.0), and be unrelated to traditional modifiable risk factors for stroke.

During the 633,210 person-year follow-up period, 823 incident stroke cases developed. This case-cohort study evaluated the ability to predict risk based on TRS only, and TRS plus GRS, which aggregates information from 16 genetic variants shown to be associated with risk of stroke in Koreans. Most previous genetic epidemiology studies have focused on total stroke events (ischemic or hemorrhagic), but the present study evaluating ischemic and hemorrhagic stroke separately demonstrated relatively smaller improvements in predictability for hemorrhagic stroke compared with ischemic stroke.

Although a genetic contribution to stroke risk is well-known, it has been unclear if the genetic contribution is larger for young-onset stroke than for old-onset stroke, or vice versa.¹⁸⁾ Moreover, it has been debatable whether a prediction model with genetic variants has higher predictability for young-onset stroke than for old-onset stroke. In this sense, this study provides important evidence that GRS for the putative stroke risk alleles identified from GWAS in Korean populations is predictive of incident stroke in conjunction with age, sex, and modifiable traditional risk factors. In particular, the contribution of GRS for the putative stroke risk alleles was much higher among younger participants aged less than 40 years old. Based on the recent evidence from European ancestry, heritabilities of young-onset stroke and old-onset stroke were estimated to be 42% ($\pm 8\%$, $p < 0.001$) and 34% ($\pm 10\%$, $p < 0.001$), respectively.¹⁸⁾¹⁹⁾ These results indicate that genetic contributions to stroke vary by age. Performing large, well-powered GWAS using various ranges of age is needed to provide more definitive evidence on the associations.

The early onset of stroke is more likely to reflect relatively higher genetic effects.¹⁸⁾ In our study, the AUROC for GRS was highest in participants with the early onset of stroke among the 3 groups of age (less than 40 years old, ages 40 to 59, and 60 years old and above). The mean ages of the 3 groups according to the age at the onset of stroke were 36.4, 50.7, and 66.9 years, respectively (**Figure 1**). This suggests that the prediction models with traditional risk factors had limitations in predicting the early onset of stroke. Individuals with the early onset of stroke may be more accurately diagnosed using our models incorporating genetic variants.

Genetic variants identified in GWAS of stroke may reflect direct genetic effects on stroke, as well as the effect of modifiable risk factors, such as hypertension, diabetes, dyslipidemia and smoking. As we run larger GWAS, some of the signals that emerge may turn out to reflect the action of modifiable (e.g., environmental and/or behavioral) factors, rather than the more direct biological effects.¹¹⁾ If this is true, most genetic variants from GWAS would also reflect the action of the modifiable risk factors mentioned above. In such a case, these genetic variants would not help to improve the accuracy of prediction models because these modifiable risk factors would already be included in the prediction models as predictors. In other words, modifiable risk factors may act as biological intermediates between genetic variants and stroke in this study.

Another issue is that genetic variants identified in GWAS have only moderate effects and small contributions to the variance of outcome variables. As such, we found 74 SNPs to be associated with stroke in our initial GWAS of stroke. However, only 16 of the 74 SNPs had an OR greater than 2.0.

In the present study, among 16 SNPs used in the GRS calculation, 2 of them were associated with the brain (rs56680016 in *CADPS* gene, rs141824980 in *GNB5* gene) and 2 other with dyslipidemia (rs191810437 in *GPAT3* gene, rs188932107 in *SOX5* gene). These genes were not

discovered previously in METASTROKE as reported in 2012.²⁰⁾ In addition, we evaluated 41 SNPs reported by Traylor et al.²⁰⁾ in our GWAS, and found that the K-CHIP we used had 26 SNPs, with 4 SNPs of them showing statistical significance with stroke ($p < 0.05$).²⁰⁾

In general, SNPs with relatively large effects are rare variants, while those with small effects are likely to be common variants. Eight of the 16 SNPs included in our stroke prediction models were rare variants. Therefore, these 8 SNP minor allele frequencies were very rare and should be combined with a heterozygous type. This approach might mislead the true mode of inheritance from codominant or additive mode to a recessive or a dominant mode. Therefore, our results need to be interpreted with caution.

In terms of measurement errors, the TRS estimated by traditional risk factors may have more measurement errors than the GRS. Repeated measures and biological variations of some traditional risk factors, such as blood pressure and fasting glucose, may help to reduce the measurement errors. In this study, in particular with younger age group of less than 40 years old, the prediction model with GRS had a 9% higher predictability than the model with TRS.

A major limitation is the use of self-reported exposures collected at a single point in time, which precluded the definitive exclusion of potential misclassification. The statistical power of the current study is modest, as genotyping was performed on a limited sample size of stroke cases and controls. However, a notable strength of our study is the case-cohort design drawn from an underlying large prospective cohort. Furthermore, case identifications were performed by record linkage to the NHIS with verification, increasing the reliability of our information. Genetic studies tend to be confounded by linkage disequilibrium, so confounding by population stratification cannot be completely ruled out. To overcome this limitation, we used a multidimensional-scaling method for population stratification analysis,²¹⁾ which did not separate out the cluster of our data using multidimensional-scaling. In addition, the majority of the Korean populations are ethnically homogenous, minimizing a negative effect.²²⁾

In conclusion, models using both traditional and genetic risk factors showed improvements in stroke risk prediction compared with models using either of the 2 alone. Therefore, both traditional and genetic risk factors for stroke should be considered when predicting stroke risk, in particular in younger individuals for the prediction of the early onset of stroke.

ACKNOWLEDGEMENTS

The authors thank the staff of the Korean National Insurance Service.

SUPPLEMENTARY MATERIALS

Supplementary Table 1

SNP list identified from GWAS, KCPS-II

[Click here to view](#)

Supplementary Table 2

Coefficient of total stroke prediction models according to different age groups

[Click here to view](#)

Supplementary Figure 1

Incident probability of stroke using logistic or Cox proportional hazard model.

[Click here to view](#)

Supplementary Figure 2

Cluster QC of 15 SNPs.

[Click here to view](#)

REFERENCES

1. Statistics Korea. *Annual Report on the Causes of Death Statistics 2012*. Daejeon: Statistics Korea; 2014.
2. Jee SH, Jung KJ, Jeon C, Kimm H, Yun YD, Kim IS. Smoking attributable risk and medical care cost in 2012 in Korea. *J Health Inform Stat* 2014;39:25-41.
3. Jee SH, Park JW, Lee SY, et al. Stroke risk prediction model: a risk profile from the Korean study. *Atherosclerosis* 2008;197:318-25.
[PUBMED](#) | [CROSSREF](#)
4. Williams FM, Carter AM, Hysi PG, et al. Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann Neurol* 2013;73:16-31.
[PUBMED](#) | [CROSSREF](#)
5. Zhuo Y, Yu H, Yang Z, Zee B, Lee J, Kuang L. Prediction factors of recurrent stroke among Chinese adults using retinal vasculature characteristics. *J Stroke Cerebrovasc Dis* 2017;26:679-85.
[PUBMED](#) | [CROSSREF](#)
6. Damen JA, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;353:i2416.
[PUBMED](#) | [CROSSREF](#)
7. Morrison AC, Bare LA, Chambless LE, et al. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol* 2007;166:28-35.
[PUBMED](#) | [CROSSREF](#)
8. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: advances and prospects. *Hum Mol Genet* 2008;17:R166-73.
[PUBMED](#) | [CROSSREF](#)
9. Hachiya T, Kamatani Y, Takahashi A, et al. Genetic predisposition to ischemic stroke: a polygenic risk score. *Stroke* 2017;48:253-8.
[PUBMED](#) | [CROSSREF](#)
10. Ibrahim-Verbaas CA, Fornage M, Bis JC, et al. Predicting stroke through genetic risk functions: the CHARGE Risk Score Project. *Stroke* 2014;45:403-12.
[PUBMED](#) | [CROSSREF](#)
11. Gage SH, Davey Smith G, Ware JJ, Flint J, Munafò MR. G = E: what GWAS can tell us about the environment. *PLoS Genet* 2016;12:e1005765.
[PUBMED](#) | [CROSSREF](#)
12. Achterberg S, Kappelle LJ, de Bakker PI, Traylor M, Algra SMART Study Group and the METASTROKE Consortium. No additional prognostic value of genetic information in the prediction of vascular events after cerebral ischemia of arterial origin: the PROMISE Study. *PLoS One* 2015;10:e0119203.
[PUBMED](#) | [CROSSREF](#)
13. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747-53.
[PUBMED](#) | [CROSSREF](#)

14. Kofler T, Thériault S, Bossard M, et al. Relationships of measured and genetically determined height with the cardiac conduction system in healthy adults. *Circ Arrhythm Electrophysiol* 2017;10:e004735.
[PUBMED](#) | [CROSSREF](#)
15. Lee SJ, Jee YH, Jung KJ, Hong S, Shin ES, Jee SH. Bilirubin and stroke risk using a mendelian randomization design. *Stroke* 2017;48:1154-60.
[PUBMED](#) | [CROSSREF](#)
16. Jee YH, Lee SJ, Jung KJ, Jee SH. Alcohol intake and serum glucose levels from the perspective of a mendelian randomization design: the KCPS-II Biobank. *PLoS One* 2016;11:e0162930.
[PUBMED](#) | [CROSSREF](#)
17. Park JK, Kim KS, Kim CB, et al. The accuracy of ICD codes for cerebrovascular diseases in medical insurance claims. *Korean J Prev Med* 2000;33:76-82.
18. Bluhner A, Devan WJ, Holliday EG, et al. Heritability of young- and old-onset ischaemic stroke. *Eur J Neurol* 2015;22:1488-91.
[PUBMED](#) | [CROSSREF](#)
19. Traylor M, Rutten-Jacobs LC, Holliday EG, et al. Differences in common genetic predisposition to ischemic stroke by age and sex. *Stroke* 2015;46:3042-7.
[PUBMED](#) | [CROSSREF](#)
20. Traylor M, Farrall M, Holliday EG, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol* 2012;11:951-62.
[PUBMED](#) | [CROSSREF](#)
21. Wang D, Sun Y, Stang P, Berlin JA, Wilcox MA, Li Q. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. *BMC Proc* 2009;3 Suppl 7:S109.
[PUBMED](#) | [CROSSREF](#)
22. Kim J, Oh B, Lim JE, Kim MK. No interaction with alcohol consumption, but independent effect of C12orf51 (HECTD4) on type 2 diabetes mellitus in Korean adults aged 40–69 years: the KoGES-Ansan and Ansung Study. *PLoS One* 2016;11:e0149321.
[PUBMED](#) | [CROSSREF](#)