Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jiali Chen Date

Gene-smoking interaction on incident stroke

and ischemic stroke

By

Jiali Chen
Master of Public Health

Epidemiology

_____

Yan V Sun
Committee Chair

Gene-smoking interaction on incident stroke and ischemic stroke

By

Jiali Chen

B.S.
University of California, Irvine
2017

Thesis Committee Chair: Yan V Sun, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2020

Abstract

Gene-smoking interaction on incident stroke and ischemic stroke

By Jiali Chen

Abstract: Cigarette smoking and genetic predisposition are established risk factors for all stroke subtypes. However, the interaction between smoking and genetic susceptibility on stroke and subtypes has been well understood. In this study, we examined the gene-smoking interaction on incident stroke and ischemic stroke using a genetic risk score (GRS) consisting of previous reported genetic loci associated with stroke. The prospective cohort study is based on the UK Biobank study, which included 392,997 Caucasians free of cardiovascular diseases and stroke at the baseline. Smoking status is classified as current vs. non-current, and ever vs. never smokers. Cox proportional hazard model is adopted to analyze the association of GRS and smoking status, as well as their interaction effect. After a median follow-up of 3,808 days, 2,847 incident stroke cases and 2,167 ischemic stroke cases were ascertained using diagnosis codes. Although GRS and smoking status were independently associated with incident stroke and ischemic stroke, the GRS-smoking interaction did not significantly predict incident stroke or ischemic stroke. These results do not support that gene-smoking interaction plays an important role in the development of stroke or ischemic stroke.

Key words: Genetic risk score, smoking, incident stroke, ischemic stroke, gene-environment interaction

Gene-smoking interaction on incident stroke and ischemic stroke

By

Jiali Chen

B.S.
University of California, Irvine
2017

Thesis Committee Chair: Yan V Sun, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2020

Introduction:

A stroke is a medical condition in which poor blood flow to the brain results in cell death. Stroke is the fifth cause of death and a leading etiology of disability in the United States (US)[1]. There are three main subtypes of stroke: ischemic stroke, due to artery blockage and lack of blood flow; hemorrhagic stroke, due to vessel rupture and bleeding; and transient ischemic attack (TIA). In 2016, about 87% of strokes were ischemic strokes caused by the blood clot blockages in the US[2]. There are various known risk factors for stroke, such as hypertension, diabetes mellitus, cigarette smoking, older age and male gender, race and ethnicity, and family history of stroke [3]. However, around 25-39% of patients' definite causes remained unidentified due to the complicated pathophysiology of stroke[4].

Previous twin studies and family history studies have shown that genetics play an essential role in the risk of stroke[5]. Genetic variants increase the risk of stroke through multiple mechanisms such as elevating the risk of stroke with comorbidity of hypertension or diabetes, by triggering specific pathophysiological processes such as atheroma or atrial fibrillation, by altering coagulation pathways and rendering patients predisposed to arterial thrombosis or bleeding, or by enhancing tolerance to brain ischemia and more largely brain injury[6].

MEGASTROKE, a large genome-wide association study (GWAS) in 521,612 individuals (67,162 cases and 454,450 controls), identifies 32 loci associated with stroke and 20 loci specifically associated with ischemic stroke[7]. Most of the loci relate to only one stroke subtype, and none of the loci are associated with all subtypes. Therefore, this study will first focus on all incident stroke cases in the UK Biobank cohort and then specific on incident ischemic stroke cases.

Cigarette smoking is an established, independent risk factor for stroke. According to the Centers for Disease Control and Prevention (CDC), there are 34.2 million U.S. adults who are current cigarette smokers[8]. Smokers had an overall increased risk of stroke compared with nonsmokers[9]. Compared with nonsmokers, cigarette smokers had 2.3 times higher risk of any stroke[10]. Current smokers had an increased risk of stroke with nonsmokers (OR: 1.46, 95% CI: 1.04–2.07) [11]. The odds ratio for the current smokers to get ischemic stroke was 1.88 times higher than the nonsmokers[12]. The relative risk of stroke associated with smoking varied by stroke types[13], ranging from 1.49 for ischemic stroke and 1.13 for TIA[14]. The odds ratio for the current smokers to get ischemic stroke was 1.88 times higher than the nonsmokers[15].

Gene-environment interaction (G×E) study formally analyzed the non-additive effects of genetic and environmental factors on a disease trait, and it potentially contributes to unexplained interindividual variability influenced by the main genetic and environmental risk factors[16]. To our knowledge, there are no previous studies that focus on whether smoking status could alter the influences of the genetic associations (i.e., G×Smoking interaction) on the stroke of different subtypes. Using a large sample size from the UK biobank cohort and a prospective design, we investigated whether G×Smoking interaction of known stroke loci helps to provide an insight into their detrimental role on the stroke of different subtypes.

Methods

Study Population

UK Biobank is a massive long-term biobank study in the United Kingdom (UK) which is used to investigate the respective contribution of genetic predisposition and environmental exposure to the progression of disease[Error! Reference source not found.]. It started in 2006 and, until 2010, recruited more than 500,000 participants aged 40 to 70 years from the general population in the UK[18]. In this study, people without genetic information are excluded from the study. Moreover, only participants who self-reported as white British and genetically identified as European ancestry are included in the study. People with a history of stroke, coronary heart disease, and heart failure at baseline are also excluded from this study (Figure 1).

Incident stroke and ischemic stroke

Incident stroke in the UK Biobank cohort was identified by hospital admission EHR and death registry. We used the stroke variables provided by UK Biobank, which were generated by combining information from International Classification of Disease (ICD) 9 codes (430.X, 431.X, 434.X, 434.0, 434.1, 434.9, 436.X) and ICD 10 codes (I60, I60.0-I60.9, I61, I61.0-I61.9, I63, I63.0-I63.9, I64.X). The UK Biobank Stroke Outcomes Group estimated the accuracy of defined stroke events based on two different systematic reviews. Ischemic stroke, a subset of incident stroke, was also based on hospital admission EHR and death register. Its combined information from ICD 9 codes (434.0, 434.1, 434.9, 436.X) and ICD 10 codes (I63, I63.1-I63.9, I64.X).

Genetic data and Construction of Genetic Risk Score (GRS)

Details of the arrays, sample processing, and quality control of the genetic resource in the UK Biobank have been described in detail elsewhere[19]. Overall, UK Biobank used the UK

BiLEVE Axiom array and the UK Biobank Axiom array, to genotype about 805,426 markers. The Haplotype Reference Consortium (HRC) and UK10K haplotype resources are used as a reference. All single nucleotide polymorphisms (SNPs) passed the Hardy-Weinberg equilibrium test with a P>0.001. There are 488,369 participants who have been imputed genotypes information.

MEGASTROKE includes participants from 29 studies with genome-wide genotypes imputed to 1000 Genomes Project (1000G) phase 1v3 or similar. Ancestry-specific meta-analysis and subsequent fixed-effects trans-ancestral meta-analyses and MANTRA trans-ancestral meta-analyses were conducted[20]. According to the latest meta-analysis which based on European ancestry, there are 32 SNPs associated with any stroke and 20 SNPs associated to ischemic stroke[21]. For each individual in the study population, we calculated the GRS, which defined as the weighted sum of its trait-associated alleles[22]. The effect size of the SNP-stroke association was based on previously MEGASTROKE studies. Table 2 provided detailed information on the SNPs used in this study. GRS are calculated respectively for the stroke and ischemic stroke by using Plink.

Phenotypes

Smoking status is defined as current, previous, and never smokers. Both diastolic and systolic blood pressure are read and recorded automatically by the Omron device. Two measures of blood pressure were carried out a few moments apart. BMI value is constructed from height and weight measured during the initial Assessment Centre visit. Hyperlipidemia is defined by ICD10 codes (E78.0, E78.1, E78.2, E78.3, E78.4, E78.5). Alcohol status is defined as more than three or four times a week and less than three or four times a week. Moderate physical

activity is defined as at least 150 minutes of moderate-intensity activity weekly or 75 minutes of vigorous activity weekly[23].

Statistical Analysis

Two binary variables for smoking status were derived as current vs. non-current smokers, and ever vs. never smokers. To examine the interaction between GRS and smoking status on the stroke, Cox Proportional Hazards models are used. Before running this model, collinearity and proportional hazards assumption were checked for all variables. For the patients having an incident stroke, the time of follow-up was calculated as the time between the baseline assessment and the first stroke even from the hospital data. For the non-stroke group, it is calculated as the time from baseline assessment to the death or loss of follow-up or 2019/09/19, which is the last update time. Age, sex, blood pressure, alcohol status, hyperlipidemia, and physical activity status are adjusted for the Cox proportional hazards models. After checking the collinearity of the original model, diastolic blood pressure was eliminated. The remaining variables met with the proportional hazard assumption. The final model included age, sex, BMI, systolic blood pressure, alcohol, hyperlipidemia, physical activity, and first ten principal components of ancestry as covariates. In addition to the main effects of GRS and smoking status, an interaction term of GRS and smoking status was included in the model. The same Cox proportional hazards models were for both incident and ischemic stroke. GRS for the ischemic stroke was calculated based on 20 associated SNPs for each individual. R studio is used to calculate the hazard rati (HR) and 95% confidence interval (CI) for the Cox proportional hazards regression.

Results

Table ***1*** shows the baseline characteristics of the study population (n=392,997). There were 2,847 incident stroke cases, and among that, 2,167 were ischemic stroke. 16.8% of the incident stroke cases were current smokers. When narrowed to the ischemic stroke, this number increased to 17.2%. The mean age for the ischemic stroke group was 61.6, 61.2 for the incident stroke group, and 56.4 for the non-incident stroke group. 40.2% of the ischemic stroke patients were female, 43.7% for the incident stroke group, and 56.3% for the non-incident stroke group. We calculated the GRS for each sample using Plink software. The unadjusted association between GRS and any stroke was 1.10 ($p<0.001$). The mean survival times for the sample population was 3808 days(Figure 3). The non-stroke group had mean follow-up time of 3,826 days. The mean time-to-event for incident stroke was 1,364 days. The mean time-to-event for the ischemic stroke was 1,369 days.

Incident Stroke

When the smoking status was defined as "never vs. ever smoker", age, systolic blood pressure, BMI, and hyperlipidemia all seem to increased the risk of stroke (Table 3). Alcohol consumption and physical activity were shown to decrease the risk of stroke. When we defined the smoking status as current and non-current, after adjusting for the other confounders including age, sex, systolic blood pressure, BMI, alcohol, hyperlipidemia, physical activity, and first ten principal components of ancestry, the hazard ratio between incident stroke and GRS was 1.11. When we defined the smoking status as never and ever smokers, the hazard ratio slightly changed from 1.11 to 1.10. Smoking present with a

stronger effect on the risk of stroke when we divided smoking status as current and non-current smokers than ever and never smokers (Current vs. non-current: HR=2.10 (95% CI:1.89,2.33), Ever vs. Never: HR=1.32 (95% CI: 1.22, 1.43)) (Table 3, Table 4).

The interaction term was statistically insignificant (Current vs. Non-current P=0.43, Never vs. Ever P=0.83) no matter how we categorized the smoking status.

Incident Ischemic Stroke

The unadjusted association between GRS and ischemic stroke was HR of 1.14 ($p<0.0001$), which was higher than that of the incident stroke (HR=1.10, $p<0.0001$) (Figure 4). Smoking had a stronger effect on the risk of stroke when we divided smoking status as current vs. non-current than ever vs. never. Besides, when we focus on the ischemic stroke, the hazard ratio between ischemic stroke and polygenetic risk score was 1.16. When we defined the smoking status as never and ever smokers instead of current and non-current, the ratio changed from 1.16 to 1.13. For the current smokers, they were 2.17 times (95% CI: 1.92, 2.45) higher risk than the non-current smokers to get the stroke. When defined the smoking status as ever and never, this number decreased to 1.28 times (95% CI: 1.17, 1.40) (Table 5 Table 6). The interaction term was also statistically insignificant (Current vs. Non-current P=0.12, Never vs. Ever P=0.73) no matter how we categorized the smoking status.

Discussion

This study focuses on the association between genetic risk of stroke, smoking, and incident risk of stroke based on the UK biobank cohort. When controlling for age, sex, systolic blood pressure, BMI, alcohol, hyperlipidemia, physical activity, and first ten principal components

of ancestry, the association between polygenetic risk score and incident stroke is statistically significant. The smoking status also exerts a statistically significant effect on the risk of the stroke. However, the interaction of polygenetic risk score and smoking status is not statistically significant. A similar result was discovered on ischemic stroke.

When defining the smoking status as current and non-current, for every 1 standard deviation (SD) increase on the genetic risk score, people had 1.10 times higher risk to develop stroke. This risk increased to 1.16 times higher per SD increase of GRS for ischemic stroke. Since most of the SNPs just related to single stroke subtype, the GRS is more precise when narrowed down to one specific subtype.

In this study, we defined the smoking status both as “current vs. non-current” and “ever vs. never.” The difference is more remarkable when smoking status is defined as current and non-current smokers. Previous studies only categorized smoking status as current and non-smokers. However, smoking has a long-time effect on stroke. Therefore, it is crucial to take those previous smoking groups into consideration. In this study, we include both definitions to see the difference. When we defined the smoking status as current vs. non-current, it indicated a higher risk of incident stroke (Current vs. Non-current: HR = 2.10, Ever vs. Never: HR = 1.32). This difference shows that the definition of the smoking status could alter the study result. Therefore, it is better to include both categorized definition in the further study.

The results of this study are not consistent with the previous research. The previous research found out the gene-environment interaction between specific SNP and smoking decreased the stroke risk in a Chinese population[24]. In this study, the interaction is not statistically

significant. However, those two studies included different sample populations. The UK Biobank includes a large sample size based on majority of Caucasian. Previous study had a relatively small population. Therefore, it is possible that the previous study had underestimated the interaction. Another probable reason for this divergence is that this study uses a GRS, which combines multiple SNPs instead of single SNPs on the regression model.

Strengths and Limitations

This study includes a large sample size from the UK biobank, which allows us to detect small genetic effects and make a more accurate estimate. Additionally, we used the latest MEGASTROKE to get the list of related SNPs. MEGASTROKE is the most extensive genome-wide association study of stroke for the time being. This could also greatly improve the accuracy of this study.

Our study also has some limitations that need to be pointed out. First, this study is based on the White British only. Most participants in this study are the elder. Despite the large sample size, the composition of the sample produced some limitations. Therefore, it might be inappropriate to apply the result to the general group. Secondly, the cases of stroke are relatively small compared to the non-stroke group. It might incur underestimation of the genetic and smoking association. However, the prevalence rate of stroke is relatively low in the European population. Therefore, it is reasonable to establish a small-size stroke group. Moreover, smoking status is self-reported. The accuracy of the smoking status is uncertain. Last but not least, the smoking status changed after baseline examinations might have had an effect on the risk estimation.

For the future study, with more SNPs related to the stroke are identified, the evaluation of the risk of stroke will be more accurate. Besides, further study could also study how the amount of cigarette smoking could affect the risk of stroke and the interaction term.

Conclusion

In this prospective cohort study of 392,997 people based on UK biobank, we found that genetic and smoking status were independently associated with the risk of incident stroke and ischemic stroke. This study provides evidence that people with a higher genetic risk score are more likely to have a high risk of stroke. It also explained that smoking status has a more substantial effect on the risk of stroke. Therefore, to lower the smoking rate is a hopefully possible approach to reduce the incidence of stroke. For the future approach, we could have more policy limitations on smoking consumption to reduce the risk of stroke.

Acknowledgement

This research has been conducted using the UK Biobank Resource under Application Number ‘34031’.

References

1. About Stroke. (n.d.). Retrieved from https://www.stroke.org/en/about-stroke.
2. Types of Stroke. (2020, January 31). Retrieved from https://www.cdc.gov/stroke/types_of_stroke.htm
3. Stroke. (n.d.). Retrieved from https://www.nhlbi.nih.gov/health-topics/stroke.
4. Norrving, B. (2009). Classification of Stroke Subtypes. *Stroke*, 50–63. doi: 10.1159/000210272
5. Brass, L. M., Isaacsohn, J. L., Merikangas, K. R., & Robinette, C. D. (1992). A study of twins and stroke. *Stroke*, *23*(2), 221–223. doi: 10.1161/01.str.23.2.221
6. Hassan A, Markus HS. Genetics and ischaemic stroke. Brain. 2000;123(Pt 9):1784–1812. doi: 10.1093/brain/123.9.1784.
7. Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., … Dichgans, M. (2018). Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics*, *50*(4), 524–537. doi:10.1038/s41588-018-0058-3
8. Burden of Cigarette Use in the U.S. (2019, December 6). Retrieved from https://www.cdc.gov/tobacco/campaign/tips/resources/data/cigarette-smoking-in-united-states.html?s_cid=OSH_tips_GL0005&utm_source=google&utm_medium=cpc&utm_campaign=TipsRegular;S;WL;BR;IMM;DTC;CO&utm_content=Smoking+-+Facts_P&utm_term=statistics+about+smoking&&gclid=Cj0KCQjwu6fzBRC6ARIsAJUwa2TFEcEOo3YyHV8QWEADux4BCX96v92F1vn2BdkwqA42tTbvZ17slIcaAnktEALw_wcB&gclsrc=aw.ds
9. Collins, T. R. (2019). Health and Lifestyle Factors Can Mitigate Dementia Risk in Those with Low Genetic Risk. *Neurology Today*, *19*(20), 10–11. doi: 10.1097/01.nt.0000604232.10286.3d
10. Zhang Y, Galloway JM, Welty TK, et al. Incidence and risk factors for stroke in American Indians: the Strong Heart Study. Circulation. 2008;118:1577–1584.
11. Pan, B., Jin, X., Jun, L., Qiu, S., Zheng, Q., & Pan, M. (2019). The relationship between smoking and stroke: A meta-analysis. *Medicine*, *98*(12), e14872. https://doi.org/10.1097/MD.0000000000014872
12. Markidan, J., Cole, J. W., Cronin, C. A., Merino, J. G., Phipps, M. S., Wozniak, M. A., & Kittner, S. J. (2018). Smoking and Risk of Ischemic Stroke in Young Men. *Stroke*, *49*(5), 1276–1278. https://doi.org/10.1161/STROKEAHA.117.018859
13. Lackland, D. T., Roccella, E. J., Deutsch, A. F., Fornage, M., George, M. G., Howard, G., … Council on Functional Genomics and Translational Biology (2014). Factors influencing the decline in stroke mortality: a statement from the American Heart Association/American Stroke Association. *Stroke*, *45*(1), 315–353. doi:10.1161/01.str.0000437068.30550.cf
14. Weikert C, Berger K, Heidemann C, et al. Joint effects of risk factors for stroke and transient ischemic attack in a German population: the EPIC Potsdam Study. J Neurol. 2007;254:315–321.
15. Huan, T., Joehanes, R., Schurmann, C., Schramm, K., Pilling, L. C., Peters, M. J., … Levy, D. (2016). A whole-blood transcriptome meta-analysis identifies gene expression

signatures of cigarette smoking. *Human molecular genetics*, *25*(21), 4611–4623. doi:10.1093/hmg/ddw288
16. Dick D. M. (2011). Gene-environment interaction in psychological traits and disorders. *Annual review of clinical psychology*, *7*, 383–409. https://doi.org/10.1146/annurev-clinpsy-032210-104518
17. About U.K. Biobank: U.K. Biobank. (n.d.). Retrieved from https://www.ukbiobank.ac.uk/about-biobank-uk/
18. UKBiobank.UKBiobank:protocolfora large-scale prospective epidemiological resource. http://www.ukbiobank.ac.uk/wp-content/uploads /2011/11/UK-Biobank-Protocol.pdf. Accessed December 15, 2015.
19. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., … Marchini, J. (2017). Genome-wide genetic data on ~500,000 UK Biobank participants. doi: 10.1101/166298
20. Morris, A. P. (2011). Transethnic meta-analysis of genomewide association studies. *Genetic Epidemiology*, *35*(8), 809–822. doi: 10.1002/gepi.20630
21. Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., Rutten-Jacobs, L., Giese, A. K., van der Laan, S. W., Gretarsdottir, S., Anderson, C. D., Chong, M., Adams, H., Ago, T., Almgren, P., Amouyel, P., Ay, H., Bartz, T. M., Benavente, O. R., Bevan, S., … Dichgans, M. (2018). Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics*, *50*(4), 524–537. https://doi.org/10.1038/s41588-018-0058-3
22. Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics*, *9*(3). doi: 10.1371/journal.pgen.1003348
23. Booth, F. W., Roberts, C. K., Thyfault, J. P., Ruegsegger, G. N., & Toedebusch, R. G. (2017). Role of Inactivity in Chronic Diseases: Evolutionary Insight and Pathophysiological Mechanisms. *Physiological reviews*, *97*(4), 1351–1402. https://doi.org/10.1152/physrev.00019.2016
24. Wu, Z., Huang, Y., Huang, J., & Fan, L. (2017). Impact of CRP gene and additional gene–smoking interaction on ischemic stroke in a Chinese Han population. *Neurological Research*, *39*(5), 442–447. doi: 10.1080/01616412.2017.1297905

Figure 1 Selection of the sample population

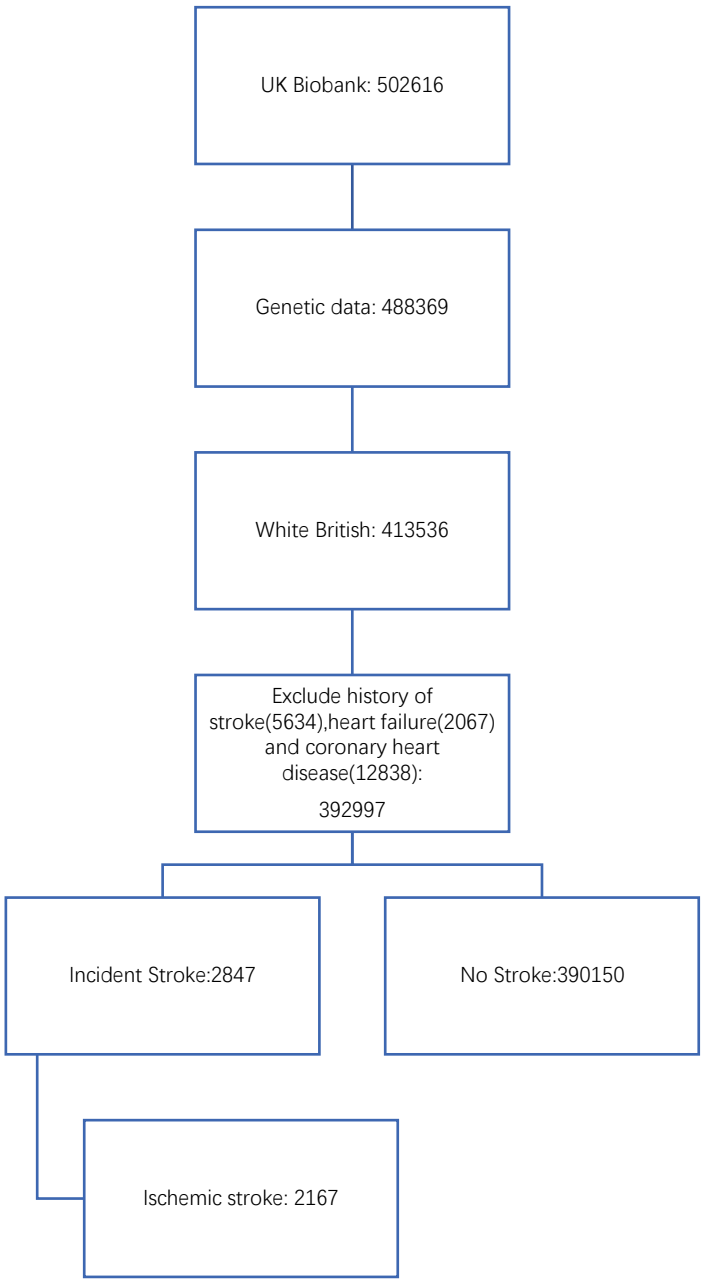


Figure 2 Density plot of GRS for any stroke and ischemic strok, standardized the GRS to z-score on the x-axis.

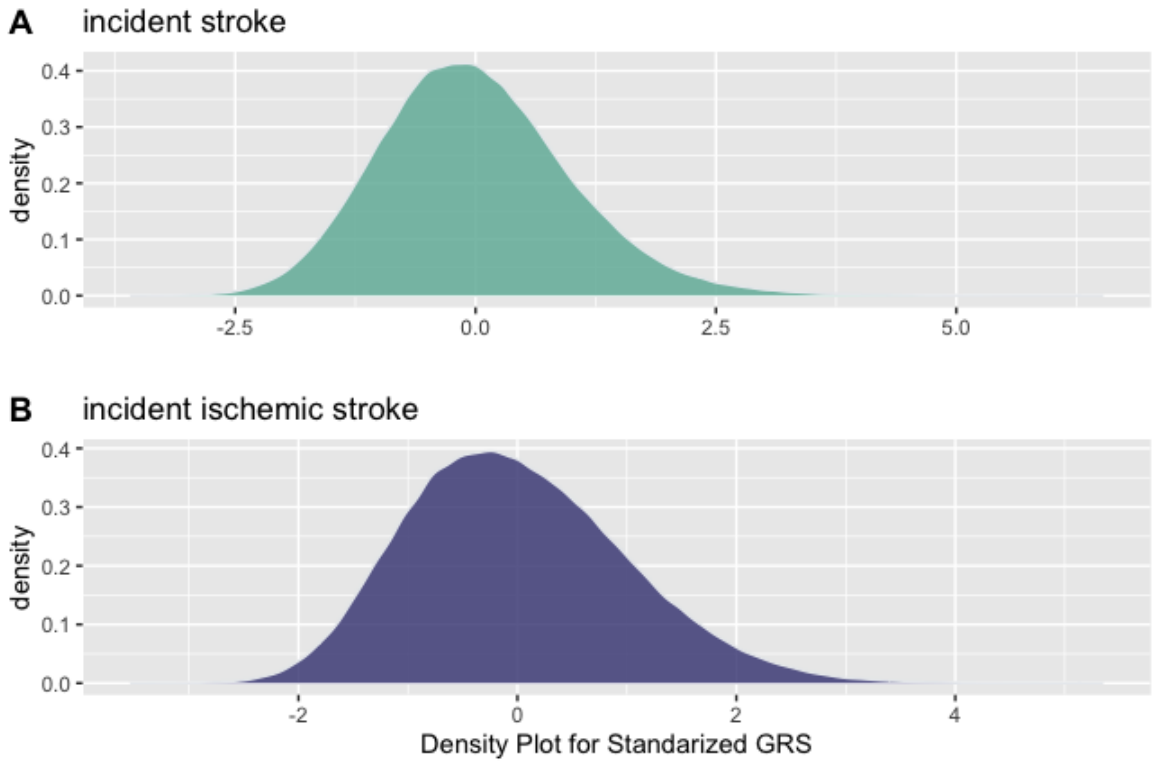


Figure 3 Density plot for the survival time for incident stroke.

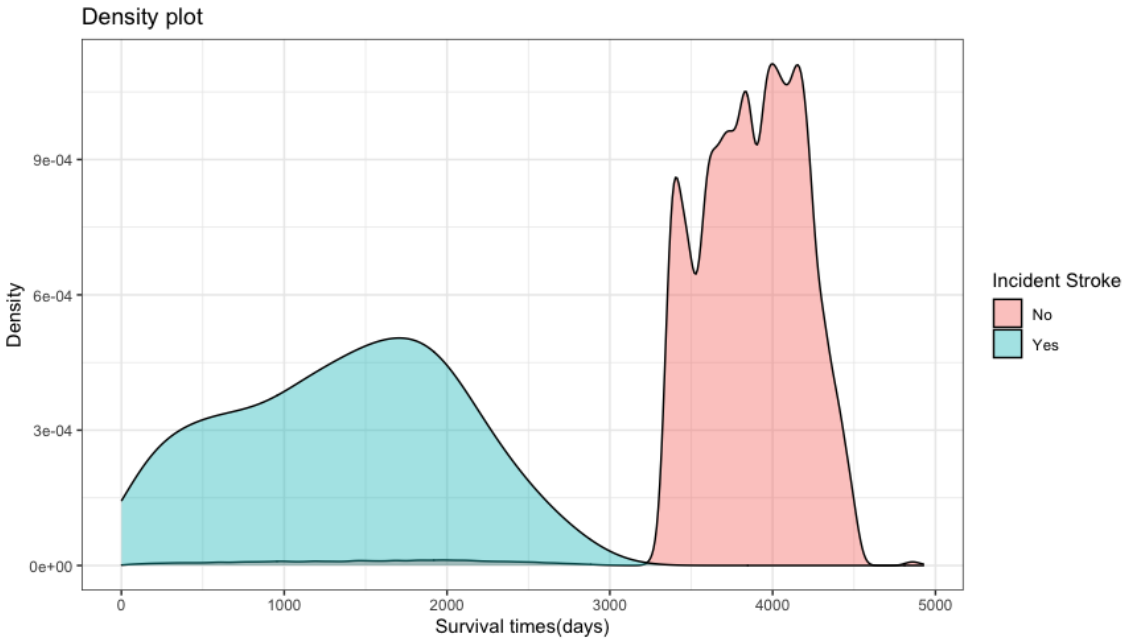
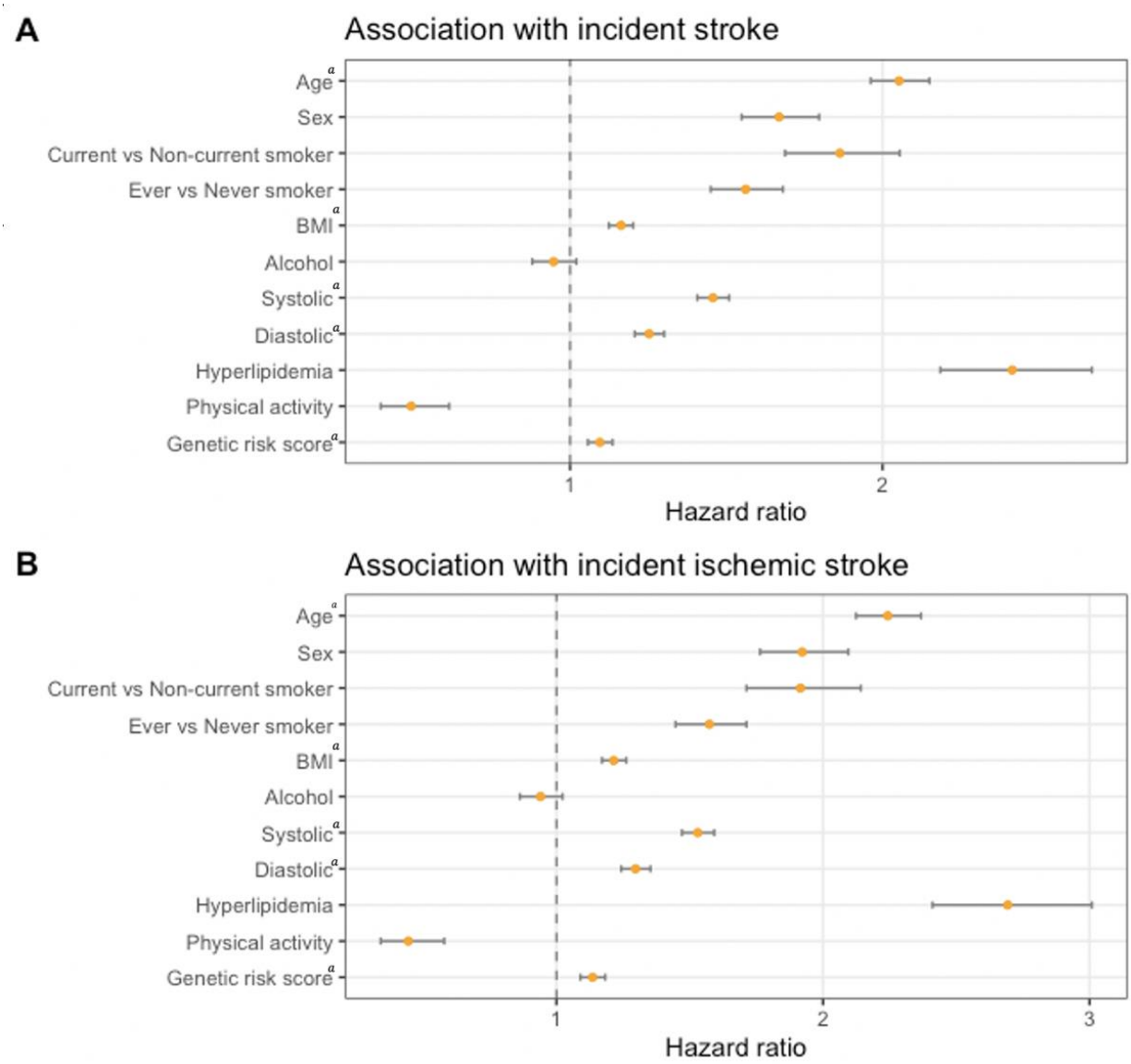


Figure 4 Unadjusted associations with incident stroke and incident ischemic stroke for all stroke risk factors.



[a]Standardized age, BMI, systolic blood pressure, diastolic blood pressure, and Genetic risk score.

Table 1 Baseline characteristics for study population

	Incident Ischemic Stroke	**All Incident Stroke**	**No Stroke**
	N=2167	**N =2847**	**N=390150**
Age (years) Mean (SD)	61.6 (6.60)	61.2 (6.77)	56.4 (8.01)
Sex:			
Female	872 (40.2%)	1242 (43.7%)	219765 (56.3%)
Male	1295 (59.8%)	1604 (56.3%)	170385 (43.7%)
Smoking Status:			
Never	970 (44.8%)	1280 (45.0%)	218277(55.9%)
Previous	825 (38.1%)	1090 (38.3%)	133508 (34.2%)
Current	372 (17.2%)	477 (16.8%)	38365 (9.83%)
Diastolic blood pressure (mmHg) Mean (SD)	85.1 (11.8)	84.7 (11.8)	82.3 (10.6)
Systolic blood pressure (mmHg) Mean (SD)	149 (22.1)	148 (21.7)	140 (19.5)
BMI (kg/m^2) Mean (SD)	28.3 (4.96)	28.0 (4.95)	27.2 (4.71)
Alcohol:			
More than three or four times a week	950 (43.9%)	1254 (44.1%)	177149 (45.4%)
Hyperlipidemia	379 (17.5%)	455 (16.0%)	28532 (7.31%)
Physical activity:			
Moderate	56 (2.58%)	81 (2.85%)	21864 (5.60%)

Non-moderate	2111(97.4%)	2762(97.2%)	273018(94.1%)

Table 2 SNPs used for calculated different GRS

A. SNPs used for incident stroke

rsID	Chromosome	Gene(s)	Location relative to gene	Risk allele	OR	95%CI
rs880315	1p36	CASZI	Intronic	C/T	1.05	(1.04, 1.07)
rs120379897	1p13	WBR2B	Intronic	C/T	1.07	(1.05, 1.10)
rs12124533	1p13	TSPAN2	Intergenic	T/C	1.17	(1.11, 1.23)
rs1052053	1q22	PMF1-SEMA4A	Exonic;nonsynonymous	G/A	1.06	(1.05, 1.08)
rs146390073	1q43	RGS7	Intronic	T/C	1.95	(1.54, 2.47)
rs12476527	2p23	KCNK3	5'-UTR	G/T	1.05	(1.03, 1.07)
rs7610618	3q25	TM4SF4-TM4Sn	Intergenic	T/C	2.33	(1.75, 3.12)
rs34311906	4q25	ANK2	Intergenic	C/T	1.07	(1.04, 1.09)
rs17612742	4q31	EDNRA	Intergenic	C/T	1.19	(1.13, 1.26)
rs6825454	4q31	FGA	Intergenic	C/T	1.06	(1.04, 1.08)
rs13143308	4q25	PITX2	Intergenic	T/G	1.32	(1.27, 1.37)
rs11957829	5q23	LOC100505841	Intronic	A/G	1.07	(1.05, 1.10)
rs6891174	5q35	NKX2-5	Intergenic	A/G	1.11	(1.07, 1.16)
rs16896398	6q21	SLC22A7-ZNF318	Intergenic	T/A	1.05	(1.03, 1.07)
rs4959130	6q25	FOXF2	Intergenic	A/G	1.08	(1.05, 1.11)
rs42039	7q21	CDK6	3'-UTR	C/T	1.07	(1.04, 1.09)
rs2107595	7q21	HDAC9-TWIST1	Intergenic	A/G	1.21	(1.15, 1.26)
rs7859727	9q21	chr9p21	ncRNA intronic	T/C	1.05	(1.03, 1.07)
rs10820405	9q31	LINC01492	ncRNA intronic	G/A	1.20	(1.12, 1.28)
rs635634	9q34	ABO	Intergenic	T/C	1.08	(1.05, 1.11)
rs2295786	10q24	SH3PXD24	Intergenic	A/T	1.05	(1.04, 1.07)
rs2005108	11q22	MMP12	Intergenic	T/C	1,08	(1.05, 1.11)
rs7304841	12q12	PDE3A	Intronic	A/C	1.05	(1.03, 1.07)
rs35436	12q24	TBX3	Intergenic	C/T	1.05	(1.03, 1.06)
rs3184504	12q24	SH2B3	Exonic;nonsynonymous	T/C	1.08	(1.06, 1.10)
rs9526212	13q14	LRCH1	Intronic	G/A	1.06	(1.04, 1.08)
rs4932370	15q26	FURIN-FES	Intergenic	A/G	1.05	(1.03, 1.07)
rs12932445	16q22	ZFHX3	Intronic	C/T	1.20	(1.15, 1.25)

rs12445022	16q24	ZCCHC14	Intergenic	A/G	1.06	(1.04, 1.08)
rs11867415	17q13	PRPF8	Intronic	G/A	1.09	(1.06, 1.13)
rs2229383	19q13	ILF3-SLC44A2	Exonic; synonymous	T/G	1.05	(1.03, 1.07)
rs8103309	19q13	SMARCA4-LDLR	Intergenic	T/C	1.05	(1.03, 1.07)

B. SNPs used for ischemic stroke

rsID	**Chromosome**	**Gene(s)**	**Location relative to gene**	**Risk allele**	**OR**	**95%CI**
rs880315	1p36	CASZI	Intronic	C/T	1.05	(1.04, 1.07)
rs1052053	1q22	PMF1-SEMA4A	Exonic;nonsynonymous	G/A	1.06	(1.05, 1.08)
rs34311906	4q25	ANK2	Intergenic	C/T	1.07	(1.04, 1.09)
rs6825454	4q31	FGA	Intergenic	C/T	1.06	(1.04, 1.08)
rs13143308	4q25	PITX2	Intergenic	T/G	1.32	(1.27, 1.37)
rs11957829	5q23	LOC100505841	Intronic	A/G	1.07	(1.05, 1.10)
rs4959130	6q25	FOXF2	Intergenic	A/G	1.08	(1.05, 1.11)
rs42039	7q21	CDK6	3’-UTR	C/T	1.07	(1.04, 1.09)
rs2107595	7q21	HDAC9-TWIST1	Intergenic	A/G	1.21	(1.15, 1.26)
rs7859727	9q21	chr9p21	ncRNA intronic	T/C	1.05	(1.03, 1.07)
rs635634	9q34	ABO	Intergenic	T/C	1.08	(1.05, 1.11)
rs2005108	11q22	MMP12	Intergenic	T/C	1,08	(1.05, 1.11)
rs7304841	12q12	PDE3A	Intronic	A/C	1.05	(1.03, 1.07)
rs35436	12q24	TBX3	Intergenic	C/T	1.05	(1.03, 1.06)
rs3184504	12q24	SH2B3	Exonic;nonsynonymous	T/C	1.08	(1.06, 1.10)
rs9526212	13q14	LRCH1	Intronic	G/A	1.06	(1.04, 1.08)
rs4932370	15q26	FURIN-FES	Intergenic	A/G	1.05	(1.03, 1.07)
rs12445022	16q24	ZCCHC14	Intergenic	A/G	1.06	(1.04, 1.08)
rs11867415	17q13	PRPF8	Intronic	G/A	1.09	(1.06, 1.13)
rs2229383	19q13	ILF3-SLC44A2	Exonic; synonymous	T/G	1.05	(1.03, 1.07)

Table 3: Cox proportional hazard model with multivariable for any stroke. Standardized GRS and smoking status defined as ever and never smoker.

Analysis of Maximum Likelihood Estimates						
Parameter	**Parameter Estimate**	**Standard error**	**chi-square**	**Pr>chisq**	**Hazard ratio**	**95% CI**
GRS	0.10	0.029	12.19	0.0005	1.10	(1.05,1.17)
Smoking status	0.28	0.041	46.47	<.0001	1.32	(1.22,1.43)
Age	0.08	0.003	595.54	<.0001	1.08	(1.08,1.09)
Sex	0.40	0.044	82.51	<.0001	1.37	(1.37,1.63)
Systolic blood pressure	0.01	0.001	81.43	<.0001	1.01	(1.01,1.01)
BMI	0.02	0.004	21.29	<.0001	1.01	(1.01,1.03)
Alcohol	-0.22	0.041	27.48	<.0001	0.74	(0.74,0.87)
Hyperlipidemia	0.24	0.060	15.94	<.0001	1.13	(1.13,1.43)
Physical activity	-0.63	0.117	28.99	<.0001	0.43	(0.43,0.67)
GRS*Smoking	-0.01	0.039	0.05	0.829		

Table 4 Cox proportional hazard model with multivariable for any stroke. Standardized GRS and smoking status defined as current and non-current smoker.

Analysis of Maximum Likelihood Estimates						
Parameter	**Parameter Estimate**	**Standard error**	**chi-square**	**Pr>chisq**	**Hazard ratio**	**95% CI**
GRS	0.10	0.021	23.51	<.0000	1.11	(1.06,1.15)
Smoking status	0.74	0.054	185.43	<.0001	2.10	(1.89,2.33)
Age	0.08	0.003	661.25	<.0001	1.08	(1.08,1.09)
Sex	0.39	0.044	80.50	<.0000	1.36	(1.36,1.63)
Systolic blood pressure	0.01	0.001	85.87	<.0001	1.01	(1.01,1.01)
BMI	0.02	0.004	30.84	<.0001	1.02	(1.02,1.03)
Alcohol	-0.18	0.041	19.82	<.0001	0.77	(0.77,0.90)
Hyperlipidemia	0.25	0.060	17.45	<.0001	1.14	(1.14,1.44)
Physical activity	-0.60	0.117	26.49	<.0001	0.44	(0.44,0.69)
GRS*Smoking	-0.04	0.053	0.63	0.429		

Table 5 Cox proportional hazard model with multivariable for ischemic stroke. Standardized GRS and smoking status defined as ever and never smoker.

Analysis of Maximum Likelihood Estimates						
Parameter	**Parameter Estimate**	**Standard error**	**chi-square**	**Pr>chisq**	**Hazard ratio**	**95% CI**
GRS	0.13	0.033	14.66	0.0001	1.13	(1.06,1.21)
Smoking status	0.25	0.047	27.54	<.0001	1.29	(1.17,1.40)
Age	0.09	0.004	532.95	<.0001	1.09	(1.08,1.10)
Sex	0.54	0.051	110.76	<.0001	1.71	(1.55,1.89)
Systolic blood pressure	0.01	0.001	85.44	<.0001	1.01	(1.01,1.01)
BMI	0.03	0.005	40.55	<.0001	1.03	(1.02,1.04)
Alcohol	-0.23	0.047	23.66	<.0001	0.80	(0.73,0.87)
Hyperlipidemia	0.23	0.066	12.63	0.0004	1.26	(1.11,1.44)
Physical activity	-0.73	0.141	27.23	<.0001	0.48	(0.36,0.63)
GRS*Smoking status	0.02	0.044	0.12	0.732		

Table 6 Cox proportional hazard model with multivariable for ischemic stroke. Standardized GRS and smoking status defined as current and non-current smoker.

Analysis of Maximum Likelihood Estimates						
Parameter	**Parameter Estimate**	**Standard error**	**chi-square**	**Pr>chisq**	**Hazard ratio**	**95% CI**
GRS	0.15	0.024	38.52	<.0001	1.16	(1.11,1.22)
Smoking status	0.78	0.062	156.58	<.0001	2.17	(1.92,2.45)
Age	0.09	0.004	586.35	<.0001	1.10	(1.09,1.10)
Sex	0.53	0.051	107.78	<.0001	1.70	(1.53,1.87)
Systolic blood pressure	0.01	0.001	90.06	<.0001	1.01	(1.01,1.01)
BMI	0.03	0.005	51.70	<.0001	1.04	(1.03,1.05)
Alcohol	-0.20	0.047	18.08	<.0001	0.82	(0.75,0.90)
Hyperlipidemia	0.24	0.066	13.59	0.0002	1.28	(1.12,1.45)
Physical activity	-0.71	0.141	25.13	<.0001	0.49	(0.38,0.65)
GRS*Smoking status	-0.09	0.060	2.38	0.123		

Table 7 Combined association of GRS and smoking status with incident stroke and incident ischemic stroke. Hazard ratio for the interaction term based on the multivariable cox proportional hazard model.

	Smoking Status							
	Non-Current		Current		Never		Ever	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
GRS – Incident Stroke		Ref	1.06	(0.97,1.17)		Ref	1.32	(1.21,1.42)
GRS – Ischemic Stroke		Ref	1.06	(0.95,1.18)		Ref	1.15	(1.09,1.22)