

Modifiable lifestyle factors and risk of stroke: a Mendelian randomization analysis

Eric L Harshfield^{1,*}, Marios K Georgakis^{2,3*}, Rainer Malik², Martin Dichgans^{2,4,5†}, Hugh S Markus^{1,†}

¹Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK

²Institute for Stroke and Dementia Research, University Hospital, Ludwig-Maximilians-University, Munich, Germany

³Graduate School for Systemic Neurosciences, Ludwig-Maximilians-University, Munich, Germany

⁴Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

⁵German Centre for Neurodegenerative Diseases (DZNE), Munich, Germany

*These authors contributed equally to this work.

†These authors jointly supervised this work.

Correspondence to: Dr Eric Harshfield
Department of Clinical Neurosciences
University of Cambridge
R3, Box 83, Cambridge Biomedical Campus
Cambridge, CB2 0QQ, United Kingdom
Email: eh457@medschl.cam.ac.uk
Phone: +44 1223 217697

Running title: Mendelian randomisation analysis of lifestyle factors and risk of stroke

Word count: Abstract: 216; Main text: 3,004

ABSTRACT

Aims: Assessing whether modifiable risk factors are causally associated with reduced stroke risk is important in planning public health measures, but determining causality can be difficult in epidemiological data. Leveraging large-scale genetic data in a technique known as Mendelian randomisation, we aimed to determine whether modifiable lifestyle factors including educational attainment, smoking, and body mass index are causal risk factors for ischaemic stroke and its different subtypes and haemorrhagic stroke.

Methods and Results: We performed two-sample and multivariable Mendelian randomization to assess the causal effect of twelve lifestyle factors on risk of stroke and whether these effects are independent. We found genetic predisposition to increased number of years of education to be inversely associated with ischaemic, large-artery, and small-vessel stroke, as well as with intracerebral haemorrhage. Genetic predisposition to ever smoking regularly, higher body mass index (BMI), and higher waist-hip ratio are also associated with ischaemic and large-artery stroke. Additionally, we found that the effects of education, BMI, and smoking on ischaemic stroke to be independent of each other.

Conclusion: Genetic predisposition to higher educational attainment can reduce the risk of ischaemic, large-artery, and small-vessel stroke, while genetic predisposition to smoking and higher anthropometry measures can increase the risk of these stroke subtypes. This suggests that lifestyle modification addressing these risk factors will reduce stroke risk.

Keywords: Mendelian randomization; education; smoking; body mass index; stroke; genetics

INTRODUCTION

Reducing the burden of stroke in the population requires identification of modifiable risk factors, and the demonstration that reducing them reduces stroke risk.¹ A large number of lifestyle factors have been associated with stroke risk,² but demonstrating whether these associations are causal, and therefore whether modification of a particular risk factor will reduce stroke risk, is less clear. For some environmental factors such as smoking, convincing evidence suggests causality, but for many others it is uncertain whether they are indeed causal. Demonstrating causality is impossible from cross-sectional studies, and can even be challenging in longitudinal epidemiological studies in which risk factor exposure may be captured only over a few years rather than the entire lifespan. One way to assess causality is with a technique known as Mendelian randomization (MR), which uses genetic variants as instrumental variables in an approach analogous to a randomised controlled trial in which allocation to the “treatment group” is determined at conception, to assess whether risk factors have a causal association with an outcome of interest.³

A further complicating factor is that stroke represents a syndrome rather than a specific disease, and can be caused by a variety of different pathologies.⁴ The majority (approximately 80%) of strokes are ischaemic, while about 20% are haemorrhagic.⁴ Ischaemic stroke can itself be further divided into aetiological subtypes, of which the main three are cardioembolic (CES), large artery atherosclerotic (LAS), and lacunar or small vessel disease (SVS).⁴ Recent genetic studies indicate that these have distinct genetic and pathophysiological characteristics.⁵ Most epidemiological data examining risk factor modification does not facilitate determination of the effect of modifiable risk factors on specific stroke subtypes.

In this study, we used an MR approach to investigate the aetiological role of modifiable lifestyle risk factors for stroke and its subtypes. We studied a range of modifiable risk factors, including educational attainment, sleep duration, physical activity, smoking status, alcohol consumption, coffee consumption, dietary components, body mass index (BMI), and waist-hip ratio (WHR). As well as examining associations with ischaemic and haemorrhagic stroke, we further investigated the association of these traits with specific ischaemic stroke subtypes (CES, LAS, and SVS). We performed two-sample MR, using genetic variants associated with lifestyle traits as instrumental variables, to determine whether these traits are causally implicated in the risk of stroke. We also performed mediation analysis to evaluate whether the effects of education on ischaemic stroke subtypes are independent to those of BMI and smoking.

METHODS

Study design

We performed a two-sample MR analysis, following guidelines for performing and reporting MR studies,^{6,7} using summary statistics from the largest publicly available genome-wide association studies (GWAS) on the following twelve lifestyle factors: educational attainment,⁸ sleep duration,⁹ physical activity,¹⁰ smoking status,¹¹ alcohol consumption,¹² coffee consumption,¹³ four dietary components,¹⁴ BMI,¹⁵ and WHR.¹⁶ The sample sizes of the included studies ranged from 235,391 to 1.1 million individuals, all of whom were of European ancestry, from cohorts that included UK Biobank, the GIANT consortium, deCODE, and 23andMe. A description of each trait is listed in **Supplementary Table 1**, along with details of the corresponding studies and genetic instruments used.

Data sources

Outcome data

We obtained summary statistics from GWAS of all stroke (AS) and ischaemic stroke subtypes from the MEGASTROKE Consortium,⁵ which consisted of data from 67,162 cases and 454,450 controls. MEGASTROKE had 60,341 cases with any ischaemic stroke (AIS) regardless of subtype, of which 9,006 were cardioembolic stroke (CES), 6,688 were large-artery stroke (LAS), and 11,710 were small vessel stroke (SVS). Stroke cases were defined based on World Health Organization criteria (i.e. sudden onset neurological changes of presumed vascular origin lasting at least 24 hours) with stroke subtypes classified according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria.¹⁷ To avoid bias due to population stratification, we restricted our analysis to Europeans since all of the GWAS of the lifestyle traits were also conducted in individuals of European ancestry. From the International

Stroke Genetics Consortium (ISGC), we also obtained GWAS summary statistics on intracerebral haemorrhage (ICH),¹⁸ which consisted of 1,545 cases and 1,481 controls.

Instrumental variable selection

For each lifestyle factor included in this analysis, we selected genetic instruments significant at $P < 1 \times 10^{-6}$ based on the largest published GWAS for that trait with publicly available summary statistics. We also conducted a sensitivity analysis in which we only included single nucleotide polymorphisms (SNPs) that reached genome-wide significance ($P < 5 \times 10^{-8}$) for association with each risk factor. We then performed linkage disequilibrium (LD) clumping, which ensured that the instruments used for each trait were independent, by selecting only the SNP with the lowest P -value amongst all SNPs with an LD $r^2 \geq 0.001$.

We obtained data on educational attainment from a GWAS of time spent in education measured in years, from which we selected 440 SNPs. The study included 1.1 million individuals from 71 cohorts (including UK Biobank, 23andMe, and deCODE).⁸ The GWAS of sleep duration analysed sleep duration measured in hours in 446,118 individuals from UK Biobank, adjusted for age, sex, 10 principal components of ancestry, genotyping array, and genetic correlation matrix, with 125 SNPs selected as genetic instruments.⁹ The GWAS of physical activity analysed moderate-to-vigorous physical activity, which was measured in MET-minutes per week with adjustment for age, sex, genotyping chip, first 10 genomic principal components, centre, and season, in 377,234 individuals from UK Biobank, and we selected 7 SNPs as instruments from this study.¹⁰ The GWAS of smoking status used a lifetime smoking index, which was constructed in 462,690 individuals from UK Biobank by developing a model that incorporated self-reported time started smoking, duration of smoking, and cigarettes per day, as well as half-life and lag time constants to capture the non-linear risk of smoking on health.

We selected 210 SNPs as instruments for the lifetime smoking index.¹¹ The GWAS of alcohol consumption was based on the number of drinks consumed per week in 941,280 individuals from a number of cohorts including UK Biobank, 23andMe, and deCODE, from which we selected 83 SNPs.¹² The GWAS of coffee consumption was based on the number of cups of coffee consumed per day as reported by 336,448 UK Biobank participants on a touchscreen questionnaire at the assessment centre, with adjustment for age, sex, BMI, total energy, proportion of 24-hour recalls self-reported as capturing “typical intake,” and top 20 principal components, for which our genetic instrument had 49 SNPs.¹³ The GWAS of dietary components consisted of four traits (fat, protein, carbohydrate, and sugar) measured in up to 268,922 participants from UK Biobank, the DietGen consortium, and several other cohorts, with 5, 7, 8, and 9 SNPs respectively as genetic instruments.¹⁴ The GWAS of BMI was performed in 681,275 individuals from UK Biobank and the Health and Retirement Study, with adjustment for age, sex, recruitment centre, genotyping batches, and 10 principal components,¹⁵ and the GWAS of WHR was performed in 694,649 individuals from UK Biobank and the GIANT consortium with adjustment for SNP array.¹⁶ These studies included 623 and 380 SNPs respectively as instruments.

Statistical analyses

We used the inverse-variance weighted (IVW) method as the primary MR analyses, which uses inverse-variance-weighted meta-analysis under a random-effects model to combine the ratio estimates from each genetic variant into a single estimate of the causal effect of the lifestyle trait on the outcome.³ We also conducted sensitivity analyses using a variety of robust MR methods, which employ different sets of assumptions to make reliable causal inferences. These approaches included MR-Egger regression, the weighted median estimator, and the simple and weighted mode-based estimators.³

For each trait, we harmonised all SNPs associated with the trait with the outcome data to ensure that the effect estimates of each SNP on the exposure and on the outcome corresponded to the same effect allele. We then performed MR using the IVW method and sensitivity analyses using additional MR approaches. We accounted for multiple testing using a false discovery rate (FDR) cut-off of $q < 0.05$. Analyses were performed in R version 3.6.1 (R Core Team, 2019) using the TwoSampleMR package version 0.4.25.¹⁹ Two-sided P -values and 95% confidence intervals are presented.

We also performed MR-based network mediation analysis²⁰ to examine the extent to which increases in BMI and increased levels of smoking mediate the protective effective of education on ischaemic stroke subtypes. We calculated the total effect of education on stroke using two-sample univariable random-effects MR using the IVW approach. To calculate the direct effects, we combined the instruments for education, BMI, and smoking, and performed multivariable MR analyses to estimate their associations with the stroke subtypes. We calculated the indirect effect of education on stroke that acts through either BMI or smoking as the difference between the direct and total effect estimates. Finally, we calculated the proportion of the total effect of education on stroke that was mediated by BMI and smoking by dividing the indirect effect by the total effect. We also used the delta method²¹ to derive standard errors for each of the effects.

RESULTS

We analysed twelve lifestyle traits for association with AS, AIS, ICH, and the three subtypes of ischaemic stroke (CES, LAS, and SVS). **Figure 1** summarises the direction and magnitude of the association estimates for each lifestyle trait with each outcome, with stronger associations indicated by darker colours (red in the positive direction and blue in the negative direction) and asterisks to indicate the level of statistical significance. **Figure 2** shows scatterplots of the associations of each genetic variant plotted against their association with corresponding outcomes for all traits that had significant ($FDR\ q < 0.05$) associations. Forest plots of associations between lifestyle traits and each stroke subtype using the random-effects inverse-variance weighted method are shown in **Figure 3**. The full results for the association of each of the twelve traits with all six outcomes, using each MR method, are provided in **Supplementary Table 2**.

We found significant inverse associations of genetically determined number of years of education with ischaemic stroke (OR: 0.68, 95% CI: 0.63-0.75) and intracerebral haemorrhage (OR: 0.34, 0.20-0.58), as well as with large-artery (OR: 0.53, 0.42-0.67) and small-vessel stroke (OR: 0.65, 0.53-0.80). There was also suggestive (but not statistically significant) evidence of an inverse association of education with cardioembolic stroke (OR: 0.83, 0.70-0.99).

Genetic predisposition to lifetime smoking was significantly associated with ischaemic stroke (OR: 1.23, 1.10-1.39), large-artery (OR: 1.72, 1.26-2.36) and small-vessel stroke (OR: 1.69, 1.27-2.25), but not with cardioembolic stroke (OR: 0.83, 0.70-0.99). The association with intracerebral haemorrhage was of similar magnitude but did not reach statistical significance (OR: 1.70, 0.73-3.94).

Genetically determined higher BMI was significantly associated with increased risk of ischaemic (OR: 1.14, 1.06-1.22) and large-artery stroke (OR: 1.32, 1.12-1.57), and there was a suggestive association with cardioembolic stroke but it was not statistically significant (OR: 1.16, 1.03-1.31). Genetically higher WHR was significantly associated with increased risk of ischaemic stroke (OR: 1.19, 1.10-1.29), large-artery (OR: 1.45, 1.21-1.75) and small-vessel stroke (OR: 1.38, 1.16-1.66), but showed no association with cardioembolic stroke (OR: 1.00, 0.87-1.16). There were no material differences in the results when the instruments were restricted to SNPs that reached genome-wide significance for association with each trait (**Supplementary Table 3**).

There were no significant associations between sleep duration, physical activity, dietary components, coffee intake, or alcohol consumption with ischaemic stroke, intracerebral haemorrhage, or any ischaemic stroke subtype.

We calculated that the proportion of the total effect of education on stroke mediated through BMI and smoking was 1% for ischaemic stroke and nearly 14% for cardioembolic and large-artery stroke. However, the indirect effects of education on each of the five stroke subtypes were not statistically significant ($q < 0.05$), and for all stroke and small-vessel stroke there was inconsistent mediation because the indirect effects and total effects were in opposing directions, which may be due to insufficient power. Our mediation analysis therefore lacked sufficient evidence to state that the effects of education on ischaemic stroke and its subtypes are mediated through either BMI or smoking. This suggests that education, BMI, and smoking each have independent effects on stroke.

DISCUSSION

In this study, we conducted MR analyses using large-scale GWAS summary statistics to investigate causal associations of modifiable lifestyle traits with risk of stroke subtypes. We found genetic predisposition to number of years of education to be inversely associated with ischaemic stroke, large-artery and small-vessel stroke, and intracerebral haemorrhage. Genetic predisposition to ever smoking regularly and having higher BMI and WHR were associated with ischaemic and large-artery stroke. We also found that the associations of education, BMI, and smoking on risk of ischaemic stroke and its subtypes were independent of each other.

Our analysis did not identify any evidence of a causal association of lifestyle factors with cardioembolic stroke. This may reflect the fact that the majority of cases of CES are due to atrial fibrillation,²² which has a different genetic and other risk factor profile to other stroke subtypes. While further research into the underlying mechanisms is needed, this could help explain why we found that lifestyle factors are implicated in the onset of large-artery and small-vessel stroke but not CES.

Our MR results provide genetic evidence for an inverse causal effect of educational attainment on ischaemic stroke and intracerebral haemorrhage. Our analysis validates previously reported associations of the effect of education on ischaemic stroke from observational studies²³ and MR analyses,²⁴ as well as a recent MR analysis showing associations of education with large-artery stroke, small-vessel stroke, and intracerebral haemorrhage.²⁵ Our findings for the association of the lifetime smoking index with increased risk of ischaemic, large-artery, and small-vessel stroke is in concordance with previous observational studies²⁶ and MR analyses.²⁷ While earlier MR analyses of BMI with smaller sample sizes did not show clear evidence of a causal association with any ischaemic stroke subtype,²⁸ our analysis in a larger sample confirms

the associations reported in a recent MR analysis of BMI with large-artery stroke and WHR with large-artery and small-vessel stroke.²⁹

We did not identify any associations of stroke or its subtypes with genetic predisposition to longer sleep duration, increased physical activity, or increased consumption of fat, protein, carbohydrates, sugar, coffee, or alcohol. The number of genetic instruments was very small for physical activity and the dietary components (fewer than 10 SNPs each), which may have resulted in weak instrument bias. Additionally, analysing each dietary component in isolation may be less effective than considering the total effect of a healthy or unhealthy diet on stroke and its subtypes. Nevertheless, our research shows that sleep, physical activity, and several diet-related factors do not exhibit evidence of a causal association with ischaemic stroke and its subtypes or intracerebral haemorrhage. In contrast, genetic liability to insomnia has been shown to be associated with ischaemic stroke in a recent MR analysis,³⁰ so it is possible that overall sleep duration is not specific enough as a phenotype.

Our findings lend support to a number of potential policy recommendations. While some individuals may have a genetic predisposition to lower educational attainment, higher frequency of smoking, or higher BMI or WHR, genetic heritability only explains a fraction of the total variation in these traits (11% for education, 2% for smoking, 6% for BMI, and 3.9% for WHR).^{8,11,15,16} Therefore, regardless of their level of genetic risk, individuals can make various lifestyle changes to reduce their risk of stroke, such as smoking cessation.^{2,31,32} This study also highlights the importance of promoting universal education, which would benefit everyone regardless of genetic makeup.²⁵ Interestingly, although BMI and WHR were significantly associated with increased risk of stroke, we did not identify significant associations for related lifestyle changes that could be adopted to help reduce adiposity

measures, such as increased physical activity and reduced fat and sugar consumption. This may be due to the small number of genetic variants that were available to use as instrumental variables for physical activity and dietary components.

Our study has several strengths. The large sample size of each study (ranging from 235,000 to 1.1 million participants) provided substantial power to detect differences in the effects of lifestyle factors on aetiological stroke subtypes and to perform multiple sensitivity analyses to test the validity of the MR assumptions, which reduced the likelihood of obtaining biased estimates. In addition, our investigation of a wide range of lifestyle factors provides a relatively comprehensive overview of the associations with subtypes of ischaemic stroke and intracerebral haemorrhage. Moreover, several of the pitfalls common to observational studies, such as reverse causation bias and potential confounding, were reduced or avoided altogether by using an MR approach since genetic variants are allocated at conception. Our estimates therefore represent the impact of a lifelong intervention in the lifestyle factors rather than lifestyle changes of a shorter duration as reported in observational studies.

Our study also has several limitations. Under the MR analysis framework, we used genetic variants as instrumental variables to evaluate whether modifiable lifestyle factors have a causal association with risk of stroke and its subtypes; however, as we only had access to summary-level data rather than individual participant data, we were not able to calculate polygenic risk scores to determine whether individuals with specific genetic variants were at increased risk of stroke. A related point is that while we did capture the impact of a lifelong intervention in each lifestyle factor, our analyses only considered the genetic predisposition to adopt a certain lifestyle habit rather than determining whether individuals who actually implement a given change to their lifestyle have a measurable impact on their risk of stroke, which may also be

influenced by various environmental factors. Additionally, our analyses were based on datasets involving individuals of European ancestry and thus might not be applicable to other ethnicities. Furthermore, although the association of BMI with stroke is J-shaped,³³ our analyses did not account for non-linear associations for continuous traits.

In conclusion, our results suggest causal associations of lower educational attainment and higher levels of smoking, BMI, and WHR with increased risk of ischaemic stroke, particularly large-artery and small-vessel stroke. The effects of modifiable lifestyle factors on large-artery and small-vessel stroke observed here may have important policy implications that support the implementation of strategies to encourage improved lifestyle habits.

FUNDING

This work was supported by the British Heart Foundation [RG/16/4/32218]; the European Union's Horizon 2020 research and innovation programme [No 667375 (CoSTREAM) and No 666881 (SVDs@target)]; a grant for strategic collaboration between Ludwig-Maximilians-University Munich and the University of Cambridge; Cambridge University Hospitals National Institute for Health Research Biomedical Research Centre [*]; the Onassis Foundation [to M.K.G.]; the Deutsche Forschungsgemeinschaft [Munich Cluster for Systems Neurology (EXC 1010 SyNergy) and Collaborative Research Center 1123 (B3) to M.D.]; the Corona Foundation [to M.D.]; Fondation Leducq [Transatlantic Networks of Excellence Program on the Pathogenesis of Small Vessel Disease of the Brain to M.D.]; and the National Institute for Health Research [Senior Investigator Award to H.S.M.] [*]. The funding organisations had no role in any of the following: design and conduct of the study; collection, management, analysis, and interpretation of the data; review, or approval of the manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication. [*] The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

ACKNOWLEDGEMENTS

This research was conducted using the UK Biobank resource under application number 36509.

We acknowledge the contributions of the MEGASTROKE consortium, the International Stroke Genetics Consortium (ISGC) Cerebrovascular Disease Knowledge Portal, the Social Science Genetic Association Consortium (SSGAC), the Sleep Disorder Knowledge Portal (SDKP), and the Genetic Investigation of ANthropometric Traits (GIANT) consortium for making their datasets publicly available. The MEGASTROKE project received funding from sources specified at <http://megastroke.org/acknowledgements.html>, and the full investigator list of the MEGASTROKE consortium is available at <http://megastroke.org/authors.html>.

Author contributions: E.L.H., M.K.G., H.S.M., and M.D. conceived and designed the study. E.L.H. and M.K.G. analysed the data. All authors participated in interpretation of the results. E.L.H. and H.S.M. drafted the manuscript. All authors critically revised the manuscript for intellectual content and approved the final submitted version.

Conflicts of interests: None reported.

REFERENCES

1. Boehme AK, Esenwa C, Elkind MSV. Stroke risk factors, genetics, and prevention. *Circ Res* 2017;**120**:472–495.
2. Rutten-Jacobs LC, Larsson SC, Malik R, Rannikmäe K, MEGASTROKE consortium M, International Stroke Genetics Consortium ISG, Sudlow CL, Dichgans M, Markus HS, Traylor M. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. *BMJ* 2018;**363**:k4168.
3. Burgess S, Foley CN, Zuber V. Inferring causal relationships between risk factors and outcomes from genome-wide association study data. *Annu Rev Genomics Hum Genet* 2018;**19**:303–327.
4. Markus H. Stroke: causes and clinical features. *Medicine* 2016;**44**:515–520.
5. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese AK, Laan SW van der, Gretarsdottir S, Anderson CD, Chong M, Adams HHH, Ago T, Almgren P, Amouyel P, Ay H, Bartz TM, Benavente OR, Bevan S, Boncoraglio GB, Brown Jr. RD, Butterworth AS, Carrera C, Carty CL, Chasman DI, Chen WM, Cole JW, Correa A, Cotlarciuc I, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018;**50**:524–537.
6. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, Hartwig FP, Holmes M V, Minelli C, Relton CL, Theodoratou E. Guidelines for performing Mendelian randomization investigations [version 1; peer review: 1 approved]. *Wellcome Open Res* 2019;**4**.
7. Smith GD, Davies NM, Dimou N, Egger M, Gallo V, Golub R, Higgins JP, Langenberg C, Loder EW, Richards JB, Richmond RC, Skrivankova VW, Swanson SA, Timpson NJ, Tybjaerg-Hansen A, VanderWeele TJ, Woolf BA, Yarmolinsky J. STROBE-MR: Guidelines for strengthening the reporting of Mendelian randomization studies. *PeerJ Prepr* 2019;**7**:e27857v1.
8. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Karlsson Linnér R, Fontana MA, Kundu T, Lee C, Li H, Li R, Royer R, Timshel PN, Walters RK, Willoughby EA, Yengo L, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Elson SL, Fontanillas P, Hinds DA, McCreight. JC, Huber KE, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018;**50**:1112–1121.
9. Dashti HS, Jones SE, Wood AR, Lane JM, Hees VT van, Wang H, Rhodes JA, Song Y, Patel K, Anderson SG, Beaumont RN, Bechtold DA, Bowden J, Cade BE, Garaulet M, Kyle SD, Little MA, Loudon AS, Luik AI, Scheer FAJL, Spiegelhalter K, Tyrrell J, Gottlieb DJ, Tiemeier H, Ray DW, Purcell SM, Frayling TM, Redline S, Lawlor DA, Rutter MK, et al. Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat Commun* 2019;**10**:1100.
10. Klimentidis YC, Raichlen DA, Bea J, Garcia DO, Wineinger NE, Mandarino LJ, Alexander GE, Chen Z, Going SB. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. *Int J Obes* 2018;**42**:1161–1176.
11. Wootton RE, Richmond RC, Stuijzand BG, Lawn RB, Sallis HM, Taylor GMJ, Hemani G, Jones HJ, Zammit S, Davey Smith G, Munafò MR. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med* 2019;**1**–9.
12. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, Datta G, Davila-Velderrain J,

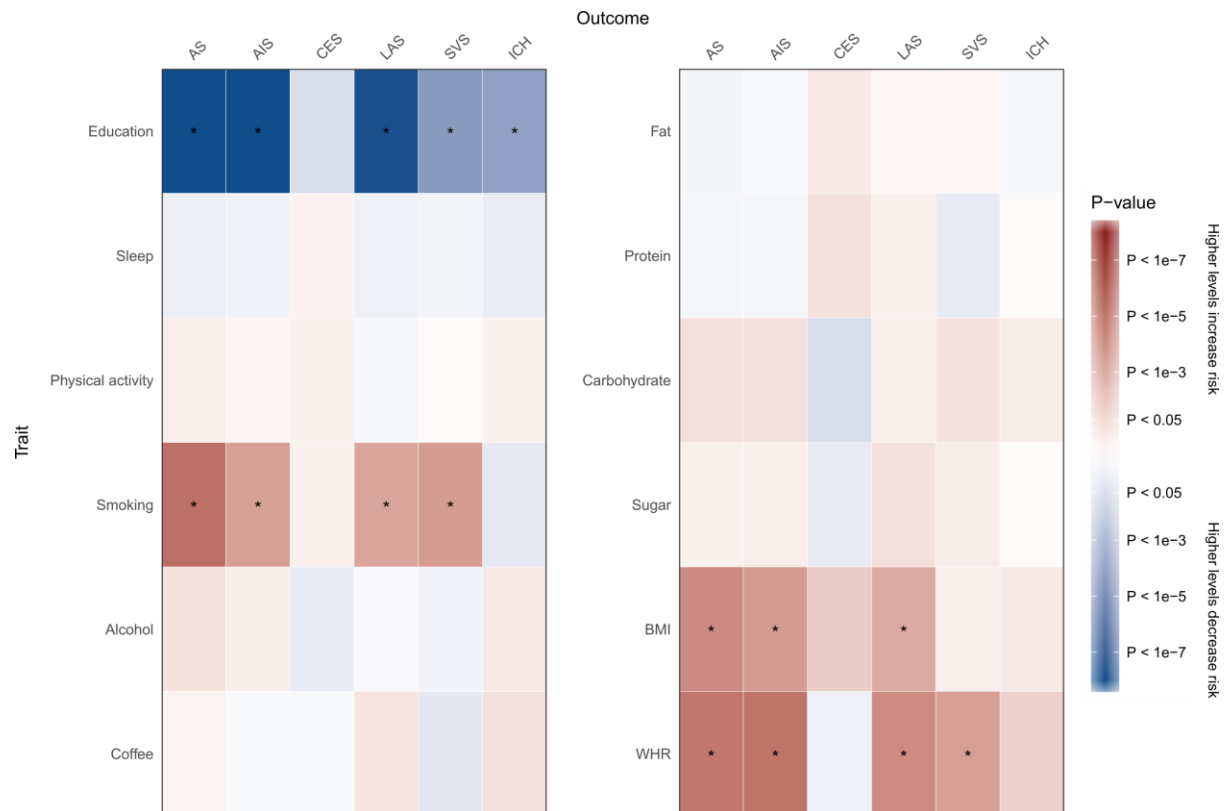
- McGuire D, Tian C, Zhan X, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Elson SL, Fontanillas P, Furlotte NA, Hinds DA, Hromatka BS, Huber KE, Kleinman A, Litterman NK, McIntyre MH, Mountain JL, Northover CAM, Sathirapongsasuti JF, Sazonova O V, Shelton JF, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019;**51**:237–244.
13. Zhong VW, Kuang A, Danning RD, Kraft P, Dam RM van, Chasman DI, Cornelis MC. A genome-wide association study of bitter and sweet beverage consumption. *Hum Mol Genet* 2019;**28**:2449–2457.
14. Meddens SFW, Vlaming R de, Bowers P, Burik CAP, Linnér RK, Lee C, Okbay A, Turley P, Rietveld CA, Fontana MA, Ghanbari M, Imamura F, McMahon G, Most PJ van der, Trudy V, Wade KH, Anderson EL, Braun KVE, Emmett PM, Esko T, Gonzalez JR, Kieft-de Jong JC, Luan J, Langenberg C, Muka T, Ring S, Rivadeneira F, Schoufour JD, Snieder H, Rooij FJA van, et al. Genomic analysis of diet composition finds novel loci and associations with health and lifestyle. *bioRxiv* 2018;**383406**.
15. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM, Consortium the G. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet* 2018;**27**:3641–3649.
16. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, Yengo L, Ferreira T, Marouli E, Ji Y, Yang J, Jones S, Beaumont R, Croteau-Chonka DC, Winkler TW, Hattersley AT, Loos RJF, Hirschhorn JN, Visscher PM, Frayling TM, Yaghootkar H, Lindgren CM. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* 2019;**28**:166–174.
17. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE. Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial. TOAST: Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993;**24**:35–41.
18. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, Anderson CD, Brouwers HB, Valant V, Battey TWK, Radmanesh F, Raffeld MR, Baedorf-Kassis S, Deka R, Woo JG, Martin LJ, Haverbusch M, Moomaw CJ, Sun G, Broderick JP, Flaherty ML, Martini SR, Kleindorfer DO, Kissela B, Comeau ME, Jagiella JM, Schmidt H, Freudenberger P, Pichler A, Enzinger C, et al. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am J Hum Genet* 2014;**94**:511–521.
19. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018;**7**:e34408.
20. Carter AR, Sanderson E, Hammerton G, Richmond RC, Smith GD, Heron J, Taylor AE, Davies NM, Howe LD. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. *bioRxiv* 2019;**835819**.
21. Thompson JR, Minelli C, Greco M F Del. Mendelian randomization using public data from genetic consortia. *Int J Biostat* 2016;**12**:20150074.
22. Lubitz SA, Parsons OE, Anderson CD, Benjamin EJ, Malik R, Weng L-C, Dichgans M, Sudlow CL, Rothwell PM, Rosand J, Ellinor PT, Markus HS, Traylor M. Atrial fibrillation genetic risk and ischemic stroke mechanisms. *Stroke* 2017;**48**:1451–1456.
23. Kroll ME, Green J, Beral V, Sudlow CLM, Brown A, Kirichek O, Price A, Yang TO, Reeves GK. Adiposity and ischemic and hemorrhagic stroke. *Neurology* 2016;**87**:1473–

- 1481.
24. Gill D, Efstathiadou A, Cawood K, Tzoulaki I, Dehghan A. Education protects against coronary heart disease and stroke independently of cognitive function: evidence from Mendelian randomization. *Int J Epidemiol* 2019;**48**:1468–1477.
25. Yuan S, Xiong Y, Michaelsson M, Michaelsson K, Larsson SC. Health related effects of education levels: a Mendelian randomization study. *medRxiv* 2020;**2020.02.01**.
26. Colpani V, Baena CP, Jaspers L, Dijk GM van, Farajzadegan Z, Dhana K, Tielemans MJ, Voortman T, Freak-Poli R, Veloso GGV, Chowdhury R, Kavousi M, Muka T, Franco OH. Lifestyle factors, cardiovascular disease and all-cause mortality in middle-aged and elderly women: a systematic review and meta-analysis. *Eur J Epidemiol* 2018;**33**:831–845.
27. Larsson SC, Burgess S, Michaëlsson K. Smoking and stroke: a Mendelian randomization study. *Ann Neurol* 2019;**86**:468–471.
28. Larsson SC, Scott RA, Traylor M, Langenberg CC, Hindy G, Melander O, Orho-Melander M, Seshadri S, Wareham NJ, Markus HS, METASTROKE Collaboration, NINDS Stroke Genetics Network (SiGN). Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study. *Neurology* 2017;**89**:454–460.
29. Marini S, Merino J, Montgomery BE, Malik R, Sudlow CL, Dichgans M, Florez JC, Rosand J, Gill D, Anderson CD. Mendelian randomization study of obesity and cerebrovascular disease. *Ann Neurol* 2020;
30. Larsson SC, Markus HS. Genetic liability to insomnia and cardiovascular disease risk. *Circulation* 2019;
31. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, Rangarajan S, Islam S, Pais P, McQueen MJ, Mondo C, Damasceno A, Lopez-Jaramillo P, Hankey GJ, Dans AL, Yusuf K, Truelsen T, Diener H-C, Sacco RL, Ryglewicz D, Czlonkowska A, Weimar C, Wang X, Yusuf S. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 2010;**376**:112–123.
32. O'Donnell MJ, Chin SL, Rangarajan S, Xavier D, Liu L, Zhang H, Rao-Melacini P, Zhang X, Pais P, Agapay S, Lopez-Jaramillo P, Damasceno A, Langhorne P, McQueen MJ, Rosengren A, Dehghan M, Hankey GJ, Dans AL, Elsayed A, Avezum A, Mondo C, Diener H-C, Ryglewicz D, Czlonkowska A, Pogosova N, Weimar C, Iqbal R, Diaz R, Yusuf K, Yusufali A, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. *Lancet* 2016;**388**:761–775.
33. Global BMI Mortality Collaboration, Angelantonio E Di, Bhupathiraju SN, Wormser D, Gao P, Kaptoge S, Berrington de Gonzalez A, Cairns BJ, Huxley R, Jackson CL, Joshy G, Lewington S, Manson JE, Murphy N, Patel A V, Samet JM, Woodward M, Zheng W, Zhou M, Bansal N, Barricarte A, Carter B, Cerhan JR, Smith GD, Fang X, Franco OH, Green J, Halsey J, Hildebrand JS, Jung KJ, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet* 2016;**388**:776–786.

Figures

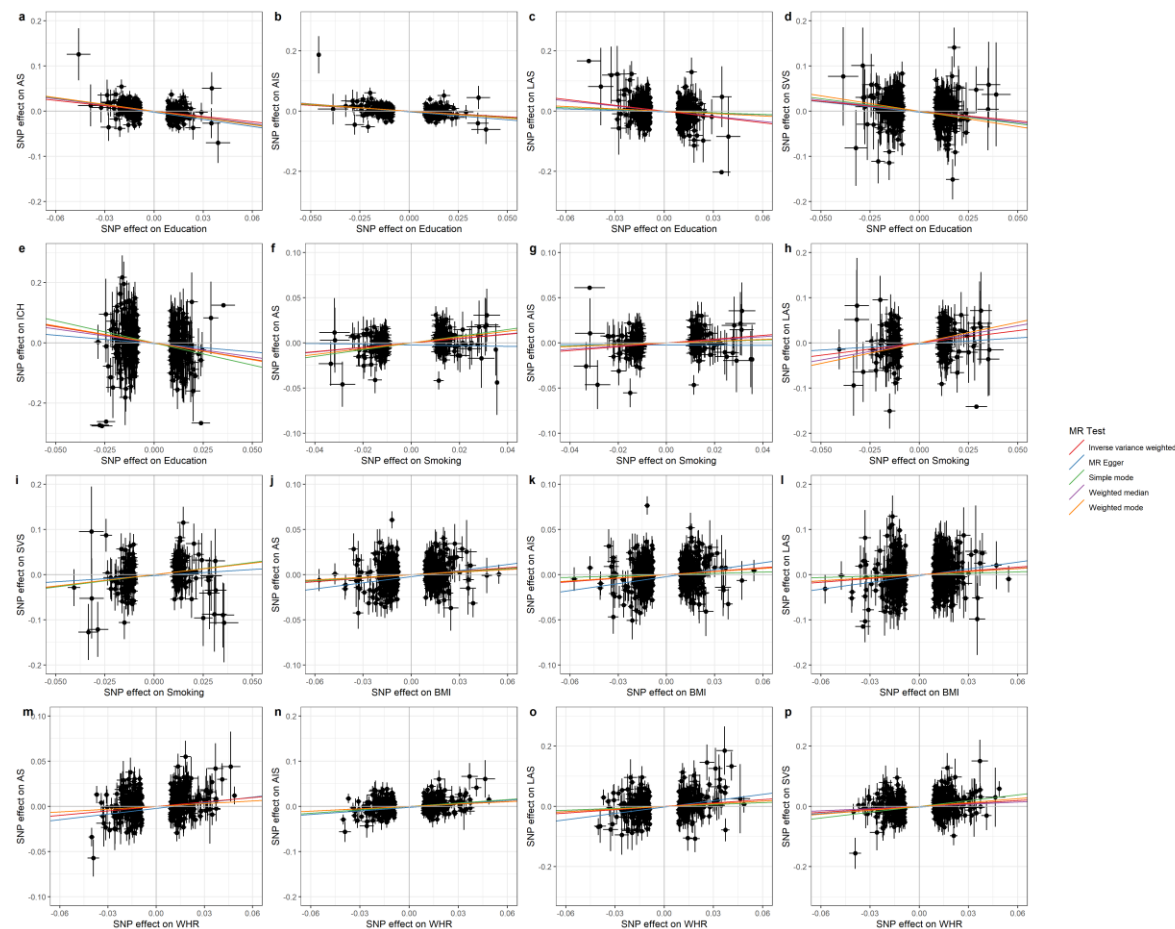
- Figure 1. Mendelian randomization results showing causal estimates for association of lifestyle traits with stroke and its subtypes
- Figure 2. Genetic associations of lifestyle traits and stroke subtypes for significant causal estimates
- Figure 3. Mendelian randomization associations between genetic predisposition to lifestyle factors and stroke subtypes

Figure 1. Mendelian randomization results showing causal estimates for association of lifestyle traits with stroke and its subtypes



Colours show magnitude and direction of P -value of association for estimate of causal effect using inverse-variance weighted MR approach. P -values less than 1×10^{-8} have been truncated in the figure. Asterisks indicate significant associations for the causal effect estimates (FDR $q < 0.05$). **AS** = All stroke; **AIS** = Any ischaemic stroke; **CES** = Cardioembolic stroke; **ICH** = Intracerebral haemorrhage; **LAS** = Large-artery stroke; **SVS** = Small vessel stroke. Please refer to **Supplementary Table 1** for a description of each trait.

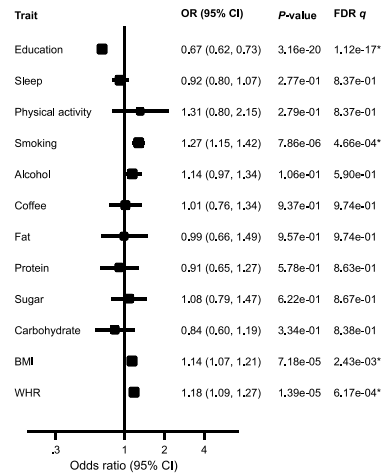
Figure 2. Genetic associations of lifestyle traits and stroke subtypes for significant causal estimates



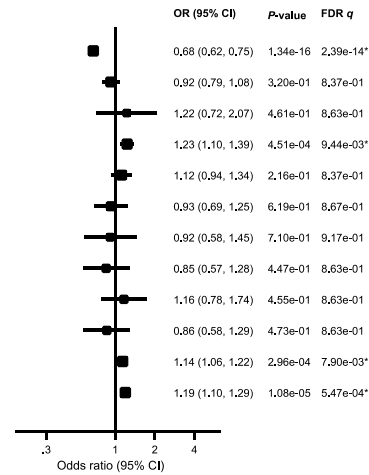
The associations of each genetic variant associated with lifestyle traits with significant ($FDR\ q < 0.05$) causal estimates are plotted against their association with the corresponding outcome. Circles represent the associated change in levels of the trait and corresponding increased risk for each variant. The horizontal and vertical lines through each circle represent the corresponding 95% confidence intervals for the genetic associations. Associations were oriented to the effect allele of each trait. Coloured lines show the slope (causal estimate) of the trait on the outcome obtained using a variety of different MR approaches. Trait / Outcome: (a) Education / AS; (b) Education / AIS; (c) Education / LAS; (d) Education / SVS; (e) Education / ICH; (f) Smoking / AS; (g) Smoking / AIS; (h) Smoking / LAS; (i) Smoking / SVS; (j) BMI / AS; (k) BMI / AIS; (l) BMI / LAS; (m) WHR / AS; (n) WHR / AIS; (o) WHR / LAS; (p) WHR / SVS. Please refer to **Supplementary Table 1** for a description of each trait.

Figure 3. Mendelian randomization associations between genetic predisposition to lifestyle factors and stroke subtypes

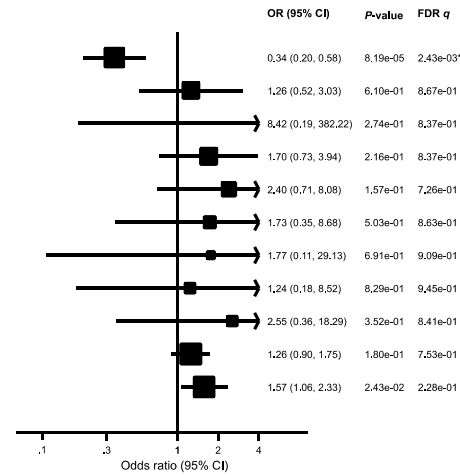
(a) AS



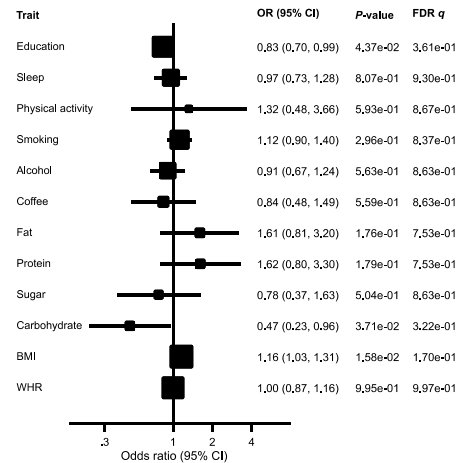
(b) AIS



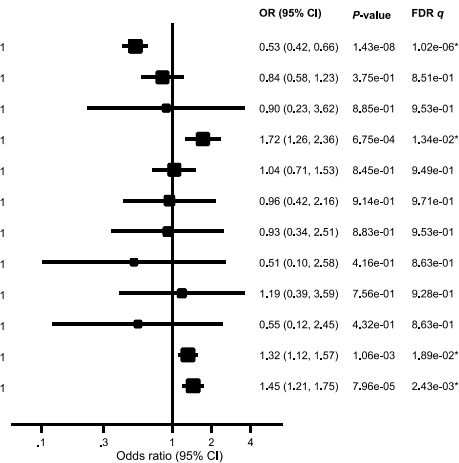
(c) ICH



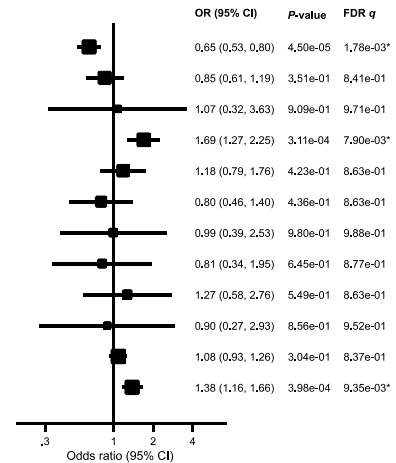
(d) CES



(e) LAS



(f) SVS



Results derived from random-effects inverse-variance weighted MR analyses. Asterisks indicate significant associations for the causal effect estimates (FDR $q < 0.05$).

Results are shown for (a) all stroke, (b) any ischaemic stroke, (c) intracerebral haemorrhage, and ischaemic stroke subtypes: (d) cardioembolic stroke, (e) large-artery stroke, and (f) small vessel stroke. Please refer to **Supplementary Table 1** for a description of each trait.