



OPEN ACCESS



Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants

Loes CA Rutten-Jacobs,^{1,2} Susanna C Larsson,³ Rainer Malik,⁴ Kristiina Rannikmäe,^{5,6} Cathie L Sudlow,^{5,6,7} Martin Dichgans,^{4,8,9} Hugh S Markus,² Matthew Traylor²

For numbered affiliations see end of article.

Correspondence to: LCA Rutten-Jacobs Loes.Rutten-Jacobs@dzne.de (@ruttenjacobs on Twitter; ORCID 0000-0003-3223-885X)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2018;363:k4168 <http://dx.doi.org/10.1136/bmj.k4168>

Accepted: 17 September 2018

ABSTRACT

OBJECTIVE

To evaluate the associations of a polygenic risk score and healthy lifestyle with incident stroke.

DESIGN

Prospective population based cohort study.

SETTING

UK Biobank Study, UK.

PARTICIPANTS

306 473 men and women, aged 40-73 years, recruited between 2006 and 2010.

MAIN OUTCOME MEASURE

Hazard ratios for a first stroke, estimated using Cox regression. A polygenic risk score of 90 single nucleotide polymorphisms previously associated with stroke was constructed at $P < 1 \times 10^{-5}$ to test for an association with incident stroke. Adherence to a healthy lifestyle was determined on the basis of four factors: non-smoker, healthy diet, body mass index $< 30 \text{ kg/m}^2$, and regular physical exercise.

RESULTS

During a median follow-up of 7.1 years (2 138 443 person years), 2077 incident strokes (1541 ischaemic stroke, 287 intracerebral haemorrhage, and 249 subarachnoid haemorrhage) were ascertained. The risk of incident stroke was 35% higher among those at high genetic risk (top third of polygenic score) compared with those at low genetic risk (bottom third): hazard ratio 1.35 (95% confidence interval 1.21 to 1.50), $P = 3.9 \times 10^{-8}$. Unfavourable lifestyle (0 or 1 healthy lifestyle factors) was associated with a 66% increased risk of stroke compared with a favourable lifestyle (3 or 4 healthy lifestyle factors): 1.66 (1.45 to 1.89), $P = 1.19 \times 10^{-13}$. The association with lifestyle was independent of genetic risk strata.

CONCLUSION

In this cohort study, genetic and lifestyle factors were independently associated with incident stroke. These results emphasise the benefit of entire populations adhering to a healthy lifestyle, independent of genetic risk.

Introduction

Stroke is one of the leading reasons for disability and death worldwide.¹ It is a complex disease, caused by both genetic and environmental factors, including diet and lifestyle.²

Early evidence supporting a role for genetics in risk of stroke came from twin studies and family history studies.^{3,4} Further evidence has emerged from genome wide association studies. MEGASTROKE, a large meta-analysis of genome wide association studies tripled the number of loci robustly associated with stroke risk.⁵

Lifestyle is an important modifiable risk factor for stroke. Clear evidence shows that adhering to a healthy lifestyle, including not smoking, reducing the risk of diabetes, regular physical activity, and a healthy diet, decreases the risk of stroke substantially.^{6,7} It might be hypothesised that adhering to a healthy lifestyle could attenuate the effect of genetics on stroke risk. A previous study in coronary artery disease—a condition closely related to ischaemic stroke, found a statistically significant interplay between genetic and lifestyle risk factors in the risk of coronary artery disease.⁸

We investigated whether a weighted genetic risk score based on the genome wide association results for stroke in MEGASTROKE is associated with incident stroke in a large population based cohort (UK Biobank). We also investigated whether adherence to a healthy lifestyle influences this association.

Methods

Study population

UK Biobank is a prospective study that recruited 500 000 community dwelling participants, aged 40-69 years, from across the United Kingdom between 2006 and 2010 (www.ukbiobank.ac.uk).⁹ Participants attended one of 22 assessment centres across England, Scotland, and Wales. The study collects extensive data from questionnaires, interviews, health records, physical measures, biological samples, and imaging. Main outcomes and exposures of interest in the current study include imputed genetic data, incident stroke, and lifestyle (smoking, diet, body mass index, and physical activity).

WHAT IS ALREADY KNOWN ON THIS TOPIC

Stroke is a complex disease caused by both genetic and environmental factors, including diet and lifestyle

Whether adhering to a healthy lifestyle could attenuate the effect of genetic background on risk of incident stroke is currently unknown

WHAT THIS STUDY ADDS

Genetic and lifestyle factors were independently associated with risk of incident stroke

An unfavourable lifestyle profile was associated with increased risk of stroke across all genetic risk strata

These findings highlight the potential of lifestyle interventions to reduce risk of stroke across entire populations, even in those at high genetic risk of stroke

In the present study we included all people who were classified as white British (all who self identified as white British, followed by the exclusion of ethnic outliers identified by principal components analysis on the genotype data), without a history of stroke or myocardial infarction on the basis of self report or medical records, or both, and with complete data on lifestyle.

Healthy lifestyle factors

We defined four healthy lifestyle factors on the basis of the American Heart Association guidelines: no current smoking, healthy diet, body mass index $<30 \text{ kg/m}^2$, and moderate physical activity two or more times weekly.¹⁰

The UK Biobank participants completed a questionnaire on their usual diet pattern. In this analysis, a healthy diet was determined according to the increased consumption of fruit, vegetables, and fish and the decreased consumption of processed meats and red meats. We defined a healthy diet as adherence to at least two of the healthy food items. Supplemental text S1 and table S1 provide additional details on the specific questions asked and the construction of a healthy diet score.

Moderate physical activity was defined as at least 150 minutes of moderate intensity activity weekly or 75 minutes of vigorous activity weekly.

Incident stroke

Incident stroke in UK Biobank was based on medical history and linkage to data on hospital admissions and mortality. We used the stroke variables provided by UK Biobank, which were created by combining information from these different data sources (see supplemental table S2). Details of the algorithms used to combine the data from different sources to identify stroke have been described previously and are available on the UK Biobank website (www.ukbiobank.ac.uk). We subtyped stroke as ischaemic stroke, intracerebral haemorrhage, or subarachnoid haemorrhage.

We took into account the occurrence of myocardial infarction during the study as this could potentially result in lifestyle changes during follow-up that affect the risk of stroke. The occurrence of myocardial infarction was defined according to the UK Biobank algorithmic definition (see supplemental table S2). Details of the myocardial infarction algorithm have been described previously and are available on the UK Biobank website (www.ukbiobank.ac.uk).¹¹

We excluded people from the analysis who self reported stroke or myocardial infarction.

Genetic data

We used the June 2017 release of the imputed genetic data from UK Biobank (downloaded 3 June 2017). Details of the design of the arrays, sample processing, and stringent quality control have been described in detail elsewhere¹² and summarised previously.¹³ Briefly, we used two closely related arrays from Affymetrix, the UK BiLEVE Axiom array (9.9% of

people) and the UK Biobank Axiom array, to genotype about 805 426 markers with good genome wide coverage. Phasing was performed using SHAPEIT3 and imputation using IMPUTE4.^{12 15} Two reference panels were used for imputation; the Haplotype Reference Consortium reference panel (39 131 578 autosomal single nucleotide polymorphisms, SNPs) and a merged UK10K and 1000 Genomes Phase 3 panel.¹⁴ Imputed genotypes were available for 488 369 participants in this study.¹² From the resulting dataset, we excluded those who self reported ancestry other than white British, related people (second degree or greater: kinship coefficient ≥ 0.884), people with high levels of heterozygosity and missingness ($>5\%$), and people whose reported sex was inconsistent with sex inferred from the genetic data. The UK Biobank core team centrally performed a check for excessive heterozygosity.¹³ Extreme heterozygosity or high rates of missingness, or both, can be indicators of poor sample quality due to, for example, DNA contamination. UK Biobank provided a list of samples with unusually high heterozygosity and we excluded those samples according to its recommendations. To evaluate a mismatch in sex self reported sex was compared with sex inferred from the genetic data (based on relative intensity of markers on the Y and X chromosomes). This sex mismatch evaluation was centrally performed by the UK Biobank core team and is described in detail elsewhere.¹² This evaluation can be used as a way to detect sample mishandling or other kinds of clerical error. However, in a dataset of this size, some such mismatches would be expected owing to transgender people or instances of real (but rare) genetic variation, such as aneuploidies in sex chromosomes.

In this analysis we only included SNPs imputed from the Haplotype Reference Consortium panel.

Polygenic risk score derived from MEGASTROKE

We derived three sets of independent ($r^2 < 0.05$ or 1000 Kb apart) SNPs associated with any stroke in people of European ancestry in MEGASTROKE (see supplemental text S2) at $P < 5 \times 10^{-8}$, $P < 1 \times 10^{-6}$, and $P < 1 \times 10^{-5}$ using an LD clumping procedure employed using plink v1.90b3.45.^{5 16} For each individual in the UK Biobank sample we calculated quantitative aggregate risk scores, defined as the sum of the number of risk alleles present at each locus weighted by the log of the odds ratio for that locus estimated from the MEGASTROKE sample using the plink “-score” command.

The three polygenic scores were tested for an association with incident any stroke, and for further analyses we used the polygenic risk score most statistically significantly associated with incident stroke. Supplemental table S3 lists the SNPs included in the MEGASTROKE risk score (all stroke, $P < 1 \times 10^{-5}$).

We repeated the previous steps while restricting to ischaemic stroke in those of European ancestry to create a genetic risk score for ischaemic stroke ($P < 1 \times 10^{-5}$).

Statistical analysis

We defined genetic risk in thirds: “low risk” (lowest third of genetic risk score), “intermediate risk” (second third), “high risk” (highest third). Lifestyle was recorded as “favourable” (three or four healthy lifestyle factors), “intermediate” (two healthy lifestyle factors), “unfavourable” (no or one healthy lifestyle factor).

To test the association of genetic and lifestyle factors with incident stroke we used Cox proportional hazards models. The duration of follow-up was calculated as time between the baseline assessment and the first event of either stroke, myocardial infarction, death, or 1 March 2016, which was the end of follow-up for the current data release. Participants who had a myocardial infarction or died before a stroke occurred were censored at the time of the respective event.

We repeated the previous steps for the ischaemic stroke genetic risk score and assessed the association between this score and incident ischaemic stroke. In this analysis we censored participants with a diagnosis of intracerebral haemorrhage, subarachnoid haemorrhage, or myocardial infarction or who died before an ischaemic stroke occurred at the time of the respective event.

Cox proportional hazards models included adjustment for age and sex for the lifestyle score models. For the models including the genetic score we additionally adjusted for the first 10 principal components of ancestry and genotyping batch. Model discrimination was evaluated with the concordance (*c*) statistic.

We included an interaction term in the regression model to test for statistical interaction between the lifestyle and genetic risk score.

To obtain cumulative incidence for lifestyle and genetic risk strata we used competing risk regression; the cumulative incidence function. We compared the hazard ratios for the genetic and lifestyle score in the risk of stroke derived from Cox proportional hazards models with the subdistribution hazard ratios, calculated using proportional subdistribution hazards regression models.¹⁷

R software version 3.4.2 was used for the Cox proportional hazards regression (package “survival”) and proportional subdistribution hazards regression (package “cmprsk”).

Patient and public involvement

The development of the research question or outcome measures was not informed by patients’ priorities, experience, or preferences. No patients were involved in the design and conduct of the present study. There are no plans to disseminate the results to study participants.

Results

Complete data for the present analysis were available for 306 473 participants in the UK Biobank Study (fig 1). Table 1 shows the baseline characteristics of the study population. During a total of 2 138 443 person years (median follow-up 7.1 years), 2077 incident

fatal or non-fatal strokes were reported as first incident vascular event or death, of which 1541 were ischaemic stroke, 287 intracerebral haemorrhage, and 249 subarachnoid haemorrhage. Furthermore, 3436 cases of fatal or non-fatal myocardial infarction and 6646 deaths due to other causes than stroke or myocardial infarction were reported as first incident vascular event or death.

Polygenic risk scores containing independent SNPs (on basis of linkage disequilibrium patterns) derived from MEGASTROKE at three different significance thresholds were tested for association with incident stroke in UK Biobank (see supplemental figure S1). The three polygenic risk scores were associated with risk of incident stroke, but the genetic risk score including all SNPs associated with stroke in MEGASTROKE at $P < 1 \times 10^{-5}$ (90 SNPs, see supplemental table S3) showed the strongest association and was therefore selected for subsequent analyses. The polygenic risk approximated a normal distribution (see supplemental figure S2).

In Cox proportional hazards analysis, the risk of incident stroke was higher for those with intermediate (hazard ratio 1.20, 95% confidence interval 1.08 to 1.34) and high genetic risk scores (1.35, 1.21 to 1.50) compared with those with a low genetic risk score (fig 2). We tested available cardiometabolic risk factors for an association with the genetic risk score, adjusting for the first 10 principal components of ancestry, genotyping batch, age, and sex. The genetic risk score was significantly associated with systolic blood pressure ($P = 1.5 \times 10^{-15}$), diastolic blood pressure ($P = 1.1 \times 10^{-7}$), use of lipid lowering drugs ($P = 7.5 \times 10^{-13}$), and diabetes (7.6×10^{-4}), but not with body mass index ($P = 0.18$).

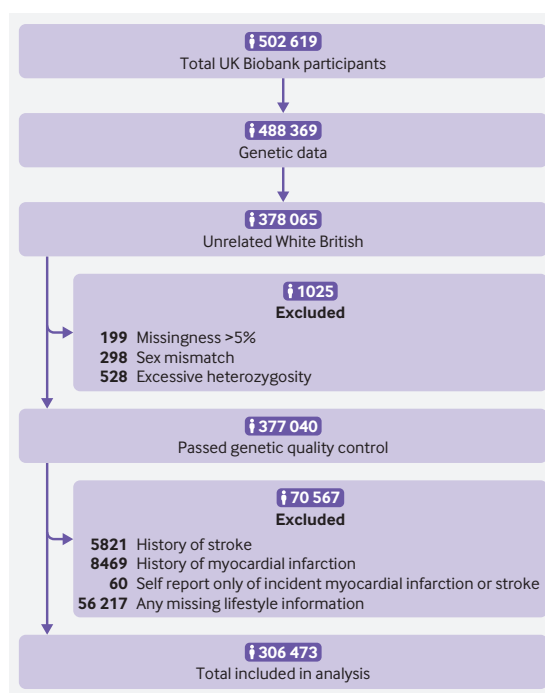


Fig 1 | Flow of participants through study

Table 1 | Characteristics of participants at baseline. Values are numbers (participants) unless stated otherwise

| Characteristic | All (n=306 473) | Incident stroke (n=2077) | No stroke (n=304 396) |
|--|-----------------|--------------------------|-----------------------|
| Mean (SD) age (years) | 56.7 (7.9) | 61.2 (6.8) | 56.6 (7.9) |
| Men | 136 654 (44.6) | 1184 (57.0) | 135 470 (44.5) |
| Mean (SD) systolic blood pressure (mm Hg) | 138 (18) | 146 (21) | 138 (19) |
| Mean (SD) diastolic blood pressure (mm Hg) | 82 (10) | 85 (11) | 82 (10) |
| Mean (SD) body mass index | 27.2 (4.7) | 28.0 (4.8) | 27.2 (4.7) |
| Diabetes | 12 927 (4.2) | 209 (10.1) | 1865 (4.2) |
| Use of lipid lowering drugs | 44 785 (14.7) | 497 (24.1) | 44 288 (14.6) |
| Use of antihypertensives | 61 218 (20.0) | 697 (33.6) | 60 521 (19.9) |
| Healthy lifestyle factors: | | | |
| No current smoking | 286 352 (93.4) | 1822 (87.7) | 284 530 (93.5) |
| Body mass index <30 | 236 326 (77.1) | 1489 (71.7) | 588 (77.1) |
| Regular moderate physical activity | 181 234 (59.1) | 1203 (57.9) | 180 031 (59.1) |
| Healthy diet | 139 328 (45.5) | 884 (42.6) | 138 444 (45.5) |
| Healthy lifestyle score: | | | |
| Favourable (3 or 4 healthy lifestyle factors) | 191 003 (62.3) | 1157 (55.7) | 189 846 (62.4) |
| Intermediate (2 healthy lifestyle factors) | 86 710 (28.3) | 652 (31.4) | 86 058 (28.3) |
| Unfavourable (0 or 1 healthy lifestyle factor) | 28 760 (9.4) | 268 (12.9) | 28 492 (9.4) |
| Genetic risk category: | | | |
| Low | 101 977 (33.3) | 589 (28.4) | 101 388 (33.3) |
| Intermediate | 102 300 (33.4) | 703 (33.8) | 101 597 (33.4) |
| High | 102 196 (33.3) | 785 (37.8) | 101 411 (33.3) |

The cumulative risk of stroke increased with age. Blood pressure, diabetes, and the use of lipid lowering drugs did not seem to influence the association between genetic risk of stroke and incident stroke (see supplemental table S4). Similarly, the risk of stroke was increased in those with an unfavourable (hazard ratio 1.66, 95% confidence interval 1.45 to 1.89) and intermediate (1.27, 1.16 to 1.40) lifestyle compared with those with a favourable lifestyle (fig 2).

Supplemental figure S3 shows the distribution of thirds of genetic risk and lifestyle scores. The genetic risk score was not associated with any of the single healthy lifestyle factors: odds ratio 0.98 (95% confidence interval 0.93 to 1.04) for body mass index score, 1.04 (0.99 to 1.09) for diet score, 0.97 (0.88 to 1.07) for smoking score, and 0.98 (0.94 to 1.03) for exercise score. Furthermore, the association of genetic risk with incident stroke was unchanged after adjustment for lifestyle (see supplemental table S5). Likewise, the association of lifestyle with incident stroke was essentially unchanged after adjustment for the genetic risk score (see supplemental table S5).

Table 2 shows the risk of incident stroke for combined genetic risk and lifestyle profiles. An additive effect was found for genetic risk and lifestyle on risk of incident stroke. Within each genetic risk stratum there was an increase in strength of association with decreasing number of favourable life style factors (see supplemental table S6). The highest risk of incident stroke was observed in participants with a high genetic risk and an unfavourable lifestyle: hazard ratio 2.30 (95% confidence interval 1.84 to 2.87); see supplemental figure S4. The test for statistical interaction between lifestyle score and genetic risk score in relation to incident stroke was not significant (P=0.57) compared with participants with low genetic risk and favourable lifestyle.

Model discrimination was similar between the main Cox proportional hazards models including the

genetic risk score (c statistic 0.69 (SE) 0.01), lifestyle score (0.69 (SE 0.01)), and the combined genetic and lifestyle score (0.70 (SE 0.01)).

We repeated the analysis to test the associations of the genetic and lifestyle scores with incident stroke while restricting only to ischaemic stroke for both genetic risk score and outcome. In addition, we compared the results derived from the Cox proportional hazards model with those derived from the competing risk proportional subdistribution hazards model. Results did not change substantially when restricting to ischaemic stroke or when using the subdistribution hazards model compared with the original analyses (see supplementary tables S7-S9).

Among individual components of the lifestyle score, smoking and body mass index ≥ 30 kg/m² contributed most to the risk of incident stroke (table 3). For all lifestyle factors, the effects were similar across genetic risk stratum.

As the effect of smoking was about twice as strong as the other individual lifestyle scores (table 3). We repeated the analysis of the risk of incident stroke for combined genetic risk and lifestyle profile in which smoking was counted twice, and this resulted in slightly increased point estimates in the unfavourable versus favourable lifestyle categories (see supplementary table S10).

The associations of genetic risk score with incident stroke were consistent in men and women (interaction P=0.70, supplemental figure S5). However, across all genetic risk stratum the absolute risk of incident stroke was lower in women than in men.

A statistically significant interaction on the multiplicative scale was found between sex and lifestyle profile in the risk of incident stroke (interaction P=0.01, supplemental figure S6). For men there was an increase in association with decreasing number of healthy lifestyle factors: hazard ratio 1.20 (95% confidence interval 1.05 to 1.36) and 1.82 (1.55

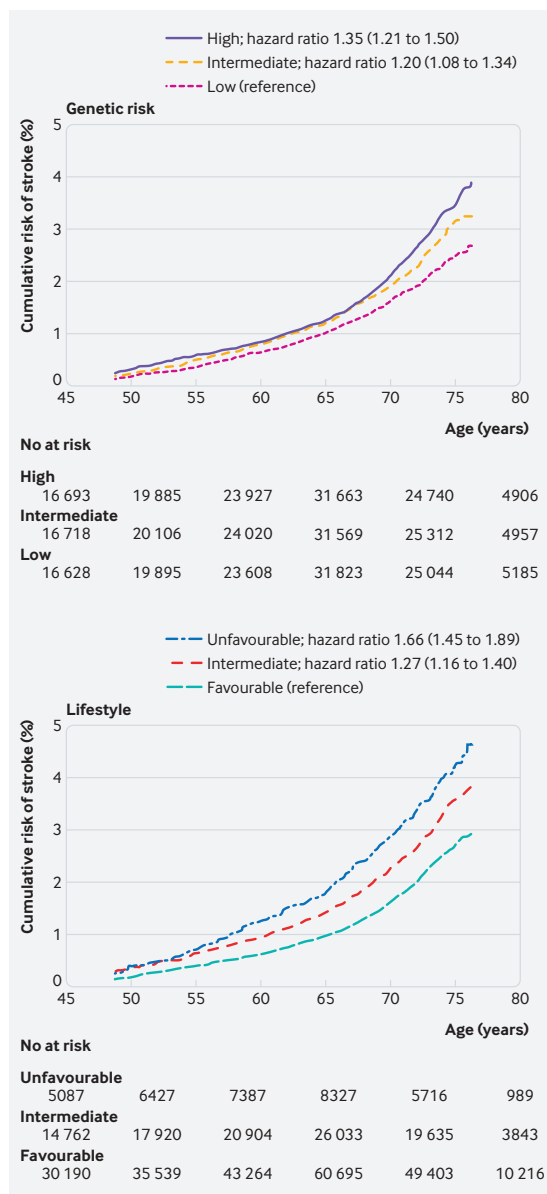


Fig 2 | Standardised risk of incident stroke according to genetic risk and lifestyle profile. Cox proportional hazards models were adjusted for age and sex, and the genetic risk models included additionally the first 10 principal components of ancestry and genotyping batch

to 2.15) for intermediate and unfavourable lifestyle versus favourable lifestyle, respectively. However, among women there was no difference between intermediate and unfavourable lifestyle versus favourable lifestyle: 1.39 (1.21 to 1.61) and 1.36 (1.08 to 1.72), respectively.

Discussion

We investigated the association between genetic risk of stroke, lifestyle, and incident risk of stroke in 306 473 people within the population based UK Biobank study. Risk of incident stroke was 35% higher among those at high genetic risk compared with those at low genetic risk, and these associations were independent of lifestyle profile. Furthermore, an unfavourable lifestyle was associated with a 66% increased risk of incident stroke compared with a favourable lifestyle, and this increased risk was present within any genetic risk category. A high genetic risk combined with an unfavourable lifestyle profile was associated with a more than twofold increased risk of stroke compared with a low genetic risk and a favourable lifestyle.

The present study provides further support that common genetic variants are implicated in the development of stroke. Our findings showing that a polygenic risk score is associated with incident stroke is in line with both clinical and population based studies.¹⁸⁻²¹

The genetic risk score was also associated with blood pressure and use of lipid lowering drugs, which suggest that the effect of the genetic variants on risk of incident stroke might at least in part be mediated by vascular risk factors. However, adjusting for those factors did not change the effect size of association between genetic risk and incident stroke, which emphasises that other mechanisms than those that involve the traditional cardiovascular risk factors are likely important. In the MEGASTROKE genome wide association analysis of stroke, only about half of the identified loci shared genetic variation with related vascular traits, including blood pressure and lipid levels, which support that the genetic risk might act through additional mechanisms.⁵

The reduction of stroke risk by adhering to a healthy lifestyle has been well reported.^{7 22-25} The

Table 2 | Relative and absolute risk of incident stroke according to genetic and lifestyle profiles

| Genetic risk | Lifestyle | | |
|---|--|--|--|
| | Favourable | Intermediate | Unfavourable |
| Low | | | |
| Hazard ratio* (95% CI) | 1 (reference) | 1.36 (1.14 to 1.63), P=7.3×10 ⁻⁰⁴ | 1.84 (1.44 to 2.35), P=8.0×10 ⁻⁰⁷ |
| 8 year cumulative incidencet (%) (95% CI) | 0.54 (0.47 to 0.60) | 0.74 (0.63 to 0.85) | 0.95 (0.74 to 1.17) |
| Intermediate | | | |
| Hazard ratio* (95% CI) | 1.26 (1.09 to 1.46), P=0.002 | 1.62 (1.37 to 1.92), P=3.2×10 ⁻⁰⁸ | 1.85 (1.46 to 2.37), P=5.4×10 ⁻⁰⁷ |
| 8 year cumulative incidencet (%) (95% CI) | 0.67 (0.60 to 0.74) | 0.82 (0.71 to 0.93) | 0.92 (0.72 to 1.12) |
| High | | | |
| Hazard ratio* (95% CI) | 1.44 (1.25 to 1.66), P=7.0×10 ⁻⁰⁷ | 1.70 (1.44 to 2.01), P=8.1×10 ⁻¹⁰ | 2.30 (1.84 to 2.87), P=3.3×10 ⁻¹³ |
| 8 year cumulative incidencet (%) (95% CI) | 0.78 (0.70 to 0.86) | 0.91 (0.78 to 1.04) | 1.11 (0.89 to 1.33) |

*Calculated using Cox proportional hazards models, adjusted for age, sex, first 10 principal components of ancestry, and genotyping batch.

†Calculated using the cumulative incidence function as implemented in the “cmprsk” R package.

Table 3 | Multivariable Cox regression analysis of age, sex, and lifestyle factors in relation to risk of stroke, stratified by genetic risk

| Lifestyle factors | Low genetic risk | | Moderate genetic risk | | High genetic risk | |
|---------------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | Hazard ratio (95% CI) | P value | Hazard ratio (95% CI) | P value | Hazard ratio (95% CI) | P value |
| Age, per year | 1.08 (1.07 to 1.09) | <2×10 ⁻¹⁶ | 1.09 (1.08 to 1.10) | <2×10 ⁻¹⁶ | 1.11 (1.10 to 1.12) | <2×10 ⁻¹⁶ |
| Male sex | 1.47 (1.25 to 1.74) | 4.6×10 ⁻⁰⁶ | 1.65 (1.42 to 1.93) | 1.2×10 ⁻¹⁰ | 1.56 (1.35 to 1.80) | 1.4×10 ⁻⁰⁹ |
| Current smoking | 2.35 (1.84 to 3.01) | 7.3×10 ⁻¹² | 2.81 (2.27 to 3.48) | <2×10 ⁻¹⁶ | 1.87 (1.48 to 2.37) | 1.5×10 ⁻⁰⁷ |
| Body mass index ≥30 | 1.43 (1.20 to 1.71) | 8.0×10 ⁻⁰⁵ | 1.38 (1.17 to 1.63) | 1.2×10 ⁻⁰⁴ | 1.19 (1.02 to 1.40) | 0.03 |
| No regular moderate physical activity | 1.09 (0.92 to 1.28) | 0.32 | 0.99 (0.85 to 1.63) | 0.91 | 1.09 (0.94 to 1.26) | 0.24 |
| Unhealthy diet | 1.10 (0.93 to 1.30) | 0.27 | 1.11 (0.77 to 1.05) | 0.16 | 1.21 (1.05 to 1.40) | 0.01 |

risk reduction associated with adherence to a healthy lifestyle in the present study was similar across all strata of genetic risk, which emphasises the benefit for entire populations of adhering to a healthy lifestyle, independent of genetic risk. Among the lifestyle factors, the most statistically significant associations were observed for smoking and body mass index ≥30 kg/m².

Comparison with previous studies

Across all categories of genetic risk and lifestyle, the risk of incident stroke was higher in men than women. This is an expected finding given the previously consistently shown higher incidence of stroke in men compared with women at the age of most of the UK Biobank participants.²⁶ Our results suggested that the relative risk of incident stroke associated with high genetic risk versus low genetic risk was similar in men and women. A family history and genome wide association studies suggested that genetic susceptibility to stroke is somewhat stronger in women than in men.^{27,28} The methodological differences of those previous studies and the current study might explain the different conclusions. In the current study, only genetic variants associated with stroke in MEGASTROKE at $P < 1 \times 10^{-5}$ were considered, whereas the other studies evaluated all genome wide variants within the study population or family history, which also includes environmental effects.

Considerable evidence from previous epidemiological studies also suggests differences in risk factors that are associated with stroke in men compared with women. Women have a higher prevalence of hypertension, whereas men have a higher prevalence of heart disease, diabetes, and unhealthy lifestyle behaviours, including smoking, obesity, and alcohol use.²⁹⁻³¹ In the present study we found a higher relative risk of stroke associated with an unfavourable lifestyle in men than women (82% v 36% increased risk, respectively).

Other possibilities for the increased relative risk include potential differences in duration of exposure to unfavourable lifestyle factors. Future studies are needed to evaluate the effect of the duration of exposure to an unfavourable lifestyle profile on the risk of stroke.

Strengths and limitations of this study

The major strengths of the current study include the large sample size of UK Biobank participants, which enabled study of the combination of genetic risk and

lifestyle in detail. Furthermore, to derive a genetic risk score for stroke, we used MEGASTROKE, which is currently the largest genome wide association study of stroke.⁵ Another distinctive feature of this analysis compared with a previous study is that we also included single nucleotide polymorphisms (SNPs) associated with stroke at a subthreshold level of significance ($P < 1 \times 10^{-5}$).¹⁹ This concurs with a previous study examining the predictive utility of genetic risk scores for incident coronary heart disease, which showed that the best performance was achieved by including SNPs that did not necessarily reach the genome wide statistical significance threshold in previous genome wide association studies.³²

Our study has several limitations. Firstly, behavioural changes before or after the baseline examinations might have had an effect on the risk estimates. We tried to reduce the effect of behavioural changes that could be related to vascular disease by excluding all those with a history of stroke and by censoring those at the time a myocardial infarction occurred. Secondly, this analysis focused on a narrow range of lifestyle factors, based on the American Heart Association guidelines.¹⁰ Expanding the range of lifestyle factors (ie, stress, sleep, alcohol and drug use) and more detailed assessment of diet and physical activity would be of interest for future studies. Thirdly, in the current study we only evaluated stroke of any cause. The effects of lifestyle and genetic variants might differ according to the cause of stroke, although some genetic risk variants and vascular risk factors are shared between two or more causal factors.⁵

Finally, in the present study we restricted our analysis to people of European descent. Therefore, our results may not be generalisable to populations with distinct ancestry. Future studies are needed that test these relations in more diverse populations.

Conclusion

In the present prospective population based cohort study of 306 473 people we found that genetic and lifestyle factors were independently associated with risk of incident stroke. These findings highlight the potential of lifestyle interventions to reduce risk of stroke across entire populations, even in those at high genetic risk of stroke.

AUTHOR AFFILIATIONS

¹German Center for Neurodegenerative diseases (DZNE), Population Health Sciences, Sigmund-Freud-Strasse 27, 53127 Bonn, Germany

²Department of Clinical Neurosciences, Stroke Research Group, University of Cambridge, UK

³Unit of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

⁴Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany

⁵Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

⁶Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

⁷Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

⁸Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

⁹German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

Contributors: LCAR-J and MT conceived the study. LCAR-J and MT wrote the first and successive drafts of the manuscript. LCAR-J and MT analysed the data. SCL contributed to study conception and design. All authors revised the manuscript for important intellectual content. LCAR-J and MT had full access to the data and take responsibility for the integrity of the data and the accuracy of the data analysis. LCAR-J is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This work was in part supported by a British Heart Foundation programme grant (RG/16/4/32218). LCAR-J was supported by a British Heart Foundation immediate research fellowship (FS/15/61/31626). CLS is chief scientist for UK Biobank. The main sources of funding for CLS's salary are UK Biobank and the Scottish funding Council. MD received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 666881 (SVDs@target) and from the DFG through the CRC 1123 (B3) and the Munich Cluster for Systems Neurology (EXC 1010 SyNergy). HSM is supported by a National Institute for Health Research (NIHR) senior investigator award, and his work is supported by the Cambridge Universities NIHR Comprehensive Biomedical Research Centre. The funding sources had no role in the design or conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript. The MEGASTROKE project received funding from sources specified at www.megastroke.org/acknowledgments.html.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: no support from any organisation for the submitted work; HSM has been paid for delivering educational presentations for AstraZeneca; no other financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: UK Biobank received ethical approval from the research ethics committee (reference 13/NW/0382). All participants provided informed consent to participate. The present analyses were conducted under UK Biobank application number 19463.

Data sharing: The genetic and phenotypic UK Biobank data are available on application to the UK Biobank (www.ukbiobank.ac.uk/). Summary statistics from the MEGASTROKE meta-analysis of genome wide association studies in stroke and stroke subtypes are available from www.megastroke.org.

Transparency: The lead author (LCAR-J) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; that discrepancies from the study as planned have been explained, and that the paper conforms to transparency policy of the International Committee of Medical Journal Editors uniform requirement for manuscripts submitted to biomedical journals.

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

1 GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;385:117-71. doi:10.1016/S0140-6736(14)61682-2

- 2 Boehme AK, Esenwa C, Elkind MS. Stroke Risk Factors, Genetics, and Prevention. *Circ Res* 2017;120:472-95. doi:10.1161/CIRCRESAHA.116.308398
- 3 Brass LM, Isaacs JH, Merikangas KR, Robinette CD. A study of twins and stroke. *Stroke* 1992;23:221-3. doi:10.1161/01.STR.23.2.221
- 4 Flossmann E, Schulz UG, Rothwell PM. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke* 2004;35:212-27. doi:10.1161/01.STR.0000107187.84390.AA
- 5 Malik R, Chauhan G, Traylor M, et al, AFGen Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, International Genomics of Blood Pressure (iGen-BP) Consortium, INVENT Consortium, STARNET, BioBank Japan Cooperative Hospital Group, COMPASS Consortium, EPIC-CVD Consortium, EPIC-InterAct Consortium, International Stroke Genetics Consortium (ISGC), METASTROKE Consortium, Neurology Working Group of the CHARGE Consortium, NINDS Stroke Genetics Network (SiGN), UK Young Lacunar DNA Study, MEGASTROKE Consortium, MEGASTROKE Consortium. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018;50:524-37. doi:10.1038/s41588-018-0058-3
- 6 Colpani V, Baena CP, Jaspers L, et al. Lifestyle factors, cardiovascular disease and all-cause mortality in middle-aged and elderly women: a systematic review and meta-analysis. *Eur J Epidemiol* 2018;33:831-45. doi:10.1007/s10654-018-0374-z
- 7 Larsson SC, Akesson A, Wolk A. Healthy diet and lifestyle and risk of stroke in a prospective cohort of women. *Neurology* 2014;83:1699-704. doi:10.1212/WNL.0000000000000954
- 8 Khera AV, Emdin CA, Drake I, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med* 2016;375:2349-58. doi:10.1056/NEJMoa1605086
- 9 Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779. doi:10.1371/journal.pmed.1001779
- 10 Benjamin EJ, Blaha MJ, Chiuve SE, et al, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation* 2017;135:e146-603. doi:10.1161/CIR.0000000000000485
- 11 Schmier C, Sudlow C. Definitions of acute myocardial infarction (MI) and main MI pathological types for UK Biobank phase 1 outcomes adjudication: UK Biobank; 2017. https://biobank.ctsu.ox.ac.uk/crystal/docs/alg_outcome_mi.pdf [accessed 1 July 2017].
- 12 Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500 000 UK Biobank participants. *BioRxiv* 2017 20-07-2017.
- 13 Rutten-Jacobs LCA, Tozer DJ, Duering M, et al. Genetic Study of White Matter Integrity in UK Biobank (N=8448) and the Overlap With Stroke, Depression, and Dementia. *Stroke* 2018;49:1340-7. doi:10.1161/STROKEAHA.118.020811
- 14 McCarthy S, Das S, Kretschmar W, et al, Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;48:1279-83. doi:10.1038/ng.3643
- 15 O'Connell J, Sharp K, Shrine N, et al. Haplotype estimation for biobank-scale data sets. *Nat Genet* 2016;48:817-20. doi:10.1038/ng.3583
- 16 Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75. doi:10.1086/519795
- 17 Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc* 1999;94:496-509. doi:10.1080/01621459.1999.10474144.
- 18 Fava C, Sjögren M, Olsson S, et al. A genetic risk score for hypertension associates with the risk of ischemic stroke in a Swedish case-control study. *Eur J Hum Genet* 2015;23:969-74. doi:10.1038/ejhg.2014.212
- 19 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. doi:10.1161/STROKEAHA.113.003044
- 20 Tada H, Shiffman D, Smith JG, et al. Twelve-single nucleotide polymorphism genetic risk score identifies individuals at increased risk for future atrial fibrillation and stroke. *Stroke* 2014;45:2856-62. doi:10.1161/STROKEAHA.114.006072
- 21 Traylor M, Rutten-Jacobs LC, Thijs V, et al. Genetic Associations With White Matter Hyperintensities Confer Risk of Lacunar Stroke. *Stroke* 2016;47:1174-9. doi:10.1161/STROKEAHA.115.011625
- 22 Chiuve SE, Rexrode KM, Spiegelman D, Logroscino G, Manson JE, Rimm EB. Primary prevention of stroke by healthy lifestyle. *Circulation* 2008;118:947-54. doi:10.1161/CIRCULATIONAHA.108.781062

- 23 Larsson SC, Åkesson A, Wolk A. Primary prevention of stroke by a healthy lifestyle in a high-risk group. *Neurology* 2015;84:2224-8. doi:10.1212/WNL.0000000000001637
- 24 Zhang Y, Tuomilehto J, Jousilahti P, Wang Y, Antikainen R, Hu G. Lifestyle factors on the risks of ischemic and hemorrhagic stroke. *Arch Intern Med* 2011;171:1811-8. doi:10.1001/archinternmed.2011.443
- 25 Myint PK, Luben RN, Wareham NJ, Bingham SA, Khaw KT. Combined effect of health behaviours and risk of first ever stroke in 20,040 men and women over 11 years' follow-up in Norfolk cohort of European Prospective Investigation of Cancer (EPIC Norfolk): prospective population study. *BMJ* 2009;338:b349. doi:10.1136/bmj.b349
- 26 Seshadri S, Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *Lancet Neurol* 2007;6:1106-14. doi:10.1016/S1474-4422(07)70291-0
- 27 Touzé E, Rothwell PM. Sex differences in heritability of ischemic stroke: a systematic review and meta-analysis. *Stroke* 2008;39:16-23. doi:10.1161/STROKEAHA.107.484618
- 28 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. doi:10.1161/STROKEAHA.115.009816
- 29 Andersen KK, Andersen ZJ, Olsen TS. Age- and gender-specific prevalence of cardiovascular risk factors in 40,102 patients with first-ever ischemic stroke: a Nationwide Danish Study. *Stroke* 2010;41:2768-74. doi:10.1161/STROKEAHA.110.595785
- 30 Reeves MJ, Bushnell CD, Howard G, et al. Sex differences in stroke: epidemiology, clinical presentation, medical care, and outcomes. *Lancet Neurol* 2008;7:915-26. doi:10.1016/S1474-4422(08)70193-5
- 31 von Sarnowski B, Putaala J, Grittner U, et al, sifap1 Investigators. Lifestyle risk factors for ischemic stroke and transient ischemic attack in young adults in the Stroke in Young Fabry Patients study. *Stroke* 2013;44:119-25. doi:10.1161/STROKEAHA.112.665190
- 32 Abraham G, Havulinna AS, Bhalala OG, et al. Genomic prediction of coronary heart disease. *Eur Heart J* 2016;37:3267-78. doi:10.1093/eurheartj/ehw450

Supplementary information: additional material
Supplementary table: table S3 showing SNPs included in MEGASTROKE risk score