White matter hyperintensity burden in acute stroke patients differs by ischemic stroke subtype

Anne-Katrin Giese, MD[1,2]; Markus D. Schirmer, PhD[1,3,4]; Adrian V. Dalca, PhD[3,5]; Ramesh Sridharan, PhD[3]; Kathleen L. Donahue, BS[1]; Marco Nardin, MS[1]; Robert Irie, PhD[5]; Elissa C. McIntosh, MA[5]; Steven J.T. Mocking, MSc[5]; Huichun Xu, MD, PhD[6]; John W. Cole, MD[7]; Eva Giralt-Steinhauer, MD, PhD[8]; Jordi Jimenez-Conde, MD, PhD[8]; Christina Jern, MD, PhD[9]; Dawn O. Kleindorfer, MD[10]; Robin Lemmens, MD, PhD[11]; Johan Wasselius, MD[12]; Arne Lindgren, MD, PhD[12,13]; Tatjana Rundek, MD, PhD[14]; Ralph L. Sacco, MD, MS[14]; Reinhold Schmidt, MD[15]; Pankaj Sharma, MD, PhD[16]; Agnieszka Slowik, MD, PhD[17]; Vincent Thijs, MD, PhD[18]; Bradford B. Worrall, MD, MSc[19]; Daniel Woo, MD, MS[10]; Steven J. Kittner, MD, MPH[7]; Patrick F. McArdle, PhD[6]; Braxton D. Mitchell, PhD, MPH[6]; Jonathan Rosand, MD, MSc[1,2,5,20,21]; James F. Meschia, MD[22]; Ona Wu, PhD[1,5]; Polina Golland, PhD[3]; Natalia S. Rost, MD, MPH[1] on behalf of the International Stroke Genetics Consortium and the MRI-GENIE Investigators.

Affiliations:

1. Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

2. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

3. Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

4. Department of Population Health Sciences, German Centre for Neurodegenerative Diseases (DZNE), Germany

5. Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital (MGH), Charlestown, MA, USA

6. Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

7. Department of Neurology, University of Maryland School of Medicine and Veterans Affairs Maryland Health Care System, Baltimore, MD, USA

8. Department of Neurology, Neurovascular Research Group (NEUVAS), IMIM-Hospital del Mar (Institut Hospital del Mar d'Investigacions Mèdiques), Universitat Autonoma de Barcelona, Barcelona, Spain

9. Institute of Biomedicine, the Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

10. Department of Neurology and Rehabilitation Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, USA

11. KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology; VIB, Vesalius Research Center, Laboratory of Neurobiology, University Hospitals Leuven, Department of Neurology, Leuven, Belgium

12. Department of Clinical Sciences Lund, Neurology, Lund University, Lund, Sweden

13. Department of Neurology and Rehabilitation Medicine, Neurology, Skåne University Hospital, Lund, Sweden

14. Department of Neurology, Miller School of Medicine, University of Miami, The Evelyn F. McKnight Brain Institute, Miami, FL, USA

15. Department of Neurology, Clinical Division of Neurogeriatrics, Medical University Graz, Graz, Austria

16. Institute of Cardiovascular Research, Royal Holloway University of London (ICR2UL), Egham, UK Ashford and St Peter's Hospital, UK

17. Department of Neurology, Jagiellonian University Medical College, Krakow, Poland

18. Stroke Division, Florey Institute of Neuroscience and Mental Health, University of Melbourne Heidelberg, Australia and Department of Neurology, Austin Health, Heidelberg, Victoria, Australia

19. Departments of Neurology and Public Health Sciences, University of Virginia, Charlottesville, VA, USA

20. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

21. Henry and Allison McCance Center for Brain Health, Boston, MA, USA

22. Department of Neurology, Mayo Clinic, Jacksonville, FL, USA

Disclosures

Anne-Katrin Giese reports no disclosures.

Markus D. Schirmer reports no disclosures.

Adrian V. Dalca reports no disclosures.

Ramesh Sridharan reports no disclosures.

Kathleen L. Donahue reports no disclosures.

Marco Nardin reports no disclosures.

Robert Irie reports no disclosures.

Elissa C. McIntosh reports no disclosures.

Steven J.T. Mocking reports no disclosures.

Huichun Xu reports no disclosures.

John W. Cole reports no disclosures.

Eva Giralt-Steinhauer reports no disclosures.

Jordi Jimenez-Conde reports no disclosures.

Christina Jern reports no disclosures.

Dawn O. Kleindorfer reports no disclosures.

Robin Lemmens reports no disclosures.

Johan Wasselius reports no disclosures.

Arne Lindgren reports no disclosures.

Tatjana Rundek reports no disclosures.

Ralph L. Sacco reports no disclosures.

Reinhold Schmidt reports no disclosures.

Pankaj Sharma reports no disclosures.

Agnieszka Slowik reports no disclosures.

Vincent Thijs reports no disclosures.

Bradford B. Worrall reports no disclosures.

Daniel Woo reports no disclosures.

Steven J. Kittner reports no disclosures.

Patrick F. McArdle reports no disclosures.

Braxton D. Mitchell reports no disclosures.

Jonathan Rosand reports no disclosures.

James F. Meschia reports no disclosures.

Ona Wu reports no disclosures.

Polina Golland reports no disclosures.

Natalia S. Rost reports no disclosures.

General information

Word count abstract: 249

Word count text: 3405

Character count title: 94

Number of references: 34

Number of tables: 3

Number of figures: 3

Correspondence and Reprints:

Natalia S. Rost, MD, MPH

J. Philip Kistler Stroke Research Center,

Massachusetts General Hospital,

175 Cambridge Street, Suite 300,

Boston MA 02114, USA.

Phone: (617) 643-3877, Fax: (617) 643-3939

Email: nrost@partners.org

Email addresses for all co-authors

Anne-Katrin Giese – anne-katrin.giese@gmx.de, Markus D. Schirmer – MSCHIRMER1@mgh.harvard.edu, Adrian V. Dalca – ADALCA@BWH.HARVARD.EDU, Ramesh Sridharan – ramesh.eecs@gmail.com, Kathleen L. Donahue – KLDONAHUE@mgh.harvard.edu, Marco Nardin – MNARDIN@mgh.harvard.edu, Robert Irie – irie@nmr.mgh.harvard.edu, Elissa C. McIntosh – mcintosh.elissa@gmail.com, Steven J.T. Mocking – s.mocking@gmail.com, Huichun Xu – hxu@som.umaryland.edu, John W. Cole – JCole@som.umaryland.edu, Eva Giralt-Steinhauer – egiralt@imim.es, Jordi Jimenez-Conde – jjimenez@imim.es, Christina Jern - christina.jern@neuro.gu.se, Dawn O. Kleindorfer – dawn.kleindorfer@uc.edu, Robin Lemmens – robin.lemmens@uzleuven.be, Johan Wasselius – johan.wasselius@gmail.com, Arne Lindgren – Arne.G.Lindgren@skane.se, Tatjana Rundek – TRundek@med.miami.edu, Ralph L. Sacco – rsacco@med.miami.edu, Reinhold Schmidt – reinhold.schmidt@medunigraz.at, Pankaj Sharma – Pankaj.Sharma@rhul.ac.uk, Agnieszka Slowik – slowik@neuro.cm-uj.krakow.pl, Vincent Thijs – vincent.thijs@florey.edu.au, Bradford B. Worrall – BBW9R@hscmail.mcc.virginia.edu, Daniel Woo – WOODL@UCMAIL.UC.EDU, Steven J. Kittner – SKittner@som.umaryland.edu, Patrick F. McArdle – pmcardle@som.umaryland.edu, Braxton D. Mitchell – bmitchel@som.umaryland.edu, Jonathan Rosand – JROSAND@mgh.harvard.edu, James F. Meschia – Meschia.James@mayo.edu, Ona Wu – ona@nmr.mgh.harvard.edu, Polina Golland – polina@csail.mit.edu, Natalia S. Rost – NROST@PARTNERS.ORG

Statistical Analysis:

Statistical analysis performed by Anne-Katrin Giese and Natalia S. Rost (Department of Neurology, Massachusetts General Hospital, Boston, MA, USA)

Search terms:

WMH, white matter hyperintensity, ischemic stroke, causative classification of stroke, small vessel disease

AUTHOR CONTRIBUTIONS

Anne-Katrin Giese: Study design, data analysis and interpretation of data, writing of manuscript

Markus D. Schirmer: Data analysis and interpretation of data, editing of manuscript

Adrian V. Dalca: Data analysis and interpretation of data, editing of manuscript

Ramesh Sridharan: Data analysis and interpretation of data, editing of manuscript

Kathleen L. Donahue: Literature search, Data analysis and interpretation of data

Marco Nardin: Data analysis and interpretation of data

Robert Irie: Data collection, editing of manuscript

Elissa C. McIntosh: Data collection, editing of manuscript

Steven J.T. Mocking: Data collection, editing of manuscript

Huichun Xu: Data interpretation, editing of manuscript

John W. Cole: Data collection, editing of manuscript

Eva Giralt-Steinhauer: Data collection, editing of manuscript

Jordi Jimenez-Conde: Data collection, editing of manuscript

Christina Jern: Data collection, editing of manuscript

Dawn O. Kleindorfer: Data collection, editing of manuscript

Robin Lemmens: Data collection, editing of manuscript

Johan Wasselius: Data collection, editing of manuscript

Arne Lindgren: Data collection, editing of manuscript

Tatjana Rundek: Data collection, editing of manuscript

Ralph L. Sacco: Data collection, editing of manuscript

Reinhold Schmidt: Data collection, editing of manuscript

Pankaj Sharma: Data collection, editing of manuscript

Agnieszka Slowik: Data collection, editing of manuscript

Vincent Thijs: Data collection, editing of manuscript

Bradford B. Worrall: Data collection, editing of manuscript

Daniel Woo: Data collection, editing of manuscript

Steven J. Kittner: Data collection, data interpretation, editing of manuscript

Patrick F. McArdle: Data collection, data interpretation, editing of manuscript

Braxton D. Mitchell: Data interpretation, editing of manuscript

Jonathan Rosand: Data interpretation, editing of manuscript

James F. Meschia: Data collection, editing of manuscript

Ona Wu: Data collections, data analysis, data interpretation, editing of manuscript

Polina Golland: Data analysis and interpretation of data, editing of manuscript

Natalia S. Rost: Study design, data analysis and interpretation of data, writing of manuscript

ABSTRACT

Objective: To examine etiologic stroke subtypes and vascular risk factor profiles and their association with WMH burden in patients hospitalized for acute ischemic stroke (AIS).

Methods: For the Magnetic Resonance Imaging and Genetics Interface Exploration (MRI-GENIE) study, we systematically assembled brain imaging and phenotypic data for 3,301 AIS patients. All cases underwent standardized web-tool-based stroke subtyping with the causative classification of stroke (CCS). WMH volume (WMHv) was measured on T2 brain MRI scans of 2,529 patients using a fully automated deep-learning trained algorithm. Univariable and multivariable linear mixed effects modelling was carried out to investigate the relationship of vascular risk factors with WMHv and CCS subtypes.

Results: AIS patients with large artery atherosclerosis, major cardioembolic stroke, small artery occlusion (SAO), other and undetermined causes of AIS differed significantly in their vascular risk factor profile (all $p<0.001$). Median WMHv in all AIS patients was 5.86 cm^3 [Interquartile range (IQR): 2.18 - 14.61 cm^3] and differed significantly across CCS subtypes ($p<0.0001$). In multivariable analysis, age, hypertension, prior stroke, smoking (all $p<0.001$) and diabetes mellitus ($p=0.041$) were independent predictors of WMHv. When adjusted for confounders, SAO patients had significantly higher WMHv compared to all other stroke subtypes ($p<0.001$).

Conclusions: In this international multi-center, hospital-based cohort of AIS patients we demonstrate that vascular risk factor profiles and extent of WMH burden differ by CCS subtype, with the highest lesion burden detected in SAO patients. These findings independently validate the small-vessel hypothesis of WMH lesions detected on brain MRI of patients with ischemic stroke.

INTRODUCTION

White matter hyperintensity (WMH) is a common radiographic marker seen in the deep and periventricular white matter on fluid-attenuated inversion recovery (FLAIR) in magnetic resonance imaging (MRI).[1] In acute ischemic stroke (AIS) patients, WMH is associated with susceptibility to infarct growth[2] and poor post-stroke outcomes.[3] As a radiographic manifestation of chronic cerebrovascular disease, WMH burden is thought to result from the impact of multiple vascular risk factors on the small vasculature and is known to be associated with ongoing injury, including higher rates of WMH accumulation and other cerebrovascular manifestations .[4] In stroke-free adults, known vascular risk factors such as hypertension,[5] carotid atherosclerosis,[6] diabetes mellitus[7] and cigarette smoking[8] are strongly associated with increased WMH volume (WMHv). However, limited data are available regarding risk factors for WMH in AIS, with age and elevated homocysteine levels having been previously identified.[9]

Given the lack of comparative data, we aimed to assess whether risk factors contributing to WMH severity in patients with AIS may in part be different from those in population-based stroke-free cohorts. However, reliable WMH assessment in AIS is compounded by the varying methodology, ranging from a semi-quantitative rating scale[10] to semi-/fully - automated WMH measurements.[11,12] Furthermore, many of these approaches are challenging to apply to large clinical datasets of WMH. To address this gap, we used a designated artificial intelligence-driven deep learning pipeline to derive robust WMHv and systematically investigate determinants of WMHv in AIS and its subtypes in a retrospective, hospital-based cohort of 3,301 AIS patients. We hypothesize that WMH burden is highest in patients with SAO and that classic vascular risk factors contribute to a higher WMH burden.

METHODS

Study design and participants

The MRI-Genetics Interface Exploration (MRI-GENIE) Study is a large, international collaboration of 12 sites contributing 3,301 AIS patients with phenotypic, radiographic and genotypic data. A detailed description of the study design and concept has been published previously.[13] In summary, each site received approval of their respective institutional review board. Patients were recruited through the Stroke Genetics Network (SiGN), with recruitment dates ranging from 1999 to 2012.[14] MRI data were assembled in a central imaging repository for assessment of neuroimaging phenotypes. Here, we assessed 2,781 patients for whom FLAIR imaging was available for automated assessment. MRI sequences were obtained in the acute phase (median time to scan: <1 day [upper quartile range 1-4 days] from symptom onset). After quality control (QC), 2,529 AIS patients with automatedly extracted WMH volume (WMHv) from clinical axial FLAIR images remained available for analysis (**Figure 1**).[15]

Standard Protocol Approvals, Registrations, and Patient Consents

All AIS subjects were recruited in a hospital-based setting. Ethics or institutional review board approval was obtained as appropriate for each individual participating study. Informed consent including sharing of de-identified demographic, imaging and genotyping data was obtained from all patients or their legally authorized representative. Demographic and genotyping data was collected by the Data Management and Genotyping Core of SiGN.

Automated WMH volume extraction

WMHv were successfully extracted from clinical axial FLAIR images of 2,529 patients (76.6%) in the MRI-GENIE cohort. All FLAIR images were acquired as per local AIS clinical imaging protocol. On average, the images had a mean resolution of 0.7mm in-plane (minimum: 0.4mm, maximum: 1.9mm) and 6.3mm through-plane (minimum: 1.0mm, maximum 65.0), whole-brain acquisition. A fully automated pipeline using deep learning was developed specifically for quantification of WMHv on clinical-grade MRI scans and applied to all patients.[15] A detailed description including assessment of the WMH pipeline performance has been previously published.[15] In brief, the pipeline has three main processing steps with two QC checks. First, an initial QC assessment based on in-plane and through-plane resolution, as well as number of slices, is performed to identify scans with insufficient information for WMH extraction. Brains are extracted from the clinical axial FLAIR images using a dedicated deep learning architecture (library found at http://github.com/adalca/neuron). After brain extraction, patients with unexpectedly low or high age-stratified brain volume, were identified and manually assessed for incomplete brain extraction. To account for differences in image acquisition, this is followed by an intensity normalization to harmonize image intensities across sites and scanners. Then, automated WMH segmentation is performed, again by using a deep learning architecture with atlas-based, spatial priors to include information on the distribution of WMH and differentiation from AIS artifacts such as edema, movement artifact, or prior stroke, with the algorithm compensating for large infarctions to avoid underestimation of WMH.[15-17] Reasons for excluding scans during the first pass QC (n=254, 10% of all subjects with FLAIR) were: low number of slices, mislabeled MRI sequence, or significant motion artifact (n=97; 3.8%). In the second pass QC, images

with different slice direction (coronal [n=72; 1.1%] or sagittal [n=3; 0.1%]), persistent motion artifact (n=8; 0.3%), incomplete brain extraction (n=62; 2.5%) or other issues (n=8, 0.3%) were excluded. Finally, we excluded four subjects with WMHv measurement of 0cc (n=4; 0.2%).

Acute Ischemic Stroke Subtyping

All patients underwent systematic stroke subtyping using the CCS.[18] A web-based standardized algorithm incorporates results of patient history, physical exam findings from the clinical stroke assessment, and diagnostic testing to systematically assign the CCS subtype. Two major CCS classifications have been established, phenotypic CCS reflecting the abnormal test results at the time of stroke and which does not rely on judgment regarding the likely etiology and causative CCS, which requires integration of all diagnostic tests and history to provide the most likely mechanism for the stroke. Throughout this article, we will refer to causative CCS as "CCS". If presented with multiple competing etiologies the web-based CCS algorithm assigns the most probable cause. CCS subtypes include large artery atherosclerosis (LAA), major cardioembolic stroke (CE major), SAO, Other, and Undetermined cause of stroke (i.e., 5-item CCS subtypes).[18] LAA, CE major, SAO, and Other categories include subjects where the subtype was considered either possible, probable or evident. "Undetermined" category included the following: cryptogenic embolism, other cryptogenic stroke, minor cardioembolic stroke, as well as subjects with incomplete information, or unclassified stroke. Each reader underwent formal training and certification prior to adjudicating CCS subtypes.[18]

Statistical Analysis

Demographic data and vascular risk factors including age, sex, race, atrial fibrillation, coronary artery disease (CAD), diabetes mellitus, hypertension, prior stroke, and smoking status were abstracted from patient records by each site. Numeric variables are expressed as mean ± standard deviation (SD) or median and IQR, depending on normal or non-parametric distribution. Categorical variables are expressed as counts and frequencies. Statistical comparison is performed across the 5-item CCS subtypes using $\chi 2$-test to compare categorical data, ANOVA for age and Kruskal-Wallis-Test for WMHv. Given the skewed distribution of WMHv, we used natural log-transformed WMHv for all regression analyses. Included patients were compared to those who failed the imaging QC using mixed logistic regression model of "included/excluded" status with the demographic variables as fixed effects variables and study site as a random effects variable for each demographic parameter. After adjustment for site as a random variable, the subjects passing or failing QC did not differ significantly (data not shown). Further, univariable linear regression was used to identify predictors of WMHv. Variables passing $p<0.1$ were included in the multivariable model. Multiple linear mixed-effect modelling was used to identify independent determinants of WMHv. The multiple linear mixed-effects model was adjusted for site as a random variable, as key variables such as age, race and distribution of vascular risk factors (atrial fibrillation, CAD, diabetes mellitus and hypertension) differed by site (data not shown). Cases with missing information for vascular risk factors were excluded from the final multivariable model (**Table 1**).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study. The corresponding author had final responsibility for the decision to submit for publication.

RESULTS

This retrospective analysis was conducted based on data collected between 1999 and 2012. Imaging data were assembled between 2012 and 2017. WMH analysis was conducted using de-identified clinical FLAIR images in June 2018. Of the 3,301 MRI-GENIE patients 518 had no FLAIR sequence and 254 failed the QC assessment for the automated WMH extraction, leaving 2,529 AIS patients for analysis (**Figure 1**). Mean age was 63.4 (SD=14.5) years and 39.3% (n=993) of all patients were female. The majority of patients were Caucasians (n=2,141; 84.7%), had a prior medical history of hypertension (n=1,668; 66.4%) and were either current or former tobacco smokers (n=1,323; 54.1%). Atrial fibrillation (n=380; 15.5%), CAD (n=444; 17.9%), diabetes mellitus (n=581; 23.2%) and a history of prior stroke (n=248; 9.8%) were less common in this AIS cohort (**Table 1**). MRI-GENIE patients excluded from this analysis due to lack or poor quality images did not differ significantly in age, sex, or other vascular risk factors from those included.

Vascular risk factor profiles in CCS Subtypes

Distribution of age, sex and vascular risk factors differed significantly across the 5-item CCS subtypes, except for history of prior stroke (**Table 1**). Patients with LAA were less likely to be female (n=176; 32.3%) or suffer from atrial fibrillation (n=30; 5.6%).Patients classified as a major CE stroke were more likely to be female (n=186; 47.2%), were on average older (mean=71.8 [SD=11.9] years) and were more likely to have atrial fibrillation (n=267; 68.8%), CAD (n=111; 28.6%), and were slightly more likely to have hypertension (n=288; 73.3%) in comparison with the other stroke subtypes, albeit not significant. Patients classified as SAO had the lowest frequency of atrial fibrillation (n=12; 3.2%) and the highest frequency of diabetes mellitus (n=107; 28.1%). Notably, patients classified as Other cause of ischemic stroke were the youngest (mean=49.0 [SD=13.6] years) and had the lowest frequency of history of CAD (n=18; 10.2%) and diabetes mellitus (n=27; 15.1%).

White Matter hyperintensity in CCS Subtypes

WMHv was extracted for each patient using an automated WMH pipeline for clinical axial FLAIR images (example outlines in **Figure 2**). The median WMHv (example outlines in **Figure 2**) in the entire AIS cohort was 5.86 cm^3 [IQR: 2.18-14.61cm^3]. The unadjusted median WMHv was the highest in patients classified as CE major (8.13 cm^3 [IQR: 3.65-17.12cm^3]), and second highest in SAO (7.53cm^3 [IQR: 2.84-18.45cm^3]). The lowest WMHv with a median volume of 2.16cm^3 [IQR: 0.93-5.29cm^3] was observed in patients classified as Other cause of stroke (**Table 1**).

Univariable associations with WMH volume in AIS

In univariable analysis, age was the strongest predictor of WMHv (β=0.05, 95%CI: 0.05-0.05, p<0.001). Further, atrial fibrillation (β=0.46, 95%CI: 0.31-0.61), CAD (β=0.41, 95%CI: 0.27-0.55), diabetes mellitus (β=0.35, 95%CI: 0.23-0.48), hypertension (β=0.82, 95%CI: 0.72-0.93) and prior stroke (β=0.55, 95%CI: 0.37-0.72) were significant univariable predictors of WMHv, all lower than p<0.001. Likewise, a history of current or former smoking contributed to a higher WMHv in univariable analysis (β=0.13, 95%CI: 0.03-0.24, p=0.015) (**Table 2**).

Multivariable associations with WMH volume in AIS

The multivariable analysis included all variables passing the threshold of p<0.1 and was adjusted for site as a random variable. Age (β=0.05, 95%CI: 0.04-0.05), hypertension (β=0.35, 95%CI: 0.27-0.47), history of prior stroke (β=0.45, 95%CI: 0.32-0.63), current or former smoking status (β=0.19, 95%CI: 0.10-0.28) (all p<0.001) as well as diabetes mellitus (β=0.11, 95%CI: 0.03-0.24, p=0.041) remained independent predictors of WMHv (**Table 2**).

Adjusted WMH volume in CCS subtypes

WMHv for the 5-item CCS subtypes was re-assessed when adjusted for the identified multivariable predictors of WMHv, as well as for site as a random variable by comparing the residuals of the multivariable linear mixed-effects model by CCS subtype. Cases classified as SAO demonstrated the highest adjusted residual WMHv (1.47cm^3 [IQR: -

1.52-3.15cm^3]) in the Bonferroni-adjusted group-wise comparison of CCS subtypes (**Figure 3**).

CCS-subtype specific associations with WMH volume

Contributors to WMH burden were also assessed by CCS subtype (**Table 3**). In LAA, age (β=0.05, 95%CI: 0.04-0.05, p<0.001) and smoking status (β=0.68, 95%CI: 0.34-1.01, p<0.001) were independently associated with higher WMHv. In CE major, only age (β=0.05, 95%CI: 0.04-0.05, p<0.001) emerged as an independent predictor of WMHv in the multivariable model. In contrast, age (β=0.04, 95%CI: 0.03-0.05, p<0.001), hypertension (β=0.43, 95%CI: 0.16-0.70, p=0.002) and prior stroke (β=0.52, 95%CI: 0.11-0.93) emerged as independent predictors of WMHv in SAO. In cases with Other causes of stroke, age (β=0.04, 95%CI: 0.03-0.06, p<0.001) and prior stroke (β=0.90, 95%CI: 0.13-1.66, p=0.021) were associated with higher WMHv, and female sex (β=-0.37, 95%CI: -0.72- -0.02, p=0.040) was associated with lower WMHv. Lastly, in patients with Undetermined cause of stroke, age (β=0.05, 95%CI: 0.04-0.05, p<0.001), hypertension (β=0.58, 95%CI: 0.42-0.74, p<0.001), prior stroke (β=0.38, 95%CI: 0.15-0.61, p=0.001) and smoking (β=0.23, 95%CI: 0.09-0.37, p=0.001) were independently associated with higher WMHv, whereas CAD was associated with lower WMHv (β=-0.26, 95%CI: -0.45- -0.05, p=0.011).

DISCUSSION

In this large multi-center, hospital-based cohort of 2,529 AIS patients, we demonstrate that vascular risk factor profiles differ across CCS subtypes. Further, we show that, as in stroke-free populations, age, hypertension and smoking (all $p<0.001$), as well as diabetes mellitus ($p=0.041$) are independent predictors of high WMHv. Additionally, when adjusted for confounding variables, patients with SAO exhibit the highest amount of WMH burden, whereas patients with CE major have the highest unadjusted WMHv.

These findings independently validate prior studies examining patients with thoroughly ascertained stroke subtypes. In a study of 891 AIS patients with the Trial of Org 10172 in Acute Stroke Treatment (TOAST) stroke subtype classification,[19] SAO patients had the highest amount of WMH burden across all AIS subtypes.[9] Recently, a large multi-center study semi-automatedly assessed WMHv in 5,035 AIS patients of Korean descent.[3] Vascular risk factor profiles were compared across WMH quintiles. Overall, hypertension, diabetes mellitus, and atrial fibrillation emerged as independently associated with WMH quintile when adjusted for age and sex. While the findings were overall similar, our study highlights the importance of assessing vascular risk factors by stroke subtype, particularly when a standardized subtyping-tool like CCS is used. We show that in patients with SAO, a prior medical history of hypertension and stroke are independently associated with larger WMH burden, highlighting potentially addressable risk factors in SAO. Additionally, prior stroke is a predictor of WMH burden in LAA, SAO, Other and Undetermined cases, possibly hinting at a “vicious cycle,” where existing cerebrovascular burden increases the risk for further cerebral tissue injury. In cases of Undetermined stroke, CAD, hypertension and smoking also are significantly

associated with WMH burden. However, this may reflect the potentially competing underlying stroke etiologies. Further, in multivariable modelling, we show that in LAA, only age and prior stroke were independently associated with WMHv, while in CE major only age independently contributed to WMH burden. Such specific CCS subtype findings support the concept of a different underlying etiologic disease processes.

Consistent with prior studies, age remained as the most significant independent determinant of WMHv. However, the effect of age on WMH burden may differ across the lifespan, and it may interact with other vascular risk factors. For example, in a cohort of 560 AIS patients at the extremes of ages in young (<55 years) and old (>75 years), different vascular risk profiles emerged.[23]

Among other vascular risk factors, hypertension has a well-established role in WMH accumulation in population-based cohorts;[5] furthermore, given its robust association with WMHv in this large cohort of AIS patients, prior studies[9,24] have most likely been underpowered. Likewise, diabetes mellitus has been implicated in the development and lesion size of WMH in stroke-free adults,[7,25] but current analysis is the first to demonstrate an association between the WMH severity and diabetes mellitus in AIS patients.

Further, we observed a relationship of AF with higher WMH burden in univariable analysis, which no longer persisted after adjusting for age and other potential confounders. Given higher incidence of AF among elderly subjects and potential co-linearity between these two risk factors, the effect of AF on WMH burden in AIS subjects could not be definitively assessed and will require further study.

Lastly, we confirmed smoking as an independent predictor of WMH burden. The importance of smoking exposure on WMH risk is further highlighted by a study from 2015 demonstrating a dose-dependent effect of smoking in stroke-free adults.[26]

Our study provides a key piece of evidence on the association between WMH burden and small-vessel stroke. This is of growing importance, as multiple mechanisms are considered for the pathophysiology of chronic cerebral ischemic changes that appear as WMH lesions on T2/FLAIR MRI. Among these competing factors, arteriolar sclerosis, capillary endothelial activation as well as immunoreactivity for hypoxia-inducible factor (HIF) 1 and 2 as a manifestation of ongoing hypoxia have been described.[27] More recently, the focus for elucidating WMH pathology shifted towards the investigation of microstructural changes in normal appearing white matter, which possibly precede formation of new WMH lesions and WMH lesion progression.[28] Such microstructural changes have been associated with hypertension, smoking, and diabetes mellitus,[29] raising the possibility of addressing modifiable vascular risk factors prior to the formation of new or the expansion of existing WMH. The association of microstructural white matter changes with worsened functional recovery after ischemic stroke further highlights the importance of white matter integrity.[30]

Recent studies also suggest that WMH has the potential to regress over time. In a study investigating 190 patients with minor stroke a repeat MRI at 1 year demonstrated WMH regression in 71 patients.[31] Patients with WMH regression also had a greater reduction in blood pressure, though this requires further validation in prospective cohorts. Patients with increasing WMH had a higher likelihood of experiencing a recurrent ischemic

event.[31] Similarly, patients with higher levels of WMH had a higher likelihood of 90-day stroke recurrence.[32]

This study has important limitations. First, imaging data were collected retrospectively, and their availability varied by site. Since brain MRIs were collected in the acute phase of AIS, variability in the quality of acquisition and clinical indication for neuroimaging is a significant factor in this study. Great care has been exercised in ascertaining the quality and operational utility of the MRI-GENIE neuroimaging database.[13] Furthermore, QC for the automated WMHv segmentation was the key feature of this innovative MRI analysis pipeline and is described in detail elsewhere.[15] An additional limitation of this analysis is related to a large proportion of AIS cases with Undetermined stroke subtype (n=1,022). As 5-item CCS-subtyping was performed with a standardized web-based protocol by trained adjudicators this may be due to either two or more equally likely competing stroke etiologies resulting in the classification “Undetermined” or due to lack of sufficient diagnostic data for a final CCS classification. The proportion of stroke cases classified as Undetermined in the presented MRI-GENIE cohort is slightly lower than in the original SiGN study (39% vs. 43%).[33] In SiGN, this category was mainly driven by cases with incomplete clinical evaluations (55%), minor cardioembolic sources (18%), other cryptogenic sources (15%) and multiple competing etiologies (9%). Overall, the proportion of completely “unclassified” patients was small (4%). Despite the number of Undetermined cases, 5-item CCS has the advantage of a standardized, reproducible web-based assessment with excellent inter-rater agreement (kappa=0.86).[34]

Notable strengths of our study include : (a) MRI-GENIE is a large imaging and genetic database specifically developed to enable future studies of genetic architecture of acute and chronic neuroimaging traits in AIS patients; (b) all cases underwent systematic stroke-subtyping using the standardized web-based CCS tool,[18] and (c) WMH was segmented using an artificial-intelligence enabled automated segmentation pipeline[15] specifically designed to analyze the multi-center, clinical brain MRI of patients with AIS. The automated WMH pipeline has the advantage that it yields robust and reproducible WMH segmentations across multiple sites and that it can be applied to other acute stroke cohorts. The pipeline has demonstrated excellent agreement with manual assessments of WMH volume in AIS patients.[15] Overall, the applied WMH pipeline contributes to the generalizability of our results to other AIS cohorts. Systematic, large-scale WMH assessment in AIS will allow for studying the underlying genetic architecture, as well as assessment of the role of WMH in AIS severity and outcome.

In conclusion, we demonstrated that patients with SAO exhibit the highest amount of WMH when adjusted for confounders compared to other AIS subtypes, supporting the hypothesis that WMH lesions seen in AIS patients are the result of small vessel disease. Furthermore, we have shown that that the vascular risk profile differs by CCS, with modifiable risk factors being important contributors to the overall WMH burden in AIS patients. Our findings in part reconcile the previously described differences in risk factor profiles for WMH in stroke-free and AIS populations. Effectively addressing these

vascular risk factors could provide an important avenue for modifying WMH disease burden and thus, potentially preventing the detrimental downstream effects of high WMH burden in AIS patients.

CONTRIBUTORS:

Lukas Holmegaard, M.D. and Katarina Jood, M.D., Ph.D, both Institute of Neuroscience and Physiology, Department of Clinical Neuroscience, the Sahlgrenska Academy at University of Gothenburg, Sweden

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

SOURCES OF FUNDING:

NIH-NINDS (MRI-GENIE: R01NS086905 - PI N. Rost; K23NS064052, R01NS082285 - N. Rost; SiGN: U01 NS069208 - J. Rosand, S. Kittner; R01NS059775, R01NS063925, R01NS082285, P50NS051343, R01NS086905, U01 NS069208 - O. Wu), NIH NIBIB (P41EB015902 – P. Golland, U01NS030678 – Kissela, Kleindorfer; EB015325 – O. Wu), ISGS: R01NS423733 – P.I. J. Meschia, Swedish Heart and Lung Foundation, Lund University, Region Skåne, the Freemasons Lodge of Instruction Eos Lund, Skåne University Hospital, the Foundation of Färs&Frosta—one of Sparbanken Skåne's ownership Foundations, and the Swedish Stroke Association – A. Lindgren, Swedish Research Council and the Swedish Heart and Lung Foundation, the Swedish State under the ALF agreement – C. Jern, Spanish Ministry of Science and Innovation, Instituto de Salud Carlos III (Funding for Research in Health (PI051737), (PI10/02064),

(PI12/01238) and (PI15/00451 – J. Jiménez-Conde)), Fondos FEDER/EDRF Red de Investigación Cardiovascular (RD12/0042/0020 – J. Jimenez-Conde), Fundació la Marató TV3 (76/C/2011 - J. Jiménez-Conde) and Recercaixa'13 (JJ086116 J. Jiménez-Conde), Wistron Corporation (P. Golland). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 753896 (M.D. Schirmer). RL is a senior clinical investigator of FWO Flanders.

Appendix 1

Name	Location	Role	Contribution
Anne-Katrin Giese, MD	Massachusetts General Hospital, Boston, USA	Author	Designed and conceptualized study; analyzed and interpreted the data; drafted the manuscript for intellectual content
Markus D. Schirmer, PhD	Massachusetts General Hospital, Boston, USA	Author	Analyzed and interpreted the data; revised the manuscript for intellectual content
Adrian V. Dalca, PhD	Massachusetts Institute of Technology, Cambridge, USA	Author	Analyzed and interpreted the data; revised the manuscript for intellectual content
Ramesh Sridharan, PhD	Massachusetts Institute of Technology, Cambridge, USA	Author	Analyzed and interpreted the data; revised the manuscript for intellectual content
Kathleen L. Donahue, BS	Massachusetts General Hospital, Boston, USA	Author	Analyzed and interpreted the data; revised the manuscript for intellectual content
Marco Nardin, MS	Massachusetts General Hospital, Boston, USA	Author	Analyzed and interpreted the data; revised the manuscript for intellectual content
Robert Irie, PhD	Massachusetts General Hospital, Boston, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Elissa C. McIntosh, MA	Massachusetts General Hospital, Boston, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Steven J.T. Mocking, MSc	Massachusetts General Hospital, Boston, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Huichun Xu, MD, PhD	University of Maryland, Baltimore, USA	Author	Interpreted the data; revised the manuscript for intellectual content
John W. Cole, MD	University of Maryland, Baltimore, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Eva Giralt-Steinhauer, MD, PhD	Universitat Autonoma de Barcelona, Barcelona, Spain	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Jordi-Jimenez-Conde, MD, PhD	Universitat Autonoma de Barcelona, Barcelona, Spain	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Christina Jern, MD, PhD	University of Gothenburg,	Author	Major role in acquisition of data; revised the manuscript for

	Gothenburg, Sweden		intellectual content
Dawn O. Kleindorfer, MD	University of Cincinnati, Cincinnati, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Robin Lemmens, MD, PhD	University of Leuven, Leuven, Belgium	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Johan Wasselius, MD	Skåne University, Lund, Sweden	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Arne Lindgren, MD, PhD	Skåne University, Lund, Sweden	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Tatjana Rundek, MD, PhD	University of Miami, Miami, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Ralph L. Sacco, MD, MS	University of Miami, Miami, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Reinhold Schmidt, MD	Medical University Graz, Graz, Austria	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Pankaj Sharma, MD, PhD	Royal Holloway University of London, Egham, UK	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Agnieszka Slowik, MD, PhD	Jagiellonian University, Krakow, Poland	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Vincent Thijs, MD, PhD	University of Melbourne, Heidelberg, Australia	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Bradford B. Worrall, MD, MSc	University of Virginia, Charlottesville, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Daniel Woo, MD, MS	University of Cincinnati, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Steven J. Kittner, MD, MPH	University of Maryland, Baltimore, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Patrick F. McArdle, PhD	University of Maryland, Baltimore, USA	Author	Major role in acquisition of data; revised the manuscript for intellectual content
Braxton D. Mitchell, PhD, MPH	University of Maryland, Baltimore,	Author	Interpreted the data; revised the manuscript for intellectual content

	USA		
Jonathan Rosand, MD, MSc	Massachusetts General Hospital, Boston, USA	Author	Major role in acquisition of data; interpreted the data; revised the manuscript for intellectual content
James F. Meschia, MD		Author	Major role in acquisition of data; revised the manuscript for intellectual content
Ona Wu, PhD	Massachusetts General Hospital, Boston, USA	Author	Major role in acquisition of data; analyzed and interpreted data; revised the manuscript for intellectual content
Polina Golland, PhD	Massachusetts Institute of Technology, Cambridge, USA	Author	Analyzed and interpreted the data; revised the manuscript for intellectual content
Natalia S. Rost, MD, MPH	Massachusetts General Hospital, Boston, USA	Author	Designed and conceptualized study; analyzed and interpreted the data; drafted the manuscript for intellectual content

Appendix 2

Name	Location	Role	Contribution
Anne-Katrin Giese, MD	Massachusetts General Hospital, Boston, USA	Site investigator	Data acquisition; data analysis; image preprocessing; data base creation
Markus D. Schirmer, PhD	Massachusetts General Hospital, Boston, USA	Site investigator	Data analysis; image preprocessing; data base creation
Adrian V. Dalca, PhD	Massachusetts Institute of Technology, Cambridge, USA	Site investigator	Data analysis; image preprocessing; data base creation
Ramesh Sridharan, PhD	Massachusetts Institute of Technology, Cambridge, USA	Site investigator	Data analysis; image preprocessing; data base creation
Kathleen L. Donahue, BS	Massachusetts General Hospital, Boston, USA	Site investigator	Data analysis; image preprocessing; data base creation
Marco Nardin, MS	Massachusetts General Hospital, Boston, USA	Site investigator	Data analysis; image preprocessing; data base creation
Robert Irie, PhD	Massachusetts General Hospital, Boston, USA	Site investigator	Image preprocessing; database creation
Elissa C. McIntosh, MA	Massachusetts General Hospital, Boston, USA	Site investigator	Image preprocessing; database creation
Steven J.T. Mocking, MSc	Massachusetts General Hospital, Boston, USA	Site investigator	Image preprocessing; database creation
Huichun Xu, MD, PhD	University of Maryland, Baltimore, USA	Site investigator	Data analysis; coordinated research at respective site
John W. Cole, MD	University of Maryland, Baltimore, USA	Site investigator	Data acquisition; coordinated research at respective site
Eva Giralt-Steinhauer, MD, PhD	Universitat Autonoma de Barcelona, Barcelona, Spain	Site investigator	Data acquisition; coordinated research at respective site
Jordi-Jimenez-Conde, MD, PhD	Universitat Autonoma de	Site investigator	Data acquisition; coordinated research

	Barcelona, Barcelona, Spain		at respective site
Christina Jern, MD, PhD	University of Gothenburg, Gothenburg, Sweden	Site investigator	Data acquisition; coordinated research at respective site
Dawn O. Kleindorfer, MD	University of Cincinnati, Cincinnati, USA	Site investigator	Data acquisition; coordinated research at respective site
Robin Lemmens, MD, PhD	University of Leuven, Leuven, Belgium	Site investigator	Data acquisition; coordinated research at respective site
Johan Wasselius, MD	Skåne University, Lund, Sweden	Site investigator	Data acquisition; coordinated research at respective site
Arne Lindgren, MD, PhD	Skåne University, Lund, Sweden	Site investigator	Data acquisition; coordinated research at respective site
Tatjana Rundek, MD, PhD	University of Miami, Miami, USA	Site investigator	Data acquisition; coordinated research at respective site
Ralph L. Sacco, MD, MS	University of Miami, Miami, USA	Site investigator	Data acquisition; coordinated research at respective site
Reinhold Schmidt, MD	Medical University Graz, Graz, Austria	Site investigator	Data acquisition; coordinated research at respective site
Pankaj Sharma, MD, PhD	Royal Holloway University of London, Egham, UK	Site investigator	Data acquisition; coordinated research at respective site
Agnieszka Slowik, MD, PhD	Jagiellonian University, Krakow, Poland	Site investigator	Data acquisition; coordinated research at respective site
Vincent Thijs, MD, PhD	University of Melbourne, Heidelberg, Australia	Site investigator	Data acquisition; coordinated research at respective site
Bradford B. Worrall, MD, MSc	University of Virginia, Charlottesville, USA	Site investigator	Data acquisition; coordinated research at respective site
Daniel Woo, MD, MS	University of Cincinnati, USA	Site investigator	Data acquisition; coordinated research at respective site
Steven J. Kittner, MD, MPH	University of Maryland,	Site investigator	Data acquisition; coordinated research

	Baltimore, USA		at respective site
Patrick F. McArdle, PhD	University of Maryland, Baltimore, USA	Site investigator	Data acquisition; coordinated research at respective site
Braxton D. Mitchell, PhD, MPH	University of Maryland, Baltimore, USA	Site investigator	Data acquisition and analysis; coordinated research at respective site
Jonathan Rosand, MD, MSc	Massachusetts General Hospital, Boston, USA	Site investigator	Data acquisition; coordinated research at respective site
James F. Meschia, MD		Site investigator	Data acquisition; coordinated research at respective site
Ona Wu, PhD	Massachusetts General Hospital, Boston, USA	Site investigator	Data acquisition; coordinated research at respective site; image preprocessing; data base creation
Polina Golland, PhD	Massachusetts Institute of Technology, Cambridge, USA	Site investigator	Data analysis; image preprocessing; data base creation
Natalia S. Rost, MD, MPH	Massachusetts General Hospital, Boston, USA	Director of Coordinating Center	Data analysis; data acquisition; coordinated research at respective site

REFERENCES

1. Pantoni L, Garcia JH. Pathogenesis of leukoaraiosis: a review. *Stroke* 1997; **28**(3): 652-659.

2. Ay H, Arsava EM, Rosand J, et al. Severity of leukoaraiosis and susceptibility to infarct growth in acute stroke. *Stroke* 2008; **39**(5): 1409–1413.

3. Ryu WS, Woo SH, Schellingerhout D, et al. Stroke outcomes are worse with larger leukoaraiosis volumes. *Brain* 2017; **140**(1): 158-170.

4. Schmidt R, Fazekas F, Kapeller P, et al. MRI white matter hyperintensities: three-year follow-up of the Austrian Stroke Prevention Study. *Neurology* 1999; **53**(1): 132–139.

5. Longstreth WT, Manolio TA, Arnold A, et al. Clinical correlates of white matter findings on cranial magnetic resonance imaging of 3301 elderly people: the cardiovascular health study. *Stroke* 1996; **27**(8): 1274–1282.

6. de Leeuw FE, de Groot JC, Bots ML, et al. Carotid atherosclerosis and cerebral white matter lesions in a population based magnetic resonance imaging study. *J Neurol* 2000; **247**(4):291–296.

7. Knopman DS, Penman AD, Catellier DJ, et al. Vascular risk factors and longitudinal changes on brain MRI: the ARIC study. *Neurology* 2011; **76**(22): 1879-1885.

8. Jeerakathil T, Wolf PA, Beiser A, et al. Stroke risk profile predicts white matter hyperintensity volume: the Framingham Study. *Stroke* 2004; **35**(8): 1857–1861.

9. Rost NS, Rahman R, Sonni S, et al. Determinants of white matter hyperintensity volume in patients with acute ischemic stroke. *J Stroke Cerebrovasc Dis* 2010; **19**(3): 230–235.

10. Fazekas F, Chawluk JB, Alavi A, et al. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR Am J Roentgenol* 1987; **149**(2): 351-356.

11. Gurol ME, Irizarry MC, Smith EE, et al. Plasma beta-amyloid and white matter lesions in AD, MCI, and cerebral amyloid angiopathy. *Neurology* 2006; **66**(1): 23-239.

12. Dadar M, Maranzano J, Misquitta K, et al. Performance comparison of 10 different classification techniques in segmenting white matter hyperintensities in aging. *NeuroImage* 2017; **157**: 233-249.

13. Giese AK, Schirmer MD, Donahue KL, et al. Design and rationale for examining neuroimaging genetics in ischemic stroke: The MRI-GENIE study. *Neurology Genetics* 2017; **3**(5): e180.

14. Meschia JF, Arnett DK, Ay H, et al. Stroke Genetics Network (SiGN) study: design and rationale for a genome-wide association study of ischemic stroke subtypes. *Stroke* 2013; **44**(10): 2694-2702.

15. Schirmer MD, Dalca AV, Sridharan R, et al., White Matter Hyperintensity Quantification in Large-Scale Clinical Acute Ischemic Stroke Cohorts - The MRI-GENIE Study. *NeuroImage: Clinical* 2019; 23: 101884.

16. Dalca AV, Sridharan R, Cloonan L, et al. Segmentation of cerebrovascular pathologies in stroke patients with spatial and shape priors. *Med Image Comput Comput Assist Interv* 2014; **17**(Pt 2): 773-780.

17. Dalca AV, Guttag J, Sabuncu MR. Anatomical Priors in Convolutional Networks for Unsupervised Biomedical Segmentation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2018: 9290-9299.

18. Meschia JF, Worrall BB, Brown RD, et al. NINDS stroke genetics network (SiGN) experience with the causative classification system. *International Journal of Stroke : Official Journal of the International Stroke Society* 2013; **8**(4): E9.

19. Adams HP Jr, Bendixen BH, Kappelle LJ, et al. 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**(1): 35-41.

20. Fornage M, Debette S, Bis JC, et al. Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. *Ann Neurol* 2011; **69**(6): 928-939.

21. Verhaaren BF, Debette S, Bis JC, et al. Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ Cardiovasc Genet* 2015; **8**(2): 398-409.

22. Adib-Samii P, Rost N, Traylor M, et al. 17q25 Locus is associated with white matter hyperintensity volume in ischemic stroke, but not with lacunar stroke status. *Stroke* 2013; **44**(6): 1609-1615.

23. Zhang CR, Cloonan L, Fitzpatrick KM, et al. Determinants of white matter hyperintensity burden differ at the extremes of ages of ischemic stroke onset. *J Stroke Cerebrovasc Dis* 2015; **24**(3): 649-654.

24. Cloonan L, Fitzpatrick KM, Kanakis AS, et al. Metabolic determinants of white matter hyperintensity burden in patients with ischemic stroke. *Atherosclerosis* 2015; **240**(1): 149-153.

25. Shen Y, Zhao B, Yan L, et al. Cerebral Hemodynamic and White Matter Changes of Type 2 Diabetes Revealed by Multi-TI Arterial Spin Labeling and Double Inversion Recovery Sequence. *Front Neurol* 2017; **8**: 717.

26. Power MC, Deal JA, Sharrett AR, et al. Smoking and white matter hyperintensity progression: the ARIC-MRI Study. *Neurology* 2015; **84**(8): 841-848.

27. Fernando MS, Simpson JE, Matthews F, et al. White matter lesions in an unselected cohort of the elderly: molecular pathology suggests origin from chronic hypoperfusion injury. *Stroke* 2006;**37**(6):1391-1398.

28. de Groot M, Verhaaren BF, de Boer R, et al. Changes in normal-appearing white matter precede development of white matter lesions. *Stroke* 2013; **44**(4): 1037-1042.

29. de Groot M, Ikram MA, Akoudad S, et al. Tract-specific white matter degeneration in aging: the Rotterdam Study. *Alzheimers Dement* 2015; **11**(3): 321-330.

30. Etherton MR, Wu O, Cougo P, et al. Integrity of normal-appearing white matter and functional outcomes after acute ischemic stroke. *Neurology*. 2017; **88**(18): 1701-1708.

31. Wardlaw JM, Chappell FM, Valdés Hernández MDC, et al. White matter hyperintensity reduction and outcomes after minor stroke. *Neurology* 2017; **89**(10): 1003-1010.

32. Kim GM, Park KY, Avery R, et al. Extensive leukoaraiosis is associated with high early risk of recurrence after ischemic stroke. *Stroke* 2014; **45**(2): 479-485.

33. Ay H, Arsava EM, Andsberg G, et al. Pathogenic ischemic stroke phenotypes in the NINDS-stroke genetics network. *Stroke* 2014 ;**45**(12): 3589-3596.

34. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke* 2007; **38**(11): 2979-2984.

Table 1: Basic demographics of the MRI-GENIE cohort and comparison by stroke subtype

	ALL	LAA	CE major	SAO	OTHER	UNDETERMINED	p
N**	2529	545	394	387	181	1022	
Female (%)	993 (39.3)	176 (32.3)	186 (47.2)	146 (37.7)	81 (44.8)	404 (39.5)	<0.001
Age (mean (sd))	63.4 (14.6)	65.6 (12.3)	71.8 (11.9)	62.7 (13.5)	49.0 (13.6)	61.9 (15.0)	<0.001
Caucasian (%)	2141 (84.7)	467 (85.7)	354 (89.8)	288 (74.4)	147 (81.2)	885 (86.6)	<0.001
Atrial Fibrillation (%)	380 (15.2)	30 (5.6)	267 (68.8)	12 (3.1)	9 (5.0)	62 (6.1)	<0.001
CAD (%)	444 (17.9)	110 (20.6)	111 (28.6)	48 (12.7)	18 (10.2)	157 (15.6)	<0.001
Diabetes Mellitus (%)	581 (23.2)	146 (26.9)	96 (24.6)	107 (28.1)	27 (15.1)	205 (20.3)	<0.001
Hypertension (%)	1668 (66.4)	391 (72.1)	288 (73.3)	274 (71.5)	87 (48.9)	628 (61.9)	<0.001
Prior Stroke (%)	248 (9.8)	52 (9.6)	46 (11.7)	31 (8.1)	10 (5.6)	109 (10.7)	0.109
Smoking (ever, %)	1323 (54.1)	336 (63.6)	182 (48.3)	214 (56.3)	86 (48.3)	505 (51.5)	<0.001
WMHv in cm^3 (median [IQR])	5.86 [2.18, 14.61]	5.76 [2.48, 14.42]	8.13 [3.65, 17.12]	7.53 [2.84, 18.45]	2.16 [0.93, 5.29]	5.14 [1.94, 13.17]	<0.001

Abbreviations: CAD – coronary artery disease, CE major – cardioembolic major, LAA – large artery atherosclerosis, SAO – small artery occlusion, WMHv – white matter hyperintensity volume

*Statistical comparison was performed across the CCS 5-item subtypes. χ^2- test was used to compare categorical data, ANOVA was used for age and Kruskal-Wallis-Test was used for WMHv.

** Missing cases: Atrial fibrillation – 30, CAD – 43, diabetes mellitus – 24, hypertension – 18, prior stroke -11, smoking status – 85.

Table 2: Univariable and multivariable predictors of WMH

	Univariable Model		Multivariable Model*	
Variable	**Estimate (95% CI)**	**P**	**Estimate (95% CI)**	**P**
Age	**0.05 (0.05, 0.05)**	**<0.001**	**0.05 (0.04, 0.05)**	**<0.001**
Male	0.06 (-0.05, 0.17)	0.257	-	
Caucasian	0.02 (-0.13, 0.16)	0.836	-	
Atrial Fibrillation	**0.46 (0.31, 0.61)**	**<0.001**	-0.09 (-0.22, 0.04)	0.185
CAD	**0.41 (0.27, 0.55)**	**<0.001**	-0.09 (-0.20, 0.04)	0.159
Diabetes Mellitus	**0.35 (0.23, 0.48)**	**<0.001**	**0.11 (0.03, 0.24)**	**0.041**
Hypertension	**0.82 (0.72, 0.93)**	**<0.001**	**0.35 (0.27, 0.47)**	**<0.001**
Prior Stroke	**0.55 (0.37, 0.72)**	**<0.001**	**0.45 (0.32, 0.63)**	**<0.001**
Smoking (ever)	**0.13 (0.03, 0.24)**	**0.015**	**0.19 (0.10, 0.28)**	**<0.001**

*linear mixed effects model adjusted for site as a random variable

Abbreviations: CAD – coronary artery disease, WMHv – white matter hyperintensity volume, 95% CI – 95% confidence interval

Table 3: Predictors of WMHv by CCS Subtype

Variable	LAA (n=545) Estimate (95% CI) [p]	CE major (n=394) Estimate (95% CI) [p]	SAO (n=387) Estimate (95% CI) [p]	Other (n=181) Estimate (95% CI) [p]	Undetermined (n=1022) Estimate (95% CI) [p]
Age	**0.05 (0.04, 0.05) [p <0.001]**	**0·.5 (0.04, 0.05) [p<0.001]**	**0.04 (0.03, 0.05) [p<0.001]**	**0.04 (0.03, 0.06) [p<0.001]**	**0·05 (0·04, 0·05) [p<0.001]**
Female	-	-	-	**-0.37 (-0.72, -0.02) [p=0.040]**	-
Caucasian	-0.01 (-0.33, 0.30) [p=0.935]	-	-0.004 (-0.29, 0.28) [p=0.980]	-	-
Atrial Fibrillation	0.02(-0.40, 0.45) [p=0.018]	-	0.23 (-0.43, 0.90) [p=0.487]	-0.25 (-1.09, 0.58) [p=0.522]	0·03 (-0·26, 0·33) [p=0.826]
CAD	-0.009 (-0.25, 0.24) [p=0.940]	-	0.30 (-0.04, 0.64) [p=0.088]	0·34 (-0·32, 1·00) [p=0.314]	**-0·26 (-0·45, -0·05) [p=0.011]**
Diabetes Mellitus	-	-	-0.02 (-0.27, 0.24) [p=0.897]	0.05 (-0.46, 0.56) [p=0.849]	0·12 (-0·06, 0·30) [p=0.192]
Hypertension	-0.12 (-0.10, 0.33) [p=0.289]	0.15 (-0.07, 0.39) [p=0.181]	**0.43 (0.16, 0.70) [p=0.002]**	0.05 (-0.33, 0.42) [p=0.808]	**0·58 (0·42, 0·74) [p<0.001]**
Prior Stroke	**0.68 (0.34, 1.01) [p=0.001]**	-	**0.52 (0.11, 0.93) [p=0.012]**	**0.90 (0.13, 1.66) [p=0.021]**	**0·38 (0·15, 0·61) [p=0.001]**
Smoking (ever)	-	-	-	-	**0·23 (0·09, 0·37) [p=0.001]**

*linear mixed effects model adjusted for site as a random variable

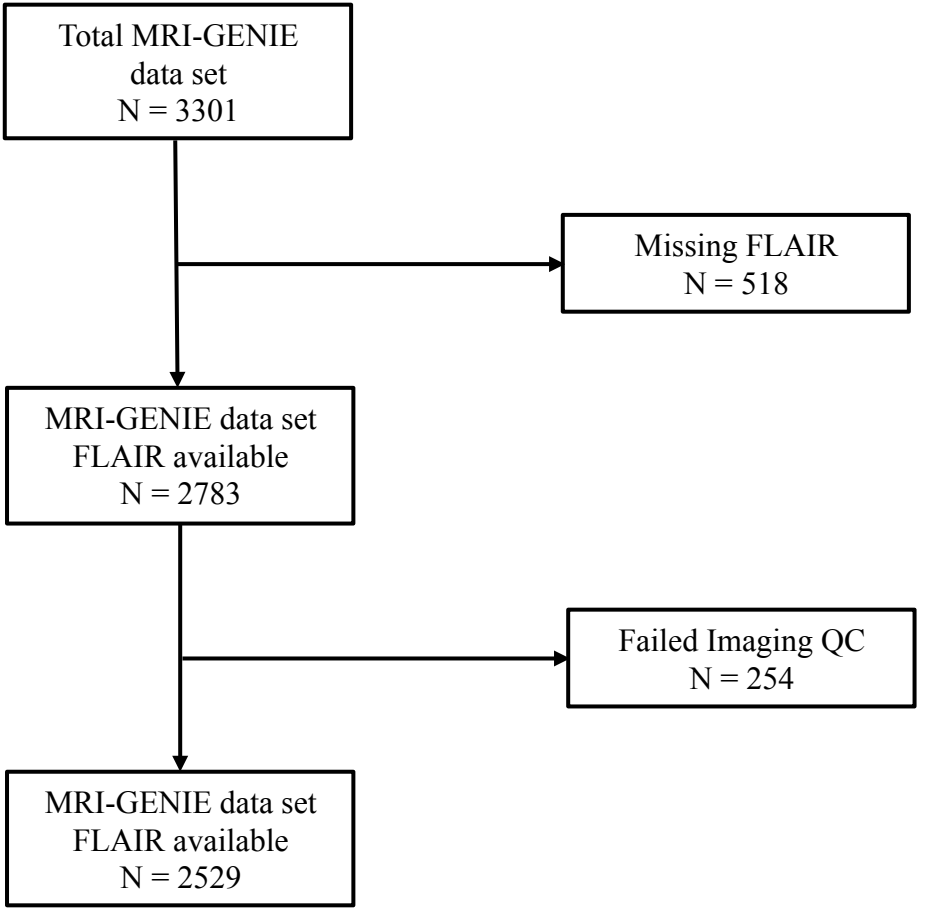
Abbreviations: CAD – coronary artery disease, CE major – major cardioembolic stroke, LAA – large artery atherosclerosis, WMHv – white matter hyperintensity volume, 95% CI – 95% confidence interval

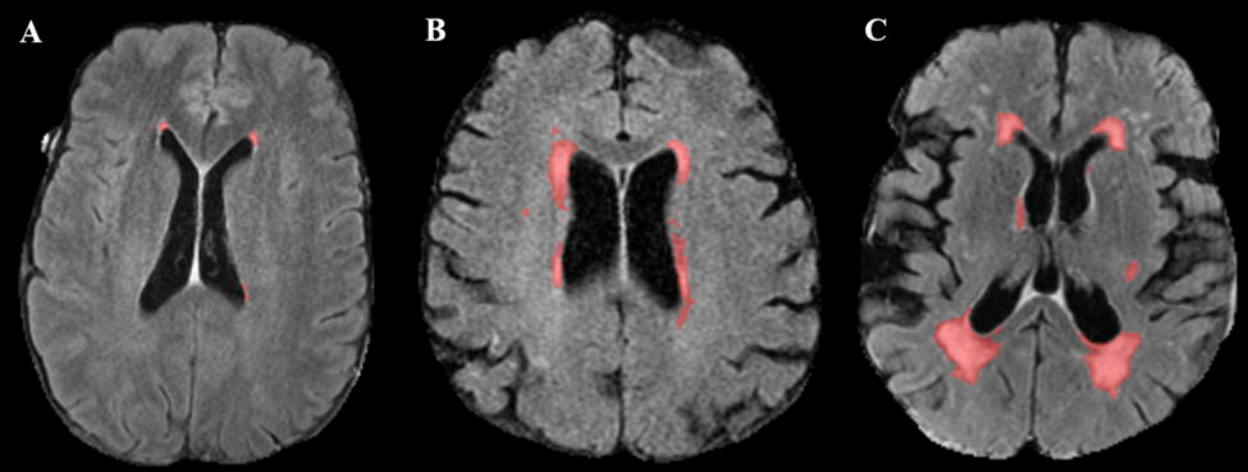
FIGURES

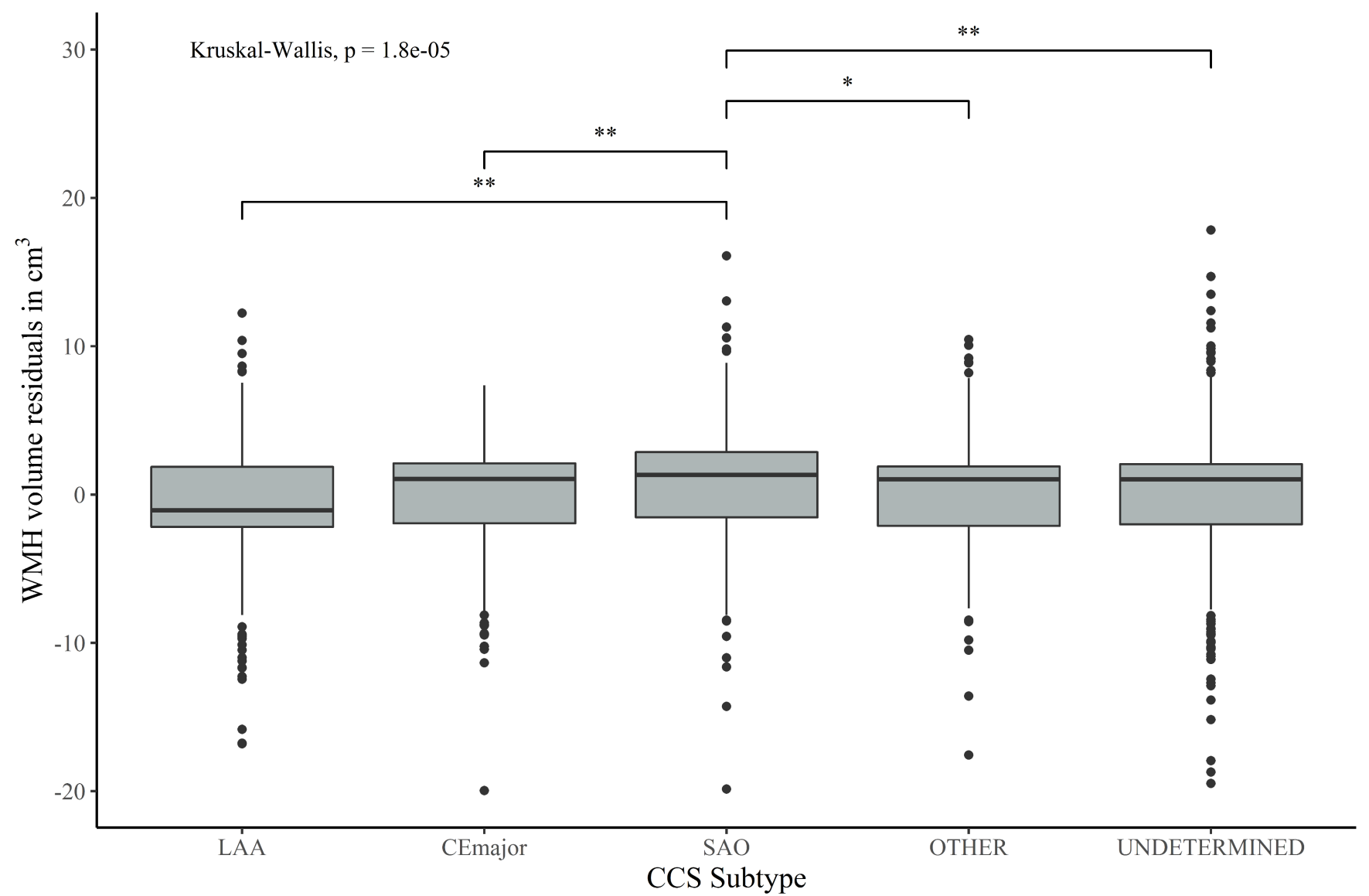
Figure 1: Flowchart of case selection for analysis

Figure 2: Automated WMH outline in ischemic stroke patients with mild (a), moderate (b) and severe (c) WMH burden. Results were extracted by the automated MRI-GENIE pipeline.

Figure 3: WMH by CCS subtype. Residuals of WMH volume adjusted for age, vascular risk factors and site. Cases with SAO have the highest WMH burden compared to other CCS subtypes (*$p<0.05$, **$p<0.005$).


Total MRI-GENIE
data set
N = 3301
Missing FLAIR
N = 518
MRI-GENIE data set
FLAIR available
N = 2783
Failed Imaging QC
N = 254
MRI-GENIE data set
FLAIR available
N = 2529


A
B
C



N	545	394	387	181	1022
Adjusted WMHv	-1.02	1.03	1.47	1.08	1.02
(median, IQR)	(-2.18-1.97)	(-2.10-2.03)	(-1.52-3.15)	(-2.31-1.99)	(-2.08-2.20)