Association of Baseline Metabolomic Profiles With Incident Stroke and Dementia and With Imaging Markers of Cerebral Small Vessel Disease

Author(s):
Eric L. Harshfield, PhD1,2; Hugh S Markus, DM FMed Sci1,2

Corresponding Author:
Eric L. Harshfield, eh457@medschl.cam.ac.uk

This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (CC BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Affiliation Information for All Authors: 1. Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom 2. Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, United Kingdom.

Equal Author Contribution:

Contributions:
Eric L. Harshfield: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data
Hugh S Markus: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data

Figure Count: 3

Table Count: 1
Search Terms:

- [2] All Cerebrovascular disease/Stroke
- [25] All Cognitive Disorders/Dementia
- [54] Cohort studies
- [118] All Imaging, metabolomics

Acknowledgment:
For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Study Funding:
This research was conducted using the UK Biobank under application number 36509. It was funded by the British Heart Foundation via a fellowship awarded to ELH from the Cambridge BHF Centre of Research Excellence (RE/18/1/34212) and a BHF program grant (RG/F/22/110052). Infrastructural support was provided by the Cambridge University Hospitals NIHR Biomedical Research Centre (BRC-1215-20014). HSM is supported by a NIHR Senior Investigator Award. The views expressed in this publication are those of the authors and not necessarily those of the NIHR, NHS, or UK Department of Health and Social Care. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclosure:
The authors report no disclosures relevant to the manuscript.
Abstract

Background and Objectives: Cerebral small vessel disease is a major cause of stroke and dementia. Metabolomics can help identify novel risk factors to better understand pathogenesis and predict disease progression and severity.

Methods: We analysed baseline metabolomics profiles from 118,021 UK Biobank participants. We examined cross-sectional associations of 325 metabolites with MRI markers of small vessel disease, evaluated longitudinal associations with incident stroke and dementia, and ascertained causal relationships using Mendelian randomization.

Results: In cross-sectional analyses, lower levels of apolipoproteins, free cholesterol, cholesteryl esters, fatty acids, lipoprotein particle concentrations, phospholipids, and triglycerides were associated with increased white matter microstructural damage on diffusion tensor MRI. In longitudinal analyses, lipoprotein subclasses of very large HDL were associated with increased risk of stroke, and acetate and 3-hydroxybutyrate were associated with increased risk of dementia. Mendelian randomization analyses identified strong evidence supporting causal relationships for many findings. A few metabolites had
consistent associations across multiple analysis types. Increased total lipids in very large HDL and increased HDL particle size were associated with increased white matter damage (Lower FA: OR: 1.44, 95% CI: 1.07-1.95, and OR: 1.19, 95% CI: 1.06-1.34, respectively; MD: OR: 1.49, 95% CI: 1.11-2.01, and OR: 1.24, 95% CI: 1.11-1.40, respectively) and increased risk of incident all stroke (HR: 4.04, 95% CI: 2.13-7.64, and HR: 1.54, 95% CI: 1.20-1.98, respectively) and ischaemic stroke (HR: 3.12, 95% CI: 1.53-6.38; HR: 1.37, 95% CI: 1.04-1.81). Valine was associated with decreased MD (OR: 0.51, 95% CI: 0.30-0.88) and had a protective association with all-cause dementia (HR: 0.008, 95% CI: 0.002-0.035). Cholesterol in small HDL had an inverse association with incident all stroke (HR: 0.17, 95% CI: 0.08-0.39) and ischaemic stroke (HR: 0.19, 95% CI: 0.08-0.46) that was supported by evidence of a causal association with MRI-confirmed lacunar stroke (OR: 0.96, 95% CI: 0.93-0.99).

Discussion: In this large-scale metabolomics study, we found multiple metabolites associated with stroke, dementia, and MRI markers of small vessel disease. Further studies may help inform development of personalised prediction models and provide insights into mechanistic pathways and future treatment approaches.

Keywords: metabolomics; stroke; dementia; MRI; small vessel disease; Mendelian randomization; UK Biobank

1. Introduction

There are 6.3 million deaths due to stroke and 1.9 million deaths due to dementia each year, and 50 million people currently living with dementia. In order to develop potential new treatments, we need improved techniques for identifying novel biomarkers that can predict disease outcomes and characterise the underlying mechanisms.

While increasing evidence implicates vascular risk factors and chronic cerebrovascular disease in the pathogenesis of vascular dementia, these factors also increase the risk of neurodegenerative dementias such as Alzheimer’s disease (AD). Recent data have suggested treating vascular risk factors, especially hypertension, may reduce overall dementia risk. This emphasises the potential for targeting cerebrovascular disease to reduce the burden of all types of dementia. Even with small treatment effects this could have a major global impact.
Cerebral small vessel disease (SVD), a subtype of cerebrovascular disease that appears closely linked to dementia risk, affects the small perforating vessels within the white matter and deep grey matter nuclei. The consequences of SVD include lacunar infarcts and more chronic changes detected on MRI, including white matter hyperintensities (WMH), cerebral microbleeds, enlarged perivascular spaces, brain atrophy, and diffuse white matter damage identified using diffusion tensor imaging (DTI).\(^5\) SVD is a major cause of vascular dementia, and it also interacts with AD pathology to increase the probability of developing clinical dementia.\(^6\) Therefore it represents an important treatment target not only to reduce vascular dementia, but also to reduce the impact of neurodegenerative dementias such as AD. MRI changes indicative of SVD are common in the general population with increasing age, and predict both stroke and dementia.\(^5\)

Metabolomics, the detailed quantification of small metabolic markers in biological samples,\(^7\) can be used to identify novel biomarkers to diagnose and monitor disease and characterise metabolic pathways underlying disease pathogenesis.\(^8\) Previous metabolomics studies in cardiovascular and dementia research have been relatively small.\(^9,10\) Studies of SVD patients have reported that plasma ceramide levels are associated with increased cerebral microbleeds, that numerous metabolites are associated with increased WMH volume, and that many glycerophospholipids, sphingolipids, and other metabolites are associated with MRI markers of SVD, cognition, and conversion to dementia.\(^11–13\) However, larger studies are required to validate and extend these findings, and these may lead to better understanding of the underlying mechanistic pathways for the development of clinical biomarkers and the identification of potential targets for intervention. In this analysis we analysed metabolomics profiles from 118,021 participants in the UK Biobank to characterise cross-sectional associations of 325 metabolites with MRI markers of SVD and longitudinal associations with future clinical diagnoses of stroke and dementia. We also used Mendelian randomization to evaluate whether these relationships are likely to be causal.
2. Methods

2.1 Data source

UK Biobank is a prospective cohort study of over 500,000 participants recruited from 22 centres throughout the United Kingdom. Participants were aged between 40-69 at baseline assessment in 2006-2010. Participants took part in a comprehensive questionnaire and verbal interview and provided blood samples for metabolomics assays and genetic analyses.

2.2 Metabolomics data

Metabolic biomarker profiling of EDTA plasma samples was performed at Nightingale Health’s laboratories in Finland from a random subset of 121,695 non-fasting participants, of which 118,021 were from baseline recruitment (2006-2010). High-throughput nuclear magnetic resonance (NMR) spectroscopy was used to obtain 325 metabolic measures, including 168 metabolites measured in absolute levels, 81 composite biomarkers and ratios provided by UK Biobank, and a further 76 biomarker ratios that we derived separately. The platform included a wide range of biomarkers including amino acids, apolipoproteins, cholesterol and cholesteryl esters, fatty acids, phospholipids, triglycerides, and detailed measurements of lipid concentrations and compositions within 14 different lipoprotein subclasses. Details of the sample processing and biomarker quantification have been described in the UK Biobank NMR metabolomics companion document (https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/nmrm_companion_doc.pdf).

To account for technical variation in metabolite levels, a multi-step processing procedure was applied as described previously. In brief, for the 168 biomarkers that were measured in absolute levels, the concentrations were log transformed and adjusted in a robust linear regression model that accounted for the time between sample preparation and sample measurement, systematic differences between rows and columns on the 96-well shipment plates, and drift over time within each of the six spectrometers. The 81 composite biomarkers and biomarker ratios provided by UK Biobank were recalculated from their adjusted parts, and an additional 76 biomarker ratios of potential biological significance were derived, resulting in a total of 325 metabolites.
2.3 Clinical and imaging endpoints

In the full set of UK Biobank participants with metabolomics data at baseline (n=118,021), we examined clinical endpoints for all stroke, ischaemic stroke, intracerebral haemorrhage, all-cause dementia, AD, and vascular dementia, using algorithmically-defined outcomes. Incident stroke and stroke subtypes and all-cause dementia were defined based on the earliest recorded date that the outcome occurred after baseline assessment, either from self-report at nurse interview, hospital admission records in the primary or secondary position, or death certificate records in the underlying cause or any other position. Incident AD and vascular dementia were based on hospital admission or death certificate records only. Identification of linked hospital admission and death certificate records for each endpoint was based on corresponding ICD-9 or ICD-10 codes (eTable 2).

In participants with metabolomics data for whom MRI had been performed (n=10,024), we examined white matter hyperintensities (WMH) volume and two DTI metrics of white matter tracts from the earliest available imaging assessment. For WMH volume we used a UK Biobank-derived phenotype, the total volume of WMH from T1 and T2 FLAIR images (measured in cubic millimetres),\(^{18}\) which we log-transformed for analysis. To obtain the DTI metrics, we performed principal component analyses on UK Biobank-derived variables for 48 markers of both mean diffusivity (MD, the degree of diffusion) and fractional anisotropy (FA, the directionality of diffusion) on the FA skeleton of the diffusion brain MRI data, and selected the first principal component of each as summary measures of MD and FA.\(^ {19}\) We rescaled the values for the three imaging markers across participants using meaning-centring and dividing by the standard deviation to ensure that the effect sizes across outcomes were comparable.

2.4 Cross-sectional and longitudinal analyses

We examined cross-sectional associations of MRI endpoints per 1-SD higher baseline metabolite levels using linear regression models adjusted for age at recruitment and sex. We also constructed regression models adjusted for a wide variety of possible confounders and vascular risk factors. These analyses were adjusted for age at recruitment, sex, UK Biobank recruitment centre, NMR spectrometer, Townsend deprivation index at recruitment (a simple census-based measure of material deprivation), taking blood pressure medications or statins at recruitment, body mass index at recruitment, smoking status at recruitment, and type 2
diabetes mellitus status (based on verbal interview, touchscreen self-report, or linked EHR or death register records). Additionally, we conducted sensitivity analyses with the above adjustments as well as further adjustment for ethnicity and APOE ε4 genotype.

We also evaluated in longitudinal analyses whether baseline metabolites predicted long-term progression to stroke and dementia and their subtypes. We constructed Cox proportional-hazards regression models adjusted for age at recruitment and sex to assess the association of progression to stroke and dementia per 1-SD higher metabolite levels. These analyses were likewise conducted with adjustment for possible confounders and vascular risk factors, as listed above.

2.5 Mendelian randomization analyses

We performed two-sample Mendelian randomization, which involves the use of genetic variants as instrumental variables in a technique that is analogous to a randomised trial, to assess whether the association of metabolites with stroke, dementia, and imaging markers were likely to be causal. The three main assumptions of Mendelian randomization are that the instrumental variables (1) are associated with the exposure, (2) are not impacted by confounding that would influence their association with the outcome, and (3) affect the outcome only through the exposure and do not affect any other trait with a downstream effect on the outcome. We used published genome-wide association study summary statistics for each metabolite measured in 115,078 participants from UK Biobank, conducted by the MRC Integrative Epidemiology Unit at the University of Bristol. We obtained summary statistics for late-onset AD from a genome-wide association study involving 35,274 cases and 59,163 controls from the International Genomics of Alzheimer's Project (IGAP), and for MRI-confirmed lacunar stroke from a genome-wide association study involving 7,338 cases and 254,798 controls. Summary statistics for all stroke and ischaemic stroke subtypes were obtained from individuals of European ancestry from the GIGASTROKE Consortium, consisting of 73,652 cases (62,100 cases of ischaemic stroke; 10,804 cases of cardioembolic stroke; 6,399 cases of large artery stroke) and 1,234,808 controls. Summary statistics from a genome-wide association study of participants from UK Biobank and the CHARGE Consortium were obtained for WMH (n=42,310), MD (n=17,467), and FA (n=17,663). To ensure that the genetic instruments were independent, we performed clumping of each exposure dataset on all bi-allelic SNPs with a minor allele frequency > 0.01 using a physical distance threshold of 10,000 kb and an r² threshold of 0.001. We also harmonised the
exposure and outcome datasets to ensure that the effect size of each SNP on an outcome and exposure were recorded relative to the same effect allele. Our primary Mendelian randomization analyses used inverse-variance weighted meta-analysis to combine the ratio estimates from each genetic variant into a single estimate of the causal effect of each metabolite on each outcome, using a random effects model to account for heterogeneity.26 We conducted sensitivity analyses using a variety of robust Mendelian randomization methods, which employ different assumptions to make reliable causal inferences. These included MR-Egger regression, the weighted median estimator, and the simple and weighted mode-based estimators. As a further sensitivity analysis we used the Mendelian randomization pleiotropy residual sum and outlier (MR-PRESSO) method to detect and account for outliers, which consists of three components: the global test, which is used to detect horizontal pleiotropy; the outlier test, which is used to correct for horizontal pleiotropy via outlier removal; and the distortion test, which is used to test for significant differences in the causal estimates before and after correction for outliers.27

All statistical analyses were conducted using R version 4.1.1 (R Core Team, 2021) and results are presented using two-sided $P$-values and 95% confidence intervals. We used a false discovery rate (FDR) threshold of $q < 0.05$ to account for multiple testing comparisons when identifying significant associations for each outcome measure. For comparison purposes, we also calculated whether the $P$-values were significant when applying a more conservative Bonferroni correction adjusted for the number of metabolites analysed. Mendelian randomization analyses were conducted using the TwoSampleMR package version 0.5.4 and the MRPRESSO package version 1.0.

### 2.6 Standard Protocol Approvals, Registrations, and Patient Consents

All UK Biobank participants provided informed consent as part of the UK Biobank recruitment process to the use of their anonymised data and samples for any health-related research, to be recontacted for further substudies, and for UK Biobank to access their electronic health records. UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval. This research was conducted using UK Biobank under application number 36509.
2.7 Data Availability

The data supporting the findings of this study are available within the article and its Supplemental Materials. The raw metabolomics data and phenotypic information can be accessed by approved researchers through application to UK Biobank (https://www.ukbiobank.ac.uk/enable-your-research). The summary statistics from the genome-wide association studies of metabolites in UK Biobank can be accessed from the MRC Integrative Epidemiology Unit (IEU) OpenGWAS Project (https://gwas.mrcieu.ac.uk/), the summary statistics for AD can be obtained from the International Genomics of Alzheimer's Project (IGAP) (https://www.niagads.org/datasets/ng00075), and the summary statistics for stroke from the GIGASTROKE Consortium can be obtained from the GWAS Catalog (https://www.ebi.ac.uk/gwas/; study accession numbers GCST90104534–GCST90104563).

3. Results

3.1 Participant characteristics

We analysed data from 118,021 participants from UK Biobank with metabolomics measurements, of whom 54% were female and 95% were white, with a mean age of 56.5 (SD: 8.1) years (Table 1). The median follow-up time was 13.1 years (5th-95th percentile, 11.8–14.4 years). The characteristics of UK Biobank participants whose blood samples were randomly selected for the metabolomics assay to be performed were broadly comparable to those that were not selected (eTable 3).

3.2 Cross-sectional associations with imaging parameters

We analysed the association of 325 baseline metabolic measures with MRI markers at the first imaging assessment in cross-sectional analyses adjusted for age at recruitment and sex (eFigure 1, eTable 4). Lower levels of apolipoproteins, cholesterol, free cholesterol, cholesteryl esters, fatty acids, lipoprotein particle concentrations, phospholipids, triglycerides, and total lipids, and higher levels of amino acids, glucose, and glycoprotein acetyl (an inflammatory marker) were associated with increased white matter microstructural damage on DTI, as indicated by higher WMH and MD and lower FA.
In cross-sectional analyses adjusted for possible confounders and additional vascular risk factors, most metabolites remained significantly associated with FA and MD (Figure 1, eTable 5). Lower levels of apolipoproteins, cholesterol, free cholesterol, cholesteryl esters, fatty acids, lipoprotein particle concentrations, phospholipids, triglycerides, and total lipids were associated with higher MD and lower FA on DTI, indicating increased white matter microstructural damage. However, increased total lipids in very large HDL (XL_HDL_L) and increased HDL particle size (HDL_size) were associated with lower FA (OR: 1.44, 95% CI: 1.07-1.95, and OR: 1.19, 95% CI: 1.06-1.34, respectively) and higher MD (OR: 1.49, 95% CI: 1.11-2.01, and OR: 1.24, 95% CI: 1.11-1.40, respectively). Higher levels of HDL cholesterol (HCL_C) and lipoproteins within large HDL and very large HDL were also associated with increased white matter microstructural damage.

### 3.3 Longitudinal analyses of stroke and dementia

In time-to-event analyses adjusted for age and sex, which took into account long-term follow-up (eFigure 2, eTable 6), higher levels of the ratio of free cholesterol to cholesteryl esters in very small VLDL (XS_VLDL_FC_by_CE) and higher levels of glycoprotein acetylts (GlycA) were associated with increased risk of all stroke. Conversely, higher levels of docosahexaenoic acid (DHA) and several lipoprotein subclasses containing cholesterol, free cholesterol, and cholesteryl esters were associated with decreased risk of all stroke. Higher levels of glucose and the percentage of phospholipids to total lipids in very small VLDL (XS_VLDL_PL_pct) were associated with increased risk of all-cause dementia, while a wide range of cholesterol, cholesteryl esters, free cholesterol, apolipoprotein B, and lipoprotein subclasses were associated with decreased risk of all-cause dementia.

In longitudinal analyses adjusted for potential confounders and vascular risk factors (Figure 2, eTable 7), total lipids in very large HDL (XL_HDL_L) were most strongly associated with all stroke (HR: 4.04, 95% CI: 2.13-7.64), followed by HDL particle size (HR: 1.54, 95% CI: 1.20-1.98), while cholesterol, cholesteryl esters, phospholipids, and total lipids in small HDL (S_HDL_C, S_HDL_CE, S_HDL_PL, and S_HDL_L, respectively) had inverse associations with all stroke. Acetate, 3-hydroxybutyrate (bOHbutyrate), and the ratio of free cholesterol to cholesteryl esters in small LDL (S_LDL_FC_by_CE) were associated with increased risk of all-cause dementia. Meanwhile, omega-3 fatty acids, DHA, leucine, isoleucine, valine, the total concentration of branched-chain amino acids (Total_BCAA), and several lipoprotein subclasses were associated with decreased risk of all-cause dementia.
Very few metabolites were significantly associated with stroke and dementia subtypes in analyses adjusted for potential confounders and vascular risk factors (eFigure 3, eTable 7). The main exceptions were that decreased levels of lipoprotein subclasses containing cholesterol, cholesteryl esters, phospholipids, and total lipids in small HDL, and increased levels of total lipids in very large HDL (XL_HDL_L) and the proportion of free cholesterol to total lipids in HDL, small HDL, and large HDL (HDL_FC_pct, S_HDL_FC_pct, and L_HDL_FC_pct, respectively), were associated with increased risk of ischaemic stroke. Additionally, increased levels of valine, leucine, and total branched-chain amino acids (Total_BCAA) were associated with decreased risk of AD.

In sensitivity analyses with further adjustment for ethnicity and APOE ε4 genotype, the results for all stroke and all-cause dementia were very similar and broadly comparable (eFigure 4, eTable 8).

### 3.4 Mendelian randomization analyses

Our Mendelian randomization analyses showed that genetically elevated levels of cholesteryl esters in very large VLDL and total lipids in small VLDL were associated with increased risk of ischaemic stroke, and genetically lowered levels of the concentration of HDL particles were associated with increased risk of lacunar stroke (Figure 3, eTables 9 and 10). Furthermore, genetically elevated levels of LDL within cholesterol, cholesteryl esters, free cholesterol, and phospholipids, and lipoprotein subclasses of LDL, VLDL, and IDL, and genetically lower levels of total lipids in medium HDL, were associated with increased risk of late-onset AD. The MR-PRESSO global test revealed the presence of horizontal pleiotropy that was statistically significant (FDR q < 0.05) in 70.6% (n=1,741) of the 2,465 tests (eTable 11). The MR-PRESSO outlier test identified false-positive relationships in 2.8% of putatively causal relationships (n=1,295 out of 46,857 total tests) using a threshold of P < 0.05 (eTable 12). The MR-PRESSO distortion test revealed that horizontal pleiotropy introduced distortions in the causal estimates with a median of 24.8% (5th-95th percentile, -356%–1,758%) (Figure 5, eTable 13). The MR-PRESSO outlier-corrected results are provided in eTable 14.
4. Discussion

In this large-scale metabolomics study of 118,021 individuals, we identified 289 metabolic markers significantly associated with MRI markers of SVD and future risk of stroke and dementia in a large population-based cohort. We found that lower levels of apolipoproteins, cholesterol, free cholesterol, cholesteryl esters, fatty acids, lipoprotein particle concentrations, phospholipids, triglycerides, and total lipids were associated with increased white matter microstructural damage on DTI. Additionally, lower levels of amino acids and fatty acids and higher levels of ketone bodies were associated with increased risk of all-cause dementia. Increased levels of lipoprotein subclasses within large HDL and very large HDL were associated with increased white matter microstructural damage and increased risk of all stroke and all-cause dementia, whereas lipoprotein subclasses within small and medium HDL were associated with decreased white matter microstructural damage and decreased risk of all stroke and all-cause dementia. We did not observe associations of metabolites with stroke and dementia subtypes, but the analyses may have lacked sufficient power due to the fewer number of cases for these outcomes.

A few metabolites had consistent associations across multiple types of analyses (cross-sectional, longitudinal, and/or Mendelian randomization), enabling a greater level of confidence in the importance of the findings. Increased total lipids in very large HDL and increased HDL particle size at baseline were significantly associated with increased white matter microstructural damage on DTI at the imaging assessment and also significantly associated with increased risk of incident all stroke and ischaemic stroke, but there was not sufficient evidence to support a causal relationship. Valine was associated with decreased MD and had a protective association with all-cause dementia, but it also lacked evidence of causality. Cholesterol in small HDL, however, had an inverse association with incident all stroke and ischaemic stroke that was supported by evidence of a causal association with MRI-confirmed lacunar stroke in the Mendelian randomization analyses. We lacked information on ischaemic stroke subtypes in the observational data, but the association of cholesterol in small HDL appears to be specific to lacunar stroke.

Another notable finding was that metabolites were much more strongly associated with DTI markers than with WMH lesion volume. This is consistent with previous studies showing DTI is more sensitive to white matter damage in SVD, and that DTI parameters correlate with cognition more strongly than WMH lesion volume.28–30
Although there is a well-established inverse association of plasma HDL cholesterol levels with coronary heart disease, stroke, and vascular brain damage, whether these relationships are causal has remained uncertain.\textsuperscript{31–38} This study provides new insights into these relationships by demonstrating that the direction and magnitude of the association of HDL with stroke and dementia and their subtypes depends on the size of the lipoprotein subclasses within HDL.

Many of the cross-sectional associations of metabolites with white matter microstructural damage were not in the direction that would normally be expected. Lower levels of LDL, VLDL, and total cholesterol, and cholesteryl esters, free cholesterol, and triglycerides in LDL and VLDL, and cholesterol, free cholesterol, cholesteryl esters, triglycerides, and phospholipids within small, medium, and large LDL, within IDL, and within very small, small, medium, large, very large, and extremely large VLDL were associated with lower FA and increased MD. However, the associations of these metabolites with stroke were not statistically significant and there was limited evidence from the Mendelian randomization analyses that these associations may be causal.

Our metabolite associations for all-cause dementia, AD, and vascular dementia confirm the findings reported in a recent preprint.\textsuperscript{39} However, we analysed an expanded set of metabolites by including 76 additionally derived biomarkers of potential biological significance, and we evaluated associations with a wider range of endpoints including all stroke, ischaemic stroke, intracerebral haemorrhage, and DTI markers.

Metabolic dysfunction is believed to play an important role in cognitive decline and AD progression.\textsuperscript{40} In AD the brain loses its ability to effectively use glucose, the primary energy substrate in the brain, so ketone bodies may be an effective alternative energy substrate.\textsuperscript{41} Previous studies have shown that elevated levels of 3-hydroxybutyrate, a ketone body also known as $\beta$-hydroxybutyrate, is associated with improved cognition and reduced risk of AD.\textsuperscript{41,42} However, our study showed that elevated levels of 3-hydroxybutyrate were associated with increased risk of all-cause dementia and AD (although the association with AD was not statistically significant after correction for multiple testing), in concordance with the recent preprint.\textsuperscript{39}

Mendelian randomization analyses identified evidence of causal associations of multiple metabolites with stroke, dementia, and MRI imaging markers. However, we detected substantial horizontal pleiotropy, meaning that some genetic variants were associated with
multiple metabolites on different pathways, which is unsurprising due to strong correlations between metabolites. MR-PRESSO indicated the presence of significant horizontal pleiotropy in 68% of the metabolite-outcome associations tested, which introduced substantial distortions in the causal estimates. This makes it difficult to assess whether the identified causal relationships are genuine or spurious and introduces challenges in drawing meaningful causal inferences, so the findings should be interpreted cautiously.

There are several important clinical implications arising from these findings. First, this research provides new insights into the metabolic pathways underlying stroke and dementia. This could be used to help inform dietary interventions or the development of novel therapies, since modifying levels of specific metabolites might help reduce the risk of vascular-related conditions. Second, clinicians might be able to use these findings to predict which patients are most likely to develop stroke and dementia and offer personalized treatment plans.

This study has several strengths. First, we had metabolomics data on nearly 120,000 individuals, which greatly increases the power to detect statistically significant associations. Second, a fully automated, comprehensive spectrum analysis was used to measure the metabolites under strict quality control, which increases the accuracy and validity of the findings. Third, the prospective study design and long follow-up period, with metabolites that were measured prior to disease onset, was particularly useful for evaluating their associations with risk of stroke and dementia. Fourth, we conducted sensitivity analyses to assess whether potential confounders and vascular risk factors, including APOE ε4 genotype, impacted the identified associations.

There are also limitations. First, we conducted numerous statistical tests. Although we corrected for multiple testing by applying an FDR threshold to reduce the possibility of identifying false positives, this may have inadvertently led us to discount some biologically and clinically meaningful associations that did not reach the threshold for statistical significance. Second, although the sample size was large it is not representative of the overall UK population. Third, we did not have endpoints for ischaemic stroke subtypes in UK Biobank (e.g. large artery, cardioembolic, and lacunar stroke), so we were unable to examine direct observational associations of metabolites with lacunar stroke. However, our Mendelian randomization analyses helped mitigate this by identifying evidence that genetically elevated levels of metabolites are associated with increased risk of lacunar stroke. Fourth, we obtained...
genome-wide association summary statistics with late-onset AD but not with all-cause dementia and vascular dementia, which limited the scope of our assessment of causal relationships of metabolites with dementia. Finally, a limitation of our two-sample Mendelian randomization analyses is that UK Biobank was used for the genetic associations with the metabolites but also contributed significantly to the datasets used for the genetic associations with the outcomes, which may contribute to bias due to overfitting, although owing to the large sample sizes of the respective studies the bias due to sample overlap is expected to be very small.44

In conclusion, our study provides evidence supporting the association of a wide range of metabolites with stroke, dementia, and MRI markers of SVD. Although further research is needed, these findings could help inform the development of personalised prediction models and novel treatment approaches.
References


### Table 1. Characteristics of UK Biobank participants with metabolomics data at baseline

<table>
<thead>
<tr>
<th>Trait</th>
<th>Total n</th>
<th>Mean (SD) or n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Risk factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female sex</td>
<td>117,938</td>
<td>63,890 (54.2%)</td>
</tr>
<tr>
<td>Age at recruitment (years)</td>
<td>118,021</td>
<td>56.5 (8.1)</td>
</tr>
<tr>
<td>White ethnicity</td>
<td>117,491</td>
<td>111,323 (94.8%)</td>
</tr>
<tr>
<td>Townsend deprivation index at recruitment</td>
<td>117,864</td>
<td>-1.3 (3.10)</td>
</tr>
<tr>
<td>Body mass index (BMI) (kg/m²)</td>
<td>117,609</td>
<td>27.4 (4.78)</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>117,902</td>
<td>137.8 (18.6)</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg)</td>
<td>117,903</td>
<td>82.2 (10.2)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>117,449</td>
<td>12,474 (10.6%)</td>
</tr>
<tr>
<td>Has Type 2 diabetes mellitus</td>
<td>118,020</td>
<td>11,277 (9.6%)</td>
</tr>
<tr>
<td>Has hypertension</td>
<td>117,902</td>
<td>53,999 (45.8%)</td>
</tr>
<tr>
<td>Taking blood pressure medication</td>
<td>117,079</td>
<td>24,305 (20.8%)</td>
</tr>
<tr>
<td>Taking statins</td>
<td>88,067</td>
<td>21,199 (24.1%)</td>
</tr>
<tr>
<td>APOE ε4 carrier</td>
<td>118,021</td>
<td>29,628 (25.1%)</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log white matter hyperintensities (WMH) volume</td>
<td>9,998</td>
<td>8.02 (1.01)</td>
</tr>
<tr>
<td>Fractional anisotropy (FA)</td>
<td>9,779</td>
<td>0.0085 (4.35)</td>
</tr>
<tr>
<td>Mean diffusivity (MD)</td>
<td>9,779</td>
<td>0.0379 (4.44)</td>
</tr>
<tr>
<td>Incident all stroke case</td>
<td>118,021</td>
<td>2,477 (2.1%)</td>
</tr>
<tr>
<td>Incident ischaemic stroke case</td>
<td>118,021</td>
<td>2,053 (1.7%)</td>
</tr>
<tr>
<td>Incident intracerebral haemorrhage case</td>
<td>118,021</td>
<td>464 (0.4%)</td>
</tr>
<tr>
<td>Incident all-cause dementia case</td>
<td>118,021</td>
<td>1,785 (1.5%)</td>
</tr>
<tr>
<td>Incident Alzheimer’s disease case</td>
<td>118,021</td>
<td>765 (0.6%)</td>
</tr>
<tr>
<td>Incident vascular dementia case</td>
<td>118,021</td>
<td>418 (0.3%)</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1. Association of MRI markers at baseline per 1-SD higher metabolite levels with adjustment for possible confounders and vascular risk factors. (A) Lipids and other metabolites. (B) Lipoprotein subclasses. (C) Relative lipid, lipoprotein, and cholesterol concentrations. Beta estimates and $P$-values were obtained from linear or logistic regression models adjusted for age at recruitment, sex, UK Biobank recruitment centre, Townsend deprivation index at recruitment, whether the person was taking blood pressure medication or statins, body mass index, smoking status, and Type 2 diabetes mellitus status. Colours show magnitude and direction of $P$-value for association of metabolite with each outcome (red indicates positive association and blue indicates inverse association). Asterisks indicate significance: *$P < 0.05$; **FDR $q < 0.05$. 
<table>
<thead>
<tr>
<th>Category/metabolite</th>
<th>VLDL</th>
<th>LDL</th>
<th>HDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acids</td>
<td>Val</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>ApoB by ApoA1</td>
<td>*</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>ApoB</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Apolipoproteins</td>
<td>VLDL C</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Total C</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Remnant C</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Non_HDL C</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>LDL C</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>HDL C</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Clinical LDL C</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Cholesteryl esters</td>
<td>VLDL CE</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LDL CE</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>HDL CE</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Fatty acids</td>
<td>Total FA</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>SFA</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>PUFA by MUFA</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>MUFA</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Omega_6</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>MUFA_pct</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LA</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Fluid balance</td>
<td>Creatinine</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>Free cholesterol</td>
<td>VLDL FC</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LDL FC</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>Lipoprotein particle concentrations</td>
<td>VLDL P</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Lipoprotein particle sizes</td>
<td>VLDL size</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Other lipids</td>
<td>TG by PG</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Phospholipids</td>
<td>VLDL PL</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LDL PL</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Total lipids</td>
<td>VLDL L</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Total L</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LDL L</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>HDL L</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>VLDL TG</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Total TG</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LDL TG</td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>
Figure 2. Adjusted hazard ratios for all stroke and all-cause dementia per 1-SD higher metabolite levels with adjustment for possible confounders and vascular risk factors. (A) All stroke. (B) All-cause dementia. Analyses were adjusted for age at recruitment, sex, UK Biobank recruitment centre, Townsend deprivation index at recruitment, whether the person was taking blood pressure medication or statins, body mass index, smoking status, and Type 2 diabetes mellitus status. Filled squares indicate associations significant at FDR $q < 0.05$. 

![Graph showing adjusted hazard ratios for metabolites]
Figure 3. Mendelian randomization results showing causal estimates for association of metabolite levels with stroke, dementia, and MRI markers. Colours show the magnitude and direction of the $P$-value of association for the estimate of the causal effect using Mendelian randomization (the method that yielded the most significant $P$-value is shown, where red indicates a positive association and blue indicates an inverse association). Asterisks indicate significance: *$P < 0.05$; **FDR $q < 0.05$. A thick black border delineates associations that are also significant using the MR-Egger approach, indicating putative causal relationship that are more robust to pleiotropy.
Association of Baseline Metabolomic Profiles With Incident Stroke and Dementia and With Imaging Markers of Cerebral Small Vessel Disease
Eric L. Harshfield and Hugh S Markus FMed Sci
Neurology published online June 8, 2023
DOI 10.1212/WNL.0000000000207458

This information is current as of June 8, 2023

Updated Information & Services
including high resolution figures, can be found at:
http://n.neurology.org/content/early/2023/06/08/WNL.0000000000207458.full

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
All Cerebrovascular disease/Stroke
http://n.neurology.org/cgi/collection/all_cerebrovascular_disease_stroke
e
All Cognitive Disorders/Dementia
http://n.neurology.org/cgi/collection/all_cognitive_disorders_dementia
All Imaging
http://n.neurology.org/cgi/collection/all_imaging
Cohort studies
http://n.neurology.org/cgi/collection/cohort_studies

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
http://www.neurology.org/about/about_the_journal#permissions

Reprints
Information about ordering reprints can be found online:
http://n.neurology.org/subscribers/advertise