

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Checklist

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	2	'In a clinical cohort of MRI-confirmed lacunar stroke cases and unrelated hospital controls recruited from stroke centers across the UK, we assayed 250 serum metabolites using NMR spectroscopy.'
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2	'We investigated the association of metabolites with MRI-confirmed lacunar stroke, neuroimaging markers, and cognitive impairment. We also conducted a genome-wide association analysis for each metabolite and applied bidirectional Mendelian randomization (MR) to investigate causality.' 'Observational analyses identified 211 metabolites that were significantly associated with lacunar stroke. MR analyses identified significance evidence to support a causal association of the proportion of cholesterol to total lipids within medium LDL (%C/Total in M-LDL) with multiple lacunar infarcts and/or leukoaraiosis (MLI/LA, >1 lacunar infarct and/or Fazekas grade ≥ 2).'
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4	'Despite its clinical significance, the pathogenesis of SVD remains poorly understood. Consequently, current treatment and prevention strategies for SVD are largely restricted to managing established vascular risk factors. Improving understanding of SVD pathophysiology and identifying novel risk factors are critical steps toward developing targeted interventions to prevent stroke and reduce cognitive decline.' 'Metabolomics profiling provides a powerful tool for identifying novel risk factors in SVD by enabling comprehensive characterization of

				metabolic pathways implicated in disease pathogenesis and potentially enhancing prediction of disease progression and severity.'
Objectives	3	State specific objectives, including any prespecified hypotheses	5	'In this study, we compared metabolite concentrations between well-phenotyped, MRI-confirmed lacunar stroke patients and stroke-free controls using data from two longitudinal, prospectively recruited, UK-based, multicenter studies. We examined associations of metabolites with lacunar stroke subtypes, neuroimaging markers of SVD, and cognitive function. By integrating evidence from observational and causal inference analyses, we aimed to identify novel and robust metabolic biomarkers for lacunar stroke.'
Methods				
Study design	4	Present key elements of study design early in the paper	7, 9	'We followed a structured, criterion-based selection process to select serum samples from 2,500 participants (673 DL1 cases, 827 DL2 cases, and 1,000 DL1 controls) for metabolomics profiling.' 'To complement the observational analyses, we conducted a genome-wide association analysis for each of the 250 metabolites using genetic data from DL1 and DL2, and obtained genetic association summary statistics for MRI-confirmed lacunar stroke, ILI, MLI/LA, and neuroimaging markers from published studies... We performed two-sample Mendelian randomization analyses (MR) using the independent SNPs for each metabolite to examine whether there was evidence to support a causal relationship of metabolites with lacunar stroke and MRI markers of SVD.'
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6	'DNA Lacunar 1 (DL1) and 2 (DL2) are prospective studies of MRI-confirmed lacunar stroke patients recruited from UK stroke centers. DL1 recruited 1,030 lacunar stroke patients of European ancestry aged ≤ 70 from 72 specialist stroke centers between 2002 and 2012. DL1 also recruited 1,973 unrelated hospital controls of European ancestry without a history of stroke, myocardial infarction, or dementia, who were randomly recruited from general practices covering the same geographical area as

the cases. DL2 had recruited 1,319 MRI-confirmed lacunar stroke patients from 54 hospitals as of July 2023 (recruitment is ongoing).'

Participants	6	(a) <i>Cohort study</i>—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i>—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i>—Give the eligibility criteria, and the sources and methods of selection of participants	6, eMethods	‘Lacunar stroke was defined by the presence of a clinical lacunar syndrome with an anatomically compatible lesion on MRI (DWI-positive lesion for acute infarcts, or a round/oval subcortical fluid-filled cavity 3-15mm in diameter for cases imaged >3 weeks after the acute event).' ‘Exclusion criteria included stenosis >50% in the extracranial or intracranial cerebral vessels, history of carotid endarterectomy, high or moderate probability of cardioembolic stroke sources defined by the Trial of Org 10172 Acute Stroke Therapy (TOAST) classification criteria, cortical infarcts on MRI, subcortical infarcts >15mm in diameter, monogenic cause of stroke, or other specific causes of stroke.' ‘DL1 also recruited 1,973 unrelated hospital controls of European ancestry without a history of stroke, myocardial infarction, or dementia, who were randomly recruited from general practices covering the same geographical area as the cases.'
		(b) <i>Cohort study</i>—For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i>—For matched studies, give matching criteria and the number of controls per case	N/A	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6, 7, 8	‘Lacunar stroke was defined by the presence of a clinical lacunar syndrome with an anatomically compatible lesion on MRI (DWI-positive lesion for acute infarcts, or a round/oval subcortical fluid-filled cavity 3-15mm in diameter for cases imaged >3 weeks after the acute event). Microbleeds were defined as focal areas of very low intensity <10mm in diameter and counted using the Brain Observer Microbleed Rating Scale. WMHs were graded using the Fazekas scale... Isolated lacunar infarcts

(ILI) and multiple lacunar infarcts (MLI) were defined as 1 or >1 lacunar infarct, respectively, without confluent WMH (Fazekas grade <2). Leukoaraiosis (LA) was defined as having confluent WMH (Fazekas grade ≥ 2) irrespective of the number of lacunar infarcts. MLI and LA were combined...'

'DL2 cases were assessed using the Brief Memory and Executive Test (BMET), a short cognitive screening tool designed and validated to be sensitive for detection of VCI in SVD... A BMET total score ≤ 13 has been validated to indicate presence of VCI.'

'To ensure approximately normal distributions, a generalized log transformation was applied to each metabolite, and concentrations were rescaled using mean centering and dividing by the SD of each metabolite across participants.'

Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7	'Serum samples were shipped on dry ice. High-throughput NMR spectroscopy was used to obtain measurements of 250 metabolites for each sample with high precision and accuracy...'
Bias	9	Describe any efforts to address potential sources of bias	6, 7	'DL1 also recruited 1,973 unrelated hospital controls of European ancestry without a history of stroke, myocardial infarction, or dementia, who were randomly recruited from general practices covering the same geographical area as the cases.' 'MRI scans, performed clinically at each recruiting site, were reviewed centrally to confirm eligibility.' 'All selected cases had lacunar stroke confirmed on central review of imaging.'

Study size	10	Explain how the study size was arrived at	7	‘We followed a structured, criterion-based selection process to select serum samples from 2,500 participants (673 DL1 cases, 827 DL2 cases, and 1,000 DL1 controls) for metabolomics profiling.’
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8	‘To ensure approximately normal distributions, a generalized log transformation was applied to each metabolite, and concentrations were rescaled using mean centering and dividing by the SD of each metabolite across participants.’ ‘Presence of confluent WMH was defined as Fazekas grade ≥ 2, and BCI was defined as BMET total score ≤ 13. Age-normalized scores for the BMET O/M and EF/PS subdomains were analyzed as continuous outcomes—the directions were reversed so that higher scores indicated evidence of cognitive impairment.’
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	8, 9	‘In our primary analyses, we studied the association of metabolites with lacunar stroke cases versus controls, and with ILI and MLI/LA subtypes versus controls, using logistic regression models adjusted for age and sex.’ ‘Primary analyses were conducted using the inverse-variance weighted (IVW) method with an LD clumping threshold of $r^2 < 0.01$.’
		(b) Describe any methods used to examine subgroups and interactions	8, 9	‘Within lacunar stroke cases (i.e. excluding controls), we also analyzed the association of each metabolite with presence of WMH, CMBs, and VCI using logistic regression, and with the BMET O/M and EF/PS subdomain scores using linear regression, all adjusted for age and sex.’ ‘We also examined whether there was significant evidence of interaction of the associations with vascular risk factors.’
		(c) Explain how missing data were addressed	8	‘We excluded samples with insufficient sample volume (minimum 350 μ L) and measurements below the limit of quantification.’

		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed	N/A	N/A
		<i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed		
		<i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy		
		(e) Describe any sensitivity analyses	8, 9	‘As sensitivity analyses, we conducted regression analyses for the association of each metabolite with the above outcomes with further adjustment for hypertension, smoking, alcohol consumption, body mass index (BMI), hyperlipidemia, and type 2 diabetes. We also conducted analyses with further adjustment for medications commonly prescribed for stroke treatment—including antihypertensives, diabetic drugs, lipid-lowering drugs, antiplatelets, and anticoagulants—that patients were taking prior to the stroke event and/or recruitment. Furthermore, we conducted analyses within cases adjusted for the number of days between the stroke event and the date of blood sample collection.’ ‘Sensitivity analyses were performed including the weighted median, simple and weighted mode, and MR-Egger regression methods, which require different assumptions to be satisfied, thus supporting the credibility of inferred causal claims.’
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible,	11, 12, 13	‘Metabolomics measurements that passed quality control were returned for 2,408 participants (DL1: 645 cases, 952 controls; DL2: 811 cases) (Figure 1; Tables 1-3).’

		included in the study, completing follow-up, and analysed		‘Among the 1,456 lacunar stroke cases, we performed cross-sectional univariate analyses of metabolites with two MRI markers of SVD—WMH and cerebral microbleeds—adjusting for age and sex.’ ‘Among 743 DL2 cases with cognitive testing, we assessed associations of 250 metabolites with VCI and scores on the O/M and EF/PS subdomains in cross-sectional analyses adjusted for age and sex.’ ‘DNA samples were genotyped in 1,958 out of 2,408 participants. After excluding samples that did not pass quality control, genetic data were available for 1,925 participants (DL1: 631 cases, 525 controls; DL2: 769 cases) (eTable 15).’
		(b) Give reasons for non-participation at each stage	7, eMethods	‘All selected cases had lacunar stroke confirmed on central review of imaging.’ ‘Exclusion criteria included stenosis >50% in the extracranial or intracranial cerebral vessels, history of carotid endarterectomy, high or moderate probability of cardioembolic stroke sources defined by the Trial of Org 10172 Acute Stroke Therapy (TOAST) classification criteria, cortical infarcts on MRI, subcortical infarcts >15mm in diameter, monogenic cause of stroke, or other specific causes of stroke.’
		(c) Consider use of a flow diagram	11, Figure 1	‘Metabolomics measurements that passed quality control were returned for 2,408 participants (DL1: 645 cases, 952 controls; DL2: 811 cases) (Figure 1; Tables 1-3).’
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	11	‘Lacunar stroke cases were slightly older than controls (Table 1). Cases were more likely to have hypertension, hyperlipidemia, type 2 diabetes, and slightly higher average BMI than controls, and more likely to smoke but less likely to consume alcohol.’
		(b) Indicate number of participants with missing data for each variable of interest	11, Tables 1-3	‘Metabolomics measurements that passed quality control were returned for 2,408 participants (DL1: 645 cases, 952 controls; DL2: 811 cases) (Figure 1; Tables 1-3).’

		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	N/A	N/A
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	N/A	N/A
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	11	‘Metabolomics measurements that passed quality control were returned for 2,408 participants (DL1: 645 cases, 952 controls; DL2: 811 cases) (Figure 1; Tables 1-3).’
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	N/A	N/A
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	11, 12, 13	‘In observational analyses, 211 metabolites out of 250 (84%) were significantly associated with lacunar stroke at an FDR-corrected threshold of $P < 0.05$ (Figures 2-3; eFigures 1-2; eTable 3).’ ‘Although 75 metabolites were nominally associated with WMH and 73 with cerebral microbleeds (uncorrected $P < 0.05$), none remained significant after correction for multiple testing, indicating no robust associations with these MRI markers.’ ‘Genetically determined levels of glycine were associated with a reduced risk of ILI (OR: 0.704, 95% CI 0.585-0.848). Genetic predisposition to a higher proportion of cholesterol to total lipids within medium LDL (%C/Total in M-LDL) was associated with a reduced risk of MLI/LA (OR: 0.541, 95% CI 0.401-0.730).’
		(b) Report category boundaries when continuous variables were categorized	6, 8	‘Isolated lacunar infarcts (ILI) and multiple lacunar infarcts (MLI) were defined as 1 or >1 lacunar infarct, respectively, without confluent WMH (Fazekas grade <2). Leukoaraiosis (LA) was defined as having confluent WMH (Fazekas grade ≥ 2) irrespective of the number of lacunar infarcts. MLI and LA were combined, consistent with previous studies.’

				‘Presence of confluent WMH was defined as Fazekas grade ≥ 2, and BCI was defined as BMET total score ≤ 13. Age-normalized scores for the BMET O/M and EF/PS subdomains were analyzed as continuous outcomes—the directions were reversed so that higher scores indicated evidence of cognitive impairment.’
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12, 13, 14	‘In sensitivity analyses adjusted for additional vascular risk factors and potential confounders, 211 metabolites remained significantly associated with lacunar stroke after FDR correction, with similar results to the basic adjustment model. In a sensitivity analysis further adjusted for pre-recruitment usage of medications commonly taken for stroke (e.g. antihypertensives, antiplatelets, anticoagulants), 203 metabolites were significantly associated with lacunar stroke after FDR correction (eTable 5), overlapping substantially with the basic adjustment results.’ ‘Likewise, there were no significant associations with MRI markers in sensitivity analyses adjusted for additional vascular risk factors and potential confounders (eTable 8), stroke medications (eTable 9), and duration between the stroke event and date of blood draw (eTable 10).’ ‘These five metabolites remained significantly associated with EF/PS in sensitivity analyses, including after adjustment for additional vascular risk factors and potential confounders (eTable 12), stroke medications (eTable 13), and the duration between the stroke event and date of blood sample collection (eTable 14).’ ‘The MR results were consistent in direction across sensitivity analyses (weighted median, simple and weighted mode, MR-Egger) with no evidence of directional or horizontal pleiotropy (eFigure 9). Reverse MR analyses, in which lacunar stroke and neuroimaging markers were used as

exposures and metabolite levels as outcomes, indicated that these associations may be bidirectional (eFigure 10; eTable 20).’

Discussion

Key results	18	Summarise key results with reference to study objectives	15	‘We conducted observational and causal inference analyses to identify metabolites associated with lacunar stroke subtypes, neuroimaging markers of SVD, and cognitive impairment. Observational analyses identified 211 metabolites associated with lacunar stroke. MR analyses identified genetic evidence consistent with causal associations of higher glycine levels with a reduced risk of ILI and higher %C/Total in M-LDL with a reduced risk of MLI/LA.’
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	17	‘This study also has limitations. First, because blood samples in cases occurred at variable intervals after stroke onset, some observed metabolite differences may reflect acute/subacute-phase physiology rather than chronic SVD-related biology. Second, neuroimaging markers and cognitive scores were only available in a subset of DL2 cases and not assessed in controls. Restricting our analyses to post-stroke patients may have introduced collider (index event) bias; additionally, due to the smaller sample size, these analyses had less power to detect associations. Third, since neuroimaging and cognitive testing were not performed in controls, the absence of SVD cannot be definitively ruled out. Fourth, WMH were not segmented and quantified but instead assessed using the semi-quantitative Fazekas score, which is less sensitive to subtle differences compared to volumetric methods, potentially reducing measurement precision. Fifth, details of cerebral microbleed location (lobar vs deep) were unavailable. Sixth, cognition was assessed using the BMET, which has been previously validated to detect VCI in SVD with high sensitivity, but the test was primarily designed as a VCI screening tool and may not provide a comprehensive neuropsychological assessment. Seventh, we applied an FDR correction for multiple testing to reduce the likelihood of identifying false-positive associations, which may have inadvertently excluded some biologically and clinically meaningful associations that did not reach the threshold for statistical

significance. Eighth, although we adjusted for stroke medications (e.g. antiplatelets, anticoagulants), some residual confounding may remain. Ninth, the observational analyses were cross-sectional since follow-up data were not available for clinical endpoints, so we could not evaluate associations with risk of incident cerebrovascular and neurological conditions. We also lacked repeated metabolite measures from multiple timepoints to enable characterization of metabolite changes over the preclinical, acute, and recovery phases of SVD. Tenth, we lacked sufficient power to perform a metabolomics pathway analysis. Eleventh, significant genetic associations were only available for a quarter of the metabolites, possibly due to the small GWAS sample size, which limited the number of metabolites that could be included in MR analyses. Twelfth, there was partial sample overlap between our metabolite GWAS and the outcome GWAS for lacunar stroke, which could potentially bias two-sample MR estimates toward the null due to weak instrument bias and winner's curse. Finally, our analyses were conducted in specific UK-based clinical cohorts of MRI-confirmed lacunar stroke cases and controls; further studies are required to validate these findings in other patient populations and ancestries.'

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	18	'In this study, we identified significant associations of two metabolites, glycine and %C/Total in M-LDL, with two lacunar stroke subtypes, with concordant associations in both observational and MR analyses. Although further research is needed to validate the findings, these findings may inform future research on SVD pathogenesis and biomarker development.'
Generalisability	21	Discuss the generalisability (external validity) of the study results	18	'Finally, our analyses were conducted in specific UK-based clinical cohorts of MRI-confirmed lacunar stroke cases and controls; further studies are required to validate these findings in other patient populations and ancestries.'
Other information				

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19	‘This research was supported by the British Heart Foundation (RG/F/22/110052). Infrastructural support was provided by the Cambridge British Heart Foundation Centre of Research Excellence (RE/18/1/34212; RE/24/130011) and the Cambridge University Hospitals National Institute for Health and Care Research (NIHR) Biomedical Research Centre (NIHR203312). ELH was supported by the Alzheimer’s Society (AS-RF-21-017).’ ‘Collection of the UK Young Lacunar Stroke DNA Study (DNA Lacunar 1) was primarily supported by the Wellcome Trust (WT072952) with additional support from the Stroke Association (TSA 2010/01). Genotyping of the DNA Lacunar 1 samples was supported by a Stroke Association grant (TSA 2013/01).’ ‘Collection and genotyping of the UK Young Lacunar Stroke DNA Study 2 (DNA Lacunar 2) was supported by British Heart Foundation Programme Grants (RG/16/4/32218; RG/F/22/110052).’
---------	----	---	----	---

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.