

Plasma lipid and lipoprotein biomarkers in LBC1936: Do they predict general cognitive ability and brain structure?

Sarah E. Harris PhD^{1*}, Stuart J Ritchie PhD^{1,2}, Gonçalo D S Correia PhD³, Beatriz Jiménez PhD³,
Chloe Fawns-Ritchie PhD¹, Alison Pattie BSc¹, Janie Corley PhD¹, Susana Muñoz Maniega PhD^{1,4,5},
Maria Valdés Hernández PhD^{1,4,5}, John M. Starr PhD⁶, Derek Hill PhD⁷, Paul Wren PhD⁸, Mark E.
Bastin PhD^{1,4,9}, Matthew R Lewis PhD³, Joanna M. Wardlaw MD^{1,4,5,9}, Ian J. Deary PhD¹

¹Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.

²Social, Genetic, and Developmental Psychiatry Centre, King's College, London SE5 8AF, UK.

³National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London SW7 2AZ, UK.

⁴Brain Research Imaging Centre, Neuroimaging Sciences, The University of Edinburgh, Chancellor's Building, 49 Little France Crescent, Edinburgh EH16 4SB, UK.

⁵UK Dementia Research Institute at the University of Edinburgh, Edinburgh BioQuarter, Edinburgh EH16 4SB, UK.

⁶Alzheimer Scotland Dementia Research Centre, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.

⁷Department of Medical Physics and Biomedical Engineering, University College London WC1E 6BT, UK.

⁸ESCAPE Bio, San Francisco, CA 94080, USA

⁹Scottish Imaging Network, A Platform for Scientific Excellence (SINAPSE) Collaboration, 300 Bath St, Glasgow G2 4LH, UK.

*Corresponding author:

Sarah.Harris@ed.ac.uk

Running title: Plasma lipid and lipoprotein biomarkers in LBC1936

Abstract

Identifying predictors of cognitive ability and brain structure in later life is an important step towards understanding the mechanisms leading to cognitive decline and dementia. This study used ultra-performance liquid chromatography mass spectrometry (UPLC-MS) and nuclear magnetic resonance (NMR) to measure targeted and untargeted metabolites, mainly lipids and lipoproteins, in ~600 members of the Lothian Birth Cohort 1936 (LBC1936) at aged ~73 years. Penalized regression models (LASSO) were then used to identify sets of metabolites that predict variation in general cognitive ability and structural brain variables. UPLC-MS-POS measured lipids, together predicted 19% of the variance in total brain volume and 17% of the variance in both grey matter and normal appearing white matter volumes. Multiple subclasses of lipids were included in the predictor, but the best performing lipid was the sphingomyelin SM(d18:2/14:0) which occurred in 100% of iterations of all three significant models. No metabolite set predicted cognitive ability, or white matter hyperintensities or connectivity. Future studies should concentrate on identifying specific lipids as potential cognitive and brain-structural biomarkers in older individuals.

Introduction

Ageing populations have an increasingly greater proportion of individuals affected by cognitive decline and dementia. To further our understanding of these important medical and societal problems, it is important to identify metabolite predictors of individual differences in cognitive ability and brain structure in later life¹⁻⁴. The present study investigates one set of potential predictors of cognitive and brain-structural differences: plasma lipid and lipoprotein biomarkers.

Lipids are important metabolites that form essential components of cell membranes and are involved in cell signalling processes⁵. Lipoproteins are responsible for transporting water-insoluble lipids in blood and also carry apolipoproteins which determine the structure and function of the lipoproteins. Plasma lipid levels can be influenced by genetic⁶ and environmental factors, including diet⁷, exercise⁸ and medication⁹. Dysregulation of lipids has been associated with poorer cognitive function, cognitive decline, Alzheimer's disease (AD), other dementias, and white matter hyperintensities (WMH), although — as one might expect for a relatively new area of research — results are not consistent¹⁰.

High total cholesterol, low-density lipoproteins, triglycerides and apolipoprotein B, and low high-density lipoproteins have been associated with poorer cognitive performance and increased risk for dementia, particularly vascular dementia, in some, but not all studies (reviewed in¹⁰). Inconsistencies may be due to the age of the subjects, the cognitive domains tested, dementia diagnosis criteria, and the lipids measured. Sex specific and non-linear associations of serum lipid levels with cognitive function have also been identified in middle- to older- aged individuals¹¹ and a study in the Lothian Birth Cohort 1936 (LBC1936) indicated that associations with cognitive function in older age may be confounded by childhood cognitive function¹².

Cholesterol is an important lipid in the brain, vital for synapse formation and development, dendrite differentiation, axonal elongation, and long-term potentiation¹³. The blood-brain barrier (BBB) separates brain metabolised cholesterol and lipoproteins from those in plasma. Increased BBB permeability is associated with cognitive decline and has been shown to occur in rats fed diets high in saturated fats and cholesterol¹⁴⁻¹⁶. Anti-inflammatory and lipid-lowering agents have been shown to reverse this phenomenon¹⁴. Increased HDL-cholesterol and decreased LDL-cholesterol were associated with increased WMH in a longitudinal study of 1919 individuals aged over 65 years¹⁷ and hyperlipidemia was associated with less severe WMH in patients with acute ischaemic stroke in two independent cohorts (N=1135)¹⁸, suggesting that hyperlipidemia may actually protect against small vessel disease. Lower HDL-cholesterol predicted increased WMH volume from age 73 to 76 years in the LBC1936¹⁹.

Extracellular plaques consisting of amyloid beta peptides are one of the main neuropathologies associated with AD. Genetic variants in genes involved in the amyloid pathway are associated with AD. Apolipoprotein E (APOE) regulates transport of cholesterol and lipids and mediates clearance of plasma lipoproteins. The *APOE* e4 allele, that is associated with an increased risk of late-onset AD as well as risk of steeper non-AD cognitive decline²⁰, impairs clearance of the plasma lipoproteins, enhancing amyloid beta aggregation²¹. Sphingolipids, including sphingomyelin and ceramide, have important functional and structural roles in cellular membranes. Sphingomyelin and cholesterol interact to influence membrane permeability²².

Modern analytical technologies such as nuclear magnetic resonance (NMR) and ultra-performance liquid chromatography mass spectrometry (UPLC-MS) enable the broadscale detection and measurement of chemical constituents within biofluids and tissue extracts. Although both techniques have been used extensively to measure small molecule metabolites,

NMR has emerged in recent years as a popular and reproducible means to quantify lipoprotein species²³. Similarly, UPLC-MS is now commonly applied for the measurement of complex lipids from human biofluids and tissue extracts. Such analyses may be targeted to measure sets of well-defined species or untargeted whereby thousands of lipid-related features are detected in a single analysis. In the latter case, those measurements showing a positive association with a trait of interest require chemical annotation prior to downstream interpretation of the results.

Large scale metabolomics studies of cognitive function and decline are starting to identify biological mechanisms and biomarkers of cognitive impairment in later life^{24,25}. A recent study of lipidome evolution in mammalian tissues showed a significant excess of lipid concentration changes in the brain cortex, with cortical changes clustered in glycerolipid, glycerophospholipid, and linoleic acid metabolism pathways, all of which have previously been implicated in neurodegenerative disorders²⁶. Serum levels of a glycerophosphocoline (a subclass of glycerophospholipids) have been suggested as potential marker of visceral fat related peripheral inflammation, associated with changes in brain structure and function²⁷.

In the present study we used penalized regression models (LASSO) to identify sets of metabolites (principally lipids and lipoproteins) in both targeted and untargeted UPLC-MS and NMR spectroscopy data sets that might be used to predict some of the variation in general cognitive ability and structural brain variables in the LBC1936.

Materials and Methods

Lothian Birth Cohort 1936 (LBC1936)

The LBC1936 consists of 1091 individuals, most of whom, at the age of ~11 years, took part in the Scottish Mental Survey of 1947, when they took a validated test of cognitive ability,

the Moray House Test (MHT) version 12. Between 2004 and 2007, they were recruited to a study to determine influences on cognitive ageing at age ~70 years and have now taken part in five waves of testing in later life (ages ~70, ~73, ~76, ~79 and ~82 years). At each wave, they have undergone a series of cognitive and physical tests, with brain magnetic resonance imaging (MRI) introduced at age 73 years. For the present study plasma was extracted from lithium heparin collected blood at Wave 2 at a mean age of 72.5 (SD 0.7) years and stored at -80°C²⁸.

Metabolomics

Plasma metabolic profiles were measured at the National Phenome Centre at Imperial College London. Plasma samples were shipped on dry ice and stored at -80°C upon receipt prior to aliquoting for subsequent analyses by thawing in batches of 80 overnight at 4°C, centrifuging, and decanting the homogenous supernatant into 96 well plates as previously described²⁹. Sample plates were stored at -80°C until needed for analysis. Quality control materials including a study reference (SR) sample (equal-parts pool of all study samples) were prepared as previously described²⁹.

Lipidomics

Plasma samples were prepared for lipidomics analysis by dilution with isopropanol³⁰ as previously described for serum analysis²⁹ with the following modification: 100 µL of sample were thawed at 4°C for 2h and protein precipitation was achieved by addition of four parts of an isopropanol solution containing reference standards to one part plasma. UPLC-MS was performed using Acquity UPLC chromatography systems (Waters Corporation, Milford, MA, USA) and Xevo G2-S Q-ToF mass spectrometers with electrospray interface (Waters Corp., Manchester, UK). Reversed-phase chromatography was conducted using an Acquity 2.1×100mm BEH C8 column, and separate analyses were performed for positive and

negative ion detection (UPLC-MS-POS and UPLC-MS-NEG respectively). Details of the chromatographic solvent system, gradient elution, mass spectrometer parameters, and data acquisition have been published previously²⁹. Targeted integration of signals arising from known lipid species in the UPLC-MS datasets was performed using the peakPanther R package³¹ and an in-house database of empirical retention time and theoretical m/z values for annotated lipids. The nPYc-Toolbox³² was then used to correct ion intensities for sample analysis-order intensity drift using a LOWESS smoother estimated using the intensities measured across repeated injections of the SR. Following drift correction, the SR samples and their serial dilution series were used to calculate feature-wise coefficients of variation (CV) and Pearson correlation coefficients with the dilution factor, respectively. Only signals with a CV lower than 30% and a correlation with dilution higher than 0.7 were kept for statistical analysis.

Biocrates AbsoluteIDQ® p180

The Biocrates AbsoluteIDQ® p180 (Biocrates Life Sciences AG, Innsbruck, Austria) kit provided absolute quantification for 53 compounds, and semi-quantitative measurements for a further 135 compounds. Data were acquired on Waters TQ-S instruments (Waters Corp., Manchester, UK) using Acquity UPLC systems (Waters Corporation, Milford, MA, USA).

Nuclear Magnetic Resonance (NMR) Spectroscopy

An adaptation of the standard metabolic profiling method for NMR³³ was developed and validated for this analysis due to the limited sample volume available. Briefly, 40 µL of plasma were mixed with 40 µL of aqueous buffer containing 20% of D₂O 75 mM of NaH₂PO₄, 4.6 mM of TSP (3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid or TMSP-d₄) and 6.2 mM of NaN₃ at pH 7.4 in an Eppendorf and vortexed. Subsequently, 60 µL of the mixture were transferred to a 4'' 1.7 mm tube using a 100 µL Hamilton syringe. Samples

were prepared in batches of 80 samples and subsequently split into two batches of 27 and one of 26 for analysis. One SR and one LTR samples were included in each of the batches with a total of 6 QCs included per 80 samples. An additional 600 μ L of LTR sample was prepared and run during the calibration of each spectrometer in order to transfer the IVDr methods to the small volume samples. Each 27-sample batch was transferred to the refrigerated sample handling robot (SampleJet, Bruker Corporation, Germany) of a one of three Bruker Avance III HD 600 NMR spectrometers for analysis and parallel acquisition was completed within 24 hours of sample preparation. 1D NMR profiling was acquired using the 1D-NOESY presat pulse sequence, a spin-echo experiment using the 1D-CPMG (Carr-Purcell-Meiboom-Gill) presat pulse sequence and *J*-res 2D experiments were run in automation at 310 K using a BBI probe with z-gradients and high degree shimming. A relaxation edited spin-echo experiment, the CPMG sequence, is often acquired for plasma samples in order to filter out fast relaxing signals belonging to proteins and other large molecules in the ^1H NMR spectrum, providing a better resolution for the slow relaxing signals belonging to metabolites and small molecules. 128 Free Induction Decays (FID) were accumulated for each experiment in 96 K points using a 30 ppm window centred at 4.78 ppm for the general profile while it was set to 20 ppm for the spin-echo, and 6 scans and 40 planes were acquired for the 2D *J*-resolved experiment. The relaxation delay was set at 4 s for the 1Ds while the *J*res was set to 2s and a water pre-saturation pulse was applied during this period to cancel the water signal.

Quality of the spectra were assessed following the criteria described in ³³. Any samples not passing the QC criteria were rerun at least once, and otherwise left outside the final dataset. Data was prepared for statistical analysis with the nPYc-Toolbox ³², by re-calibrating the chemical shift scale using the α -glucose signal at δ 5.233 and re-interpolating all spectra to a common chemical shift axis.

B.I.-LISA platform

One hundred and five specific lipoproteins were measured using a modified version of the Bruker *B.I.-Lisa* platform (Bruker IVDr Lipoprotein Subclass Analysis). The platform is developed and optimised for the analysis of 300 μ L aliquot plasma samples as described in³⁴. In order to calculate the correction factor needed to extend the applicability of the method to the current analysis, an independent set of 622 samples was analysed both using the standard protocol and the reduced volume one used for this project. The error associated to the current measurements is increased with respect to the precision and accuracy reported in²³. The lipoproteins measured were cholesterol, free cholesterol, phospholipids, triglycerides, apolipoproteins A1, A2, B and particle numbers for the overall concentrations contained in plasma but also information on the concentration of the subclasses of those lipoproteins as obtained by ultracentrifugation was provided.

Cognitive Tests

Cognitive ability was assessed using 13 cognitive tests taken by the LBC1936 sample, all at age 72.5 (SD 0.7) years (Wave 2). The tests were the Matrix Reasoning (MR), Block Design (BD), Digit Span Backward (DSB), Symbol Search (SS), and Digit-Symbol Substitution (DSS) tests from the Wechsler Adult Intelligence Scale, Third UK Edition³⁵; the Logical Memory (LM), Verbal Paired Associates (VPA), and Spatial Span (SSP) tests from the Wechsler Memory Scale, Third UK Edition³⁶; the National Adult Reading Test (NART)³⁷; the Wechsler Test of Adult Reading (WTAR)³⁸; a phonemic Verbal Fluency (VF) test³⁹; a Choice Reaction Time (CRT) test⁴⁰; and an Inspection Time test (IT)⁴¹. Structural equation modeling was used to produce a latent variable from the variance shared among these tests in the following manner. On the basis of previous factor-analytic work in this sample⁴², the tests were organized into a higher-order confirmatory factor-analytic model that included latent domains of visuospatial ability (indicated by the tests of MR, BD, and SSP), verbal

memory (indicated by LM, VPA, and DSB), processing speed (indicated by SS, DSS, CRT, and IT), and crystallized ability (indicated by NART, WTAR, and VF), all of which were indicators of the highest-level latent variable, which we named “general cognitive ability”. From this model, the factor scores for the general cognitive ability factor were saved, and were used to index each individual participant’s general cognitive ability in our prediction models. Structural equation modelling was performed, and the latent variables were produced, using the lavaan package for R ⁴³.

Brain MRI

Whole-brain structural and diffusion tensor MRI data were acquired using a 1.5T GE Signa Horizon scanner (General Electric, Milwaukee, WI, USA) at the Brain Research Imaging Centre, the University of Edinburgh, shortly after cognitive testing and plasma collection. Mean age at scanning was 72.7 (SD 0.7) years. Full details are given in ⁴⁴. Total brain, grey matter, normal appearing white matter and WMH volumes were calculated using a semi-automated multispectral fusion method, described previously ^{45–47}.

The diffusion tensor MRI protocol employed a single-shot spin-echo echo-planar (EP) diffusion weighted sequence in which diffusion-weighted EP volumes ($b = 1000 \text{ s mm}^{-2}$) were acquired in 64 non-collinear directions, together with seven T₂-weighted EP volumes ($b = 0 \text{ s mm}^{-2}$). This protocol was run with 72 contiguous axial slices with a field of view of 256 × 256 mm, an acquisition matrix of 128 × 128 and 2mm isotropic voxels ⁴⁷.

White matter connectivity data were created using the BEDPOSTX/ProbTrackX algorithm in FSL (<https://fsl.fmrib.ox.ac.uk>) and 12 major tracts of interest were segmented using Tractor (<https://www.tractor-mri.org.uk>) scripts: the genu and splenium of the corpus callosum, and bilateral anterior thalamic radiations, cingulum bundles, uncinate, arcuate and inferior longitudinal fasciculi. Tract-average white matter fractional anisotropy (FA) and mean

diffusivity (MD) were derived as the average of all voxels contained within the resultant tract maps. General factors of FA (*gFA*) and MD (*gMD*) were derived from a confirmatory factor analysis with a single general variable indicated by either FA or MD values from across all 12 tracts, to reflect the well-replicated phenomenon of common microstructural properties of brain white matter⁴⁸⁻⁵⁰. As above, the confirmatory factor analytic models used the *lavaan* package.

Brain imaging data and metabolite measures were available for a maximum of 668 and 595 individuals, respectively.

Statistical analyses

LASSO prediction models

In order to produce metabolomic predictors of general cognitive function and of neuroimaging measures, Least Absolute Shrinkage and Selection Operator (LASSO) penalized regression models were used⁵¹. These models produce a sparse solution to a prediction problem, dealing with a large number of predictors and the potential for collinearity among them. They do so by shrinking (penalizing) the coefficients of many of the predictors to zero in their regularisation process; in this sense, they also allow for variable selection, since a smaller (and thus more manageable) number of predictors might survive the process than at the outset.

We ran a series of LASSO models, attempting to predict variation in each brain MRI variable or cognitive outcome from all of the features derived from each metabolomic array in turn. For each model, the LBC1936 data were split at random into a training sample (80% of the data) and a testing sample (20% of the data). We first performed a LASSO regression in the training sample alone. Using 10-fold cross-validation, the penalization parameter (often denoted as λ) that provided the lowest prediction error in the training set was identified. The

predicted values from a regression with this cross-validated penalization parameter were then taken forward to the testing sample, where they formed a predictor that was regressed on the relevant outcome variable (cognitive ability or brain volume). The R^2 (proportion of variance accounted for by the predictor, expressed in the results below as a percentage) was recorded for each model. This procedure was repeated 1,000 times, with the LBC1936 dataset being split into a different random training and testing set each time. The mean R^2 from the 1,000 iterations was calculated, and the standard deviation across all the iterations was used as the (bootstrapped) standard error for this mean value. Where the computational demands were too high for this iterative procedure to successfully be performed, the number of model iterations was reduced.

The LASSO models were run using the *glmnet* package for R ⁵², with wrapper functions from the *caret* package ⁵³.

In a final feature-selection stage, metabolites that occurred in > 75% of the 1000 iterations were identified for those models where the metabolites explained a significant proportion of the variance in the outcome (that is, the confidence interval of their R^2 value, derived using the bootstrapped standard error, did not cross zero). Since a large number of R^2 values were calculated across all the arrays, we corrected their significance levels using the False Discovery Rate correction. This adjustment for multiple comparisons meant that, at the end of our analysis, only the strongest predictions would remain statistically significant.

Sensitivity analyses

For models where the metabolites explained a significant proportion of the variance in the outcome, two sensitivity analyses were performed: 1) using rank-based inverse normal transformed metabolites (*ranknorm()* function from the R package *RNOmni* ⁵⁴) to ensure that predictions were not being driven by outlying measurements; 2) correcting brain volumes for

intracranial volume (a proxy measure of previous brain volume) to identify specific associations with age-related brain volume atrophy. Correlations between total brain, grey matter and normal appearing white matter volumes, and intracranial volume were also calculated.

Results

Descriptive statistics for general cognitive ability and brain MRI variables are shown in Table 1. Eight hundred and sixty-six subjects provided cognitive data, and between 656 and 668 provided data on the brain imaging variables.

For the Biocrates AbsoluteIDQ[®] p180 platform, 171 of 188 metabolites were detected in LBC1936 (Supplementary Table 1). All 105 B.I.-LISA platform metabolites were detected. (Supplementary Table 2). From the UPLC-MS features, 193 and 55 metabolites were annotated, respectively, in the UPLC-MS-POS and UPLC-MS-NEG datasets (Supplementary Tables 3 and 4). Untargeted NMR-NOESY and NMR-CPMG sequences each detected 18,646 features. Metabolites detected by each platform were used in platform-specific penalized regression models (LASSO) as described above.

Prediction results are displayed in Table 2. Only a small number of predictions, all from the UPLC-MS-POS array, remained statistically significant after correction for multiple comparisons. However, in those cases, the predicted variance from the LASSO models was substantial.

UPLC-MS-POS lipids together predicted 19% (SE 5%, $p=0.00025$) of the variance in total brain volume and 17% (SE 5%, $p=0.001$) of the variance in both grey matter and normal appearing white matter volumes (Table 2). Nineteen of the 193 UPLC-MS-POS lipids occurred in more than 75% of iterations for at least two of the significant models

(Supplementary Table 5). These included lipids from the following lipid subclasses: sphingomyelins, ceramides, fatty acylcarnitines, monoacylglycerols, diacylglycerophosphocholines and monoacylglycerophosphocholines. Sphingomyelin, SM(d18:2/14:0), occurred in 100% of iterations of all three statistically significant models.

No set of metabolites predicted cognitive ability, WMH volume, or either measure of white matter connectivity.

Using rank-based inverse normal transformed metabolites in the models, indicated that significant predictions were not being driven by outlying metabolite measurements, as results remained essentially the same (Supplementary Table 6). Correlations between total brain, grey matter and normal appearing white matter volumes, and intracranial volume ranged from 0.8-0.9 (Supplementary Table 7). Correcting total brain, grey matter and normal appearing white matter volumes for intracranial volume greatly attenuated the results (Supplementary Table 6), indicating that the associations were mainly with absolute brain volumes and not age-related brain volume atrophy.

Discussion

This study tested a variety of metabolite arrays for their ability to predict important cognitive and brain structure phenotypes in later life. UPLC-MS array analysis followed by penalized regression models (LASSO) identified sets of lipids that made statistically significant and substantial predictions of total brain, grey matter and normal appearing white matter volumes in the LBC1936 sample at the approximate age of 73 years.

The best performing single lipid, SM(d18:2/14:0), occurred in 100% of iterations of all three significant models, indicating a robust association which stood out beyond other lipids. SM(d18:2/14:0) is a sphingomyelin previously shown, in combination with three other

compounds, to discriminate between stable and unstable angina⁵⁵ and to be associated with obesity⁵⁶. To our knowledge, it has not previously been associated with brain volume or other brain imaging related phenotypes. A second sphingomyelin, SM(d18:0/16:0), occurred in >85% of iterations of at least two of the significant models. The ceramide Cer(d18:1/16:0), which occurred in 99 or 100% of iterations of all three significant models, was previously associated with major adverse cardiovascular events⁵⁷. A second ceramide, Cer(d19:1/24:0), occurred in >79% of iterations of all three significant models.

Sphingolipids, including sphingomyelin and ceramide, are enriched in the central nervous system and sphingomyelin is a component of myelin sheaths. Levels of sphingomyelin and ceramide increase with age⁵⁸⁻⁶⁰ and both levels and ratios of these sphingolipids have previously been associated with neurodegenerative disease⁶¹⁻⁶³, cognitive function, cognitive decline, and changes in white matter microstructure⁶⁴. Sphingomyelin degradation leads to increased ceramide⁶⁵. Ceramide regulates cell growth, differentiation, and apoptotic signaling in cells types, including neurons⁶⁶. Sphingolipids may, therefore, contribute to the prediction of brain tissue volumes in the LBC1936 via these mechanisms.

A fatty acylcarnitine [CAR(26:0)] occurred in 99 or 100% of iterations of all three significant models. Fatty acylcarnitines are important for energy production and levels are often altered in fatty acid oxidation disorders⁶⁷. Five other fatty acylcarnitines occurred in >75% of iterations of at least two of the significant models. Other lipid classes occurring in >75% of iterations of at least two of the significant models included, monoacylglycerols, diacylglycerophosphocholines and monoacylglycerophosphocholines, indicating that a number of lipid subclasses may be important in predicting brain volume in later-life.

Although sets of lipids predict later-life brain tissue volumes in the LBC1936, this study provides no evidence that they can be used to predict later-life general cognitive ability, WMH volume or white matter connectivity. However, as we refine the lipid sets, prediction

of these may be possible in the future. Future analyses could also investigate specific cognitive domains, rather than general cognitive ability, as these may be more sensitive to the effects of lipids.

Sets of lipoproteins, as measured using NMR, did not predict general cognitive ability or brain structure in this study. It may be that they are not as predictive as the lipids themselves, or that they are not as sensitive to detection by the methods used in this study.

Limitations of this study included the fact that the metabolites were measured in blood plasma rather than cerebrospinal fluid. However, plasma is a much less invasive material to collect and therefore biomarkers that can be measured in blood are particularly useful. The implications of our choice of LASSO models over other forms of penalized regression should also be borne in mind. The LASSO, compared to the related Elastic Net regression, tends to provide a sparser—and thus more readily-interpretable—set of selected features. However, given the very large number of predictors included in some of our models, the LASSO was more computationally feasible, since it does not require calculation of the additional hyperparameters required for an Elastic Net analysis. Other strengths of this study include the fact that the metabolites, cognitive ability, and structural brain MRI variables were measured at about the same time in ~600 individuals with a narrow age range and from an ancestrally homogeneous population. Our penalized regression analysis allowed us to run prediction models that were robust to overfitting (since we split our overall sample into training and test subsets, and used cross-validation in the training subset), and multicollinearity. The form of modelling we employed to deal with the cognitive variables, latent variable structural equation modelling, reduced the impact of measurement error specific to any individual task or domain of cognitive ability, and allowed us to examine more directly each participant's cognitive ability at a highly general level.

In conclusion, sets of lipids, particularly sphingolipids measured using UPLC-MS, predicted up to 19% of the variance in brain tissue volumes in an older relatively healthy Scottish cohort. These results suggest that future studies should concentrate on identifying specific lipids as potential cognitive and brain-structural biomarkers in older individuals.

Acknowledgments

We thank the LBC1936 participants and team members who contributed to these studies. Phenotype collection was supported by Age UK (The Disconnected Mind project). MRI brain imaging was supported by Medical Research Council (MRC) grants G0701120, G1001245, MR/M013111/1 and MR/R024065/1. Metabolomics were supported by the Dementia Platform UK MRC grant MR/L023784/2. This work was supported by the MRC and National Institute for Health Research [grant number MC_PC_12025] and infrastructure support was provided by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC). The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative; funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and MRC is gratefully acknowledged (MR/K026992/1). We gratefully acknowledge the contribution of co-author Professor John M. Starr, who died prior to the publication of this manuscript.

Conflict of Interest

The authors have nothing to disclose.

Supplementary Information

Supplementary information is available here.

References

- 1 Wardlaw JM, Valdés Hernández MC, Muñoz-Maniega S. What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. *J Am Heart Assoc* 2015; **4**: 001140.
- 2 Ritchie SJ, Bastin ME, Tucker-Drob EM, Maniega SM, Engelhardt LE, Cox SR *et al.* Coupled Changes in Brain White Matter Microstructure and Fluid Intelligence in Later Life. *J Neurosci* 2015; **35**: 8672–8682.
- 3 Vibha D, Tiemeier H, Mirza SS, Adams HHH, Niessen WJ, Hofman A *et al.* Brain Volumes and Longitudinal Cognitive Change. *Alzheimer Dis Assoc Disord* 2018; **32**: 43–49.
- 4 Ritchie SJ, Dickie DA, Cox SR, Valdes Hernandez M del C, Corley J, Royle NA *et al.* Brain volumetric changes and cognitive ageing during the eighth decade of life. *Hum Brain Mapp* 2015; **36**: 4910–4925.
- 5 Hannun YA, Obeid LM. Principles of bioactive lipid signalling: lessons from sphingolipids. *Nat Rev Mol Cell Biol* 2008; **9**: 139–150.
- 6 Peloso GM, Natarajan P. Insights from population-based analyses of plasma lipids across the allele frequency spectrum. *Curr Opin Genet Dev* 2018; **50**: 1–6.
- 7 Fernandez ML, West KL. Mechanisms by which Dietary Fatty Acids Modulate Plasma Lipids. *J Nutr* 2005; **135**: 2075–2078.
- 8 Wang Y, Xu D. Effects of aerobic exercise on lipids and lipoproteins. *Lipids Health Dis* 2017; **16**: 132.
- 9 Russell C, Sheth S, Jacoby D. A Clinical Guide to Combination Lipid-Lowering

- Therapy. *Curr Atheroscler Rep* 2018; **20**: 19.
- 10 Li R, Wang T-J, Lyu P-Y, Liu Y, Chen W-H, Fan M-Y *et al*. Effects of Plasma Lipids and Statins on Cognitive Function. *Chin Med J (Engl)* 2018; **131**: 471.
 - 11 Lu Y, An Y, Yu H, Che F, Zhang X, Rong H *et al*. Sex-specific nonlinear associations between serum lipids and different domains of cognitive function in middle to older age individuals. *Metab Brain Dis* 2017; **32**: 1089–1097.
 - 12 Corley J, Starr JM, Deary IJ. Serum cholesterol and cognitive functions: The Lothian Birth Cohort 1936. *Int Psychogeriatrics* 2015; **27**: 439–453.
 - 13 Iuliano L, Crick PJ, Zerbinati C, Tritapepe L, Abdel-Khalik J, Poirot M *et al*. Cholesterol metabolites exported from human brain. *Steroids* 2015; **99**: 189–193.
 - 14 Pallegage-Gamarallage M, Lam V, Takechi R, Galloway S, Clark K, Mamo J. Restoration of dietary-fat induced blood–brain barrier dysfunction by anti-inflammatory lipid-modulating agents. *Lipids Health Dis* 2012; **11**: 117.
 - 15 Freeman LR, Granholm A-CE. Vascular changes in rat hippocampus following a high saturated fat and cholesterol diet. *J Cereb Blood Flow Metab* 2012; **32**: 643–53.
 - 16 Davidson TL, Monnot A, Neal AU, Martin AA, Horton JJ, Zheng W. The effects of a high-energy diet on hippocampal-dependent discrimination performance and blood-brain barrier integrity differ for diet-induced obese and diet-resistant rats. *Physiol Behav* 2012; **107**: 26–33.
 - 17 Longstreth WT, Arnold AM, Beauchamp NJ, Manolio TA, Lefkowitz D, Jungreis C *et al*. Incidence, Manifestations, and Predictors of Worsening White Matter on Serial Cranial Magnetic Resonance Imaging in the Elderly: The Cardiovascular Health

- Study. *Stroke* 2005; **36**: 56–61.
- 18 Jimenez-Conde J, Biffi A, Rahman R, Kanakis A, Butler C, Sonni S *et al.* Hyperlipidemia and reduced white matter hyperintensity volume in patients with ischemic stroke. *Stroke* 2010; **41**: 437–42.
- 19 Dickie DA, Ritchie SJ, Cox SR, Sakka E, Royle NA, Aribisala BS *et al.* Vascular risk factors and progression of white matter hyperintensities in the Lothian Birth Cohort 1936. *Neurobiol Aging* 2016; **42**: 116–123.
- 20 O’Donoghue MC, Murphy SE, Zamboni G, Nobre AC, Mackay CE. APOE genotype and cognition in healthy individuals at risk of Alzheimer’s disease: A review. *Cortex*. 2018; **104**: 103–123.
- 21 Karch CM, Cruchaga C, Goate AM. Alzheimer’s disease genetics: from the bench to the clinic. *Neuron* 2014; **83**: 11–26.
- 22 Needham D, Nunn RS. Elastic deformation and failure of lipid bilayer membranes containing cholesterol. *Biophys J* 1990; **58**: 997–1009.
- 23 Jiménez B, Holmes E, Heude C, Tolson RF, Harvey N, Lodge SL *et al.* Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by ¹H NMR Spectroscopy in a Multilaboratory Trial. *Anal Chem* 2018; **90**: 11962–11971.
- 24 Bressler J, Yu B, Mosley TH, Knopman DS, Gottesman RF, Alonso A *et al.* Metabolomics and cognition in African American adults in midlife: the atherosclerosis risk in communities study. *Transl Psychiatry* 2017; **7**: e1173–e1173.
- 25 Shi M, Bazzano LA, He J, Gu X, Li C, Li S *et al.* Novel serum metabolites associate

- with cognition phenotypes among Bogalusa Heart Study participants. *Aging (Albany NY)* 2019; **11**: 5124–5139.
- 26 Khrameeva E, Kurochkin I, Bozek K, Giavalisco P, Khaitovich P. Lipidome evolution in mammalian tissues. *Mol Biol Evol* 2018. doi:10.1093/molbev/msy097.
- 27 Syme C, Pelletier S, Shin J, Abrahamowicz M, Leonard G, Perron M *et al.* Visceral fat-related systemic inflammation and the adolescent brain: a mediating role of circulating glycerophosphocholines. *Int J Obes (Lond)* 2019; **43**: 1223–1230.
- 28 Taylor AM, Pattie A, Deary IJ. Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol* 2018; **47**: 1042-1042r.
- 29 Izzi-Engbeaya C, Comninou AN, Clarke SA, Jomard A, Yang L, Jones S *et al.* The effects of kisspeptin on β -cell function, serum metabolites and appetite in humans. *Diabetes, Obes Metab* 2018; **20**: 2800–2810.
- 30 Sarafian MH, Gaudin M, Lewis MR, Martin F-P, Holmes E, Nicholson JK *et al.* Objective Set of Criteria for Optimization of Sample Preparation Procedures for Ultra-High Throughput Untargeted Blood Plasma Lipid Profiling by Ultra Performance Liquid Chromatography–Mass Spectrometry. *Anal Chem* 2014; **86**: 5766–5774.
- 31 Wolfer AM, jaketmp, gscorreia89, Turaga N. phenomecentre/peakPantheR: Bioc 3.10 release version (v1.0.0). 2019. doi:10.5281/ZENODO.3523406.
- 32 Sands CJ, Wolfer AM, Correia GDS, Sadawi N, Ahmed A, Jiménez B *et al.* The nPYc-Toolbox, a Python module for the pre-processing, quality-control and analysis of metabolic profiling datasets. *Bioinformatics* 2019; **35**: 5359–5360.
- 33 Dona AC, Jiménez B, Schafer H, Humpfer E, Spraul M, Lewis MR *et al.* Precision

- high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem* 2014; **86**: 9887–9894.
- 34 Gerrish A, Russo G, Richards A, Moskvina V, Ivanov D, Harold D *et al.* The role of variation at A β PP, PSEN1, PSEN2, and MAPT in late onset Alzheimer’s disease. *J Alzheimer’s Dis* 2012; **28**. doi:10.3233/JAD-2011-110824.
- 35 Wechsler D. *WAIS-III UK administration and scoring manual*. Psychological Corporation: London, 1998.
- 36 Wechsler D. *WMS-III : Wechsler memory scale administration and scoring manual*. 3rd ed. Psychological Corp.: London, UK, 1998<https://www.worldcat.org/title/wms-iii-wechsler-memory-scale-administration-and-scoring-manual/oclc/38729493>.
- 37 Nelson HE, Willison JR. *National Adult Reading Test (NART) Test Manual (Part II)*. NFER-Nelson: Windsor, England, 1991.
- 38 Holdnack JA. *Wechsler Test of Adult Reading manual*. Psychological Corporation: San Antonio Tex., 2001.
- 39 Lezak M. *Neuropsychological testing*. Oxford University Press: Oxford, England, 2004.
- 40 Deary IJ, Der G, Ford G. Reaction times and intelligence differences: A population-based cohort study. *Intelligence* 2001; **29**: 389–399.
- 41 Deary IJ, Simonotto E, Meyer M, Marshall A, Marshall I, Goddard N *et al.* The functional anatomy of inspection time: an event-related fMRI study. *Neuroimage* 2004; **22**: 1466–1479.
- 42 Tucker-Drob EM, Briley DA, Starr JM, Deary IJ. Structure and correlates of cognitive

- aging in a narrow age cohort. *Psychol Aging* 2014; **29**: 236–249.
- 43 Rosseel Y. **lavaan** : An R Package for Structural Equation Modeling. *J Stat Softw* 2012; **48**: 1–36.
- 44 Wardlaw JM, Bastin ME, Valdés Hernández MC, Maniega SM, Royle NA, Morris Z *et al*. Brain aging, cognition in youth and old age and vascular disease in the Lothian Birth Cohort 1936: rationale, design and methodology of the imaging protocol. *Int J Stroke* 2011; **6**: 547–59.
- 45 Hernández M del CV, Ferguson KJ, Chappell FM, Wardlaw JM. New multispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images. *Eur Radiol* 2010; **20**: 1684–91.
- 46 Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R *et al*. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol* 2013; **12**: 822–38.
- 47 Cox SR, Allerhand M, Ritchie SJ, Muñoz Maniega S, Valdés Hernández M, Harris SE *et al*. Longitudinal serum S100 β and brain aging in the Lothian Birth Cohort 1936. *Neurobiol Aging* 2018; **69**. doi:10.1016/j.neurobiolaging.2018.05.029.
- 48 Telford EJ, Cox SR, Fletcher-Watson S, Anblagan D, Sparrow S, Pataky R *et al*. A latent measure explains substantial variance in white matter microstructure across the newborn human brain. *Brain Struct Funct* 2017; **222**: 4023–4033.
- 49 Penke L, Maniega SM, Murray C, Gow AJ, Valdes Hernandez MC, Clayden JD *et al*. A General Factor of Brain White Matter Integrity Predicts Information Processing Speed in Healthy Older People. *J Neurosci* 2010; **30**: 7569–7574.

- 50 Cox SR, Ritchie SJ, Tucker-Drob EM, Liewald DC, Hagenaars SP, Davies G *et al.* Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nat Commun* 2016; **7**: 13629.
- 51 Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Ser B* 1996; **58**: 267–288.
- 52 Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010; **33**: 1–22.
- 53 Kuhn M. caret: Classification and Regression Training (version 6.0-84). 2018. <https://rdrr.io/cran/caret/> (accessed 26 Sep2019).
- 54 McCaw Z. Rank Normal Transformation Omnibus Test [R package RNOmni version 0.7.1]. <https://cran.r-project.org/package=RNOmni> (accessed 1 Jul2020).
- 55 Calderón-Santiago M, Priego-Capote F, Galache-Osuna JG, Luque de Castro MD. Analysis of serum phospholipid profiles by liquid chromatography-tandem mass spectrometry in high resolution mode for evaluation of atherosclerotic patients. *J Chromatogr A* 2014; **1371**: 154–162.
- 56 Hellmuth C, Kirchberg FF, Brandt S, Moß A, Walter V, Rothenbacher D *et al.* An individual participant data meta-analysis on metabolomics profiles for obesity and insulin resistance in European children. *Sci Rep* 2019; **9**. doi:10.1038/s41598-019-41449-x.
- 57 Mantovani A, Dugo C. Ceramides and risk of major adverse cardiovascular events: A meta-analysis of longitudinal studies. *J. Clin. Lipidol.* 2020. doi:10.1016/j.jacl.2020.01.005.

- 58 Cutler RG, Kelly J, Storie K, Pedersen WA, Tammara A, Hatanpaa K *et al.*
Involvement of oxidative stress-induced abnormalities in ceramide and cholesterol metabolism in brain aging and Alzheimer's disease. *Proc Natl Acad Sci U S A* 2004; **101**: 2070–5.
- 59 Mielke MM, Bandaru VVR, Han D, An Y, Resnick SM, Ferrucci L *et al.* Factors affecting longitudinal trajectories of plasma sphingomyelins: the Baltimore Longitudinal Study of Aging. *Aging Cell* 2015; **14**: 112–121.
- 60 Mielke MM, Bandaru VVR, Han D, An Y, Resnick SM, Ferrucci L *et al.*
Demographic and clinical variables affecting mid- to late-life trajectories of plasma ceramide and dihydroceramide species. *Aging Cell* 2015; **14**: 1014–1023.
- 61 Koal T, Klavins K, Seppi D, Kemmler G, Humpel C. Sphingomyelin SM(d18:1/18:0) is significantly enhanced in cerebrospinal fluid samples dichotomized by pathological amyloid- β 42, tau, and phospho-tau-181 levels. *J Alzheimers Dis* 2015; **44**: 1193–201.
- 62 Mielke MM, Haughey NJ, Han D, An Y, Bandaru VVR, Lyketsos CG *et al.* The Association Between Plasma Ceramides and Sphingomyelins and Risk of Alzheimer's Disease Differs by Sex and APOE in the Baltimore Longitudinal Study of Aging. *J Alzheimer's Dis* 2017; **60**: 819–828.
- 63 van Kruining D, Luo Q, van Echten-Deckert G, Mielke MM, Bowman A, Ellis S *et al.*
Sphingolipids as prognostic biomarkers of neurodegeneration, neuroinflammation, and psychiatric diseases and their emerging role in lipidomic investigation methods. *Adv Drug Deliv Rev* 2020. doi:10.1016/j.addr.2020.04.009.
- 64 Gonzalez CE, Venkatraman VK, An Y, Landman BA, Davatzikos C, Ratnam Bandaru VV *et al.* Peripheral sphingolipids are associated with variation in white matter

- microstructure in older adults. *Neurobiol Aging* 2016; **43**: 156–163.
- 65 Goñi FM, Alonso A. Sphingomyelinases: enzymology and membrane activity. *FEBS Lett* 2002; **531**: 38–46.
- 66 Kolesnick RN, Krönke M. REGULATION OF CERAMIDE PRODUCTION AND APOPTOSIS. *Annu Rev Physiol* 1998; **60**: 643–665.
- 67 Knottnerus SJG, Bleeker JC, Wüst RCI, Ferdinandusse S, IJlst L, Wijburg FA *et al.* Disorders of mitochondrial long-chain fatty acid oxidation and the carnitine shuttle. *Rev. Endocr. Metab. Disord.* 2018; **19**: 93–106.

Variable	N	Mean (SD)
Age at cognitive testing and plasma collection (years)	866	72.49 (0.71)
General fluid cognitive ability*	866	0 (1)
Age at brain scan (years)	731	72.68 (0.72)
Total brain volume (cm³)	657	990.32 (89.40)
Grey matter volume (cm³)	657	472.43 (44.68)
Normal appearing white matter volume (cm³)	657	476.89 (50.55)
White matter hyperintensity volume (cm³)	656	12.22 (12.18)
general fractional anisotropy*	668	0 (1)
general mean diffusivity*	668	0 (1)
Sex (male; female)	448; 418	-

Note: *These variables were standardized (with a mean of zero and standard deviation of one).

Table 1 Summary descriptive data for LBC1936

	Biocrates			B.I.-LISA			UPLC-MS-POS			UPLC-MS-NEG			NMR-NOESY			NMR-CPMG		
	R ²	SE	P	R ²	SE	P	R ²	SE	P	R ²	SE	P	R ²	SE	P	R ²	SE	P
<i>g</i>	0.079	0.041	0.054	0.010	0.013	0.477	0.071	0.038	0.064	0.046	0.030	0.120	0.006 [#]	0.009	0.538	0.016	0.018	0.355
TBV	0.033	0.026	0.213	0.106	0.046	0.020	0.185	0.051	0.00025	0.111	0.045	0.015	0.098	0.043	0.023	NA	NA	NA
GMV	0.051	0.032	0.113	0.108	0.046	0.018	0.171	0.049	0.001	0.108	0.047	0.022	0.108	0.052	0.037	0.110	0.052	0.036
NAWM	0.014	0.017	0.403	0.083	0.041	0.045	0.170	0.052	0.001	0.096	0.042	0.023	0.094	0.046	0.041	0.063	0.034	0.067
WMHV	0.002	0.006	0.807	0.040	0.032	0.212	0.002 [*]	0.006	0.747	0.002	0.006	0.732	0.004	0.008	0.591	0.003	0.007	0.707
<i>gFA</i>	0.005	0.008	0.540	0.005	0.007	0.496	0.005	0.009	0.563	0.016	0.016	0.344	0.004	0.009	0.667	0.004	0.008	0.583
<i>gMD</i>	0.012	0.015	0.425	0.018	0.017	0.290	0.004	0.007	0.550	0.008	0.011	0.444	0.003	0.006	0.634	0.005	0.007	0.526

Table 2 Results from LASSO prediction models. *g* = general cognitive ability, TBV = total brain volume, GMV = grey matter volume, NAWM = normal appearing white matter volume, WMHV = white matter hyperintensity volume *gFA* = general fractional anisotropy, *gMD* = general mean diffusivity. ^{*}500 iterations, [#]100 iterations were run rather than 1000 due to high computational demands. FDR corrected significant findings are indicated in b

